

First-order traffic flow models incorporating capacity drop: Overview and real-data validation

Maria Kontorinaki^{a,*}, Anastasia Spiliopoulou^a, Claudio Roncoli^b, Markos Papageorgiou^a

^a Dynamic Systems and Simulation Laboratory, Technical University of Crete, University Campus Kounoupidiana, Chania, 73100, Greece,

^b Department of Built Environment, Aalto University, Otakaari 4, Espoo, 02150, Finland.

* Corresponding author.

E-mail addresses: mkontorinaki@dssl.tuc.gr, natasa@dssl.tuc.gr, claudio.roncoli@aalto.fi, markos@dssl.tuc.gr.

ABSTRACT

Keywords:
Capacity drop,
LWR model,
First-order traffic flow
models,
Model calibration

First-order traffic flow models of the LWR (Lighthill-Whitham-Richards) type are known for their simplicity and computational efficiency and have, for this reason, been widely used for various traffic engineering tasks. However, these first-order models are not able to reproduce significant traffic phenomena of great interest, such as the capacity drop and stop-and-go waves. This paper presents an overview of modeling approaches, which introduce the ability to reflect the capacity-drop phenomenon into discretized LWR-type first-order traffic flow models; and also proposes a new approach. The background and main characteristics of each approach are analyzed with particular emphasis on the practical applicability of such models for traffic simulation, management and control. The presented modeling approaches are tested and validated using real data from a motorway network in the U.K.

1. Introduction

Among numerous phenomena characterizing traffic flow behavior, one of the most known and puzzling is the so-called capacity drop. This phenomenon breeds the reduction in the mainstream flow of a motorway when a queue starts forming upstream of a bottleneck location (Banks, 1991; Hall and Agyemang-Duah, 1991). Bottleneck locations can be motorway merge areas, areas with particular infrastructure layout (such as lane drops, strong grade or curvature, tunnels, etc.), areas with specific traffic conditions (e.g., strong weaving of traffic streams), areas with external capacity-reducing events (e.g. work-zones, incidents) etc. (Chung et al., 2007; Dixon et al., 1996; Smith et al., 2003). If the arriving demand is higher than the bottleneck capacity, i.e., the maximum flow that can pass during a certain time period, the bottleneck is activated, i.e., congestion is formed at the bottleneck location and spreads upstream. Empirical observations show that, whenever a bottleneck is activated, the maximum outflow that materializes (also called discharge flow) may be some 5 to 20 percent lower than the nominal bottleneck capacity. The capacity drop is then defined as the difference between these two values of flow, i.e., the capacity and the discharge flow. Certainly, the capacity drop reflects infrastructure performance degradation, leading to increased congestion space-time extent and accordingly longer vehicle delays. To avoid or delay the activation of a bottleneck, and the related capacity drop phenomenon, various traffic control measures have been proposed and applied (Cassidy and Rudjanakanoknad, 2005; Papageorgiou et al., 1991, 2003).

Designing and testing efficient traffic control strategies, or assessing their performance, require the usage of traffic flow models that are able to reproduce the motorway traffic conditions with satisfactory accuracy, specifically to reproduce infrastructure degrading phenomena such as the capacity drop. Macroscopic first-order traffic flow models of the Lighthill-Whitham-Richards (LWR) type (Lighthill and Whitham, 1955; Richards, 1956), where the dynamics are described by the conservation equation of vehicles only, represent a valuable tool for the study of traffic behavior, as they are simple, yet effective in reproducing not only free-flow conditions, but also wave formation and propagation under congested conditions. However, they do not allow for capturing more complex traffic phenomena, such as the capacity drop. In order to incorporate this important feature, different approaches have been proposed, which include higher-order extensions and first-order extensions; the former include second-order traffic flow models, while the latter are formulated via introduction of complex fundamental diagrams (Zhang, 2001).

Various second-order macroscopic traffic flow models have been proposed (e.g., Payne, 1971; Messmer and Papageorgiou, 1990; Aw and Rascle, 2000; Zhang, 2002; Whitham, 2011; Delis et al., 2014), which contain an additional dynamic equation to describe the speed evolution, being thereby capable to reproduce traffic instabilities, such as stop-and-go waves, as well as the capacity drop phenomenon at active bottlenecks. Second-order models have been consistently found in diverse calibration exercises to reflect more accurately real traffic data (Cremer and Papageorgiou, 1981; Papageorgiou et al., 1989; Michalopoulos et al., 1992; Frejo et al., 2012; Spiliopoulou et al., 2014; Fan and Seibold, 2012; Fan and Seibold, 2013). On the other hand, also second-order traffic model present some drawbacks, such as: (i) they may (under rare circumstances, as shown by Helbing and Johansson (2009)) produce negative speeds or flows (see Daganzo (1995a)); (ii) they usually include a higher number of parameters (some of which without clear physical significance), that need to be appropriately calibrated; and (iii) any optimization problem built upon second-order models is characterized by a nonlinear formulation, which implies a higher computation effort and the impossibility to guarantee convergence to a global optimum (Kotsialos and Papageorgiou, 2004). With respect to the last drawback, also first-order models include non-linearities, which, however, may be more efficiently tackled, while defining an optimization problem, by using computationally-efficient mixed-integer linear formulations (Ferrara et al., 2015) or, under specific assumptions, by using only linear or piecewise linear constraints (Ziliaskopoulos, 2000; Roncoli et al., 2015a).

The LWR model is composed by a single partial differential equation reflecting the conservation of vehicles and a steady-state flow-density relationship known as the Fundamental Diagram (FD) of traffic flow. While analytical solutions of the LWR model can be obtained for simple traffic settings using the method of characteristics, a significant amount of literature proposes and extends discrete approximations of the continuous LWR model applying the Godunov discretization scheme (Godunov, 1959), where the FD is transformed into two flux functions known as the demand (flow that can be sent from upstream) and the supply (flow that can be received downstream) functions (Lebacque, 1996). The most referenced among these discretized models is the Cell Transmission Model (CTM) (Daganzo, 1994), where the flow is expressed as a function of density via the definition of a triangular FD, and the space and time increments are selected according to the free speed. Remarkably, CTM realistically predicts shockwave propagations, while all the parameters have a physical interpretation, which also implies that they can be easily calibrated using real traffic data (Munoz et al., 2004). Furthermore, it has been employed for the study of different applications, such as dynamic traffic assignment (Lebacque et al., 1996; Ziliaskopoulos, 2000), traffic prediction, signal control and ramp metering (Alecsandru et al., 2011; Gomes and Horowitz, 2006; Zhang et al., 1996). Finally, CTM is characterized by relatively low computational requirements (Gomes and Horowitz, 2006; Lo, 2001) and it can be easily employed for large-scale motorway and urban network simulation (Lebacque et al., 1996).

It should be noted that CTM is not the only computationally efficient and reasonably accurate discretized first-order model. Models such as the Point-Queue (PQ) model, proposed by Smith (1983) and Kuwahara and Akamatsu (1997), as well as the Link Transmission Model (LTM), proposed by Yperman et al. (2006), have been seen as effective tools for traffic flow representation. As discussed by Nie and Zhang (2005), although the PQ model results in less computational cost and leads to equivalent results for a number of initial/boundary conditions, compared to CTM, its solutions differ in cases of queue spill-backs,

since the PQ model does not consider the physical length of the queue. Moreover, as discussed by Jin (2015b), LTM, in which the demand and supply functions are defined from cumulative flows, appears to be less computationally demanding but more memory consuming, in comparison with CTM.

Despite the increasing interest from the research community in integrating capacity drop in LWR-type first-order models, a limited number of effective approaches have been proposed, and only a few are actually tested using real traffic data to evaluate their behavior in case a bottleneck is activated. This study, as an extension of the work by Kontorinaki et al. (2016) in various respects, aims to fill this gap, gathering the state-of-the-art related to capacity drop modeling within LWR-type models (first-order extensions), contributing with further insights about their implications, and testing their capability to reproduce correctly the desired traffic pattern at an active bottleneck due to on-ramp merging. In particular, a discretized space-time modeling framework, which comprises CTM as a special case, is used. This is because CTM (and its extensions) serves as basis for most of the proposed capacity drop approaches, which use a discretized space-time framework. It should be noted that each approach introduces additional terms and parameters to the original CTM, and that the computational efficiency of all approaches remains virtually unaltered.

The selection criteria for the CTM-based approaches that are described and analyzed in this paper are two: First, the selected models should include a low number of parameters, which implies a limited effort in calibration and easier application; second, the selected models should be capable of reproducing the capacity drop for a typical on-ramp merge scenario. In addition, based on the above considerations, a new modeling approach is introduced, which extends also the discretized LWR model to reflect capacity drop. An analysis that links this model with the continuous LWR model is also provided in the Appendix.

The rest of the paper is structured as follows: Section 2 presents a literature review of first-order models that reflect capacity drop. Section 3 describes in more detail the selected approaches, highlighting the necessary modifications of a basic discretized-LWR formulation; in addition, a first-order model that is not of LWR-type is included for comparison. In Section 4, the approaches are tested using real traffic data from a motorway network in UK. Finally, in Section 5 some concluding remarks are provided.

2. Literature review on capacity drop in LWR models

Several modeling ideas have been proposed in the research community with the purpose of incorporating the capacity drop phenomenon into LWR-type models using different assumptions. In this section, a literature review of the proposed approaches is presented, along with an attempt to categorize them according to their underlying basic ideas and the qualitative issues arising from their implementation.

From a theoretical viewpoint, several issues might arise from including capacity drop, mainly in the continuum LWR model, via lower-order extensions. Some of them are briefly presented here since they might be useful in the following, but the reader is encouraged to read Zhang (2001) for a more detailed discussion. A first issue may be whether or not the considered extensions violate the anisotropic property; roughly speaking, the anisotropic property imposes that vehicles are only affected by the traffic conditions downstream of their location. Furthermore, in many cases, behavioral assumptions must be taken into account in order to consider one out of multiple mathematical solutions imposed by the extended model. More specifically, since lower-order extensions target also the consideration of the so-called non-equilibrium phase transitions, one should restrict the range of solutions by introducing intermediate states (additional conditions) on the model and selecting the preferable solution using specific criteria, such as, for example, assumptions on driver behavior. It is also important to investigate whether or not the extended model can indeed be reduced to an equivalent kinematic wave model, but with a special (“effective”) FD. Although these issues are not of major importance in the utilized discrete environment, this study also contributes with a theoretical analysis that discusses the latter issue (reduction of a model to a kinematic wave model) for the new proposed capacity drop approach (see Section 3.8 and Appendix).

Since the pioneering work by Edie (1961), it has been observed that the flow-density relation can be discontinuous, featuring a sharp speed drop within a small density range, when a critical density value is exceeded. Specifically, this discontinuity in the FD generally arises when congestion appears in the area of an active bottleneck and reflects the capacity drop phenomenon. The resulting particular shape of the FD

has sometimes been referred as “inverse lambda”, see Koshi et al. (1983). This behavior can be theoretically modeled via definition of two flow values for a specific range of densities around the critical density, where the different flows appear in dependence of the current and past traffic conditions. Nevertheless, some research works suggested that it is more appropriate to employ a continuous FD rather than a discontinuous one, since the latter may cause (in a continuous environment) infinite shock-wave and characteristic wave speeds, as discussed by Jin et al. (2015a); moreover such a discontinuity differs from observations in the field (Cassidy, 1998).

Various approaches to enable the description of the aforementioned behavior have been proposed in different works. Muralidharan and Horowitz (2015) and Li et al. (2015) proposed to reduce the outflow of a cell to a fixed value (lower than its nominal capacity) by modifying the demand function when the density of the cell becomes overcritical, leading to a FD of the “inverse lambda” shape. More specifically, Muralidharan and Horowitz (2015) proposed an augmented Link-Node Cell Transmission Model (LN-CTM), which is utilized in the formulation of an optimization problem; while Li et al. (2015) utilized a CTM variation accompanied with a stochastic component (added in order to reproduce stop-and-go waves) and proposed a methodology to reduce crash risks via Variable Speed Limits (VSL). Furthermore, Jin (2010) proposed a model that takes into account lateral and longitudinal movements of vehicles, in order to study the aggregated traffic dynamics of a motorway, including lane-changing effects. More specifically, a modified FD using an “inverse-lambda” shape is proposed by adding a parameter, which captures the intensity of lane-changing effects.

Other works considered the capacity drop mechanism being produced as a consequence of microscopic phenomena, such as lane-changing maneuvers, slow vehicles entering a merge cell, and heterogeneous lane behavior due to the variations of traffic states at merges, which prevent the system to reach the full motorway capacity before the breakdown (Cassidy and Ahn, 2005; Laval and Daganzo, 2006; Treiber et al., 2006). Again, Muralidharan and Horowitz (2015) proposed an additional weaving parameter affecting the supply function within the LN-CTM in order to capture the intensity of lane changing maneuvers. As previously mentioned, a similar feature is also present in the model proposed by Jin (2010). Moreover, Wong and Wong (2002) presented an extension of the LWR model incorporating the distribution of heterogeneous road users. The developed multi-class model assumes the existence of drivers with different speed choice behaviors and is capable to reproduce capacity drop, hysteresis of traffic flow and platoon dispersion as well.

Other studies incorporated the ability of reproducing capacity drop into LWR-type models by considering explicitly its phenomenological aspects, which consists in the appearance of the capacity drop immediately after queues are forming upstream of the bottleneck location, by letting the downstream supply be smaller than the upstream demand. Torné et al. (2014) proposed a modeling approach, within a so-called Capacity-Lagged CTM (CL-CTM), which modifies the FD from a triangular to a trapezoidal shape in case VSL are applied. This is materialized via definition of appropriate rules to switch from capacity to a reduced discharge flow, where the two flow values (capacity and discharge flow) are chosen a priori. Similarly, Jin et al. (2015a) modifies the supply function by introducing an exogenously specified reduced capacity which is activated when the demand of the cell is lower than its supply. Han et al. (2016) and Han et al. (2017) employed a similarly modified FD, with the difference that the capacity of the bottleneck cell reduces (linearly) as the density of the upstream cell (the congested one) increases. Furthermore, Landman et al. (2015) employed the macroscopic first-order multilane model proposed by van Lint et al. (2008), which enables capacity drop by decreasing the supply function of the cells located downstream of a congested one, using a pre-specified factor. Srivastava and Geroliminis (2013) proposed a memory-based methodology, where two different values of density are chosen in order to determine whether a cell is in free-flow or congested state; whenever a cell is congested, the corresponding supply function is bounded by a pre-specified flow lower than its capacity. Thus, in this formulation, (at least) two additional parameters need to be specified: one (or two) density threshold(s), characterizing the interval where flow capacity drop is appearing, and the discharge flow. The main disadvantage of this approach is the generation of high-frequency fluctuations between congested and uncongested states that does not allow a cell to remain in congested state for a long period. The authors suggest using moving averages to smooth the density variations while applying the switching logic; however, this may not be sufficient to eliminate the

aforementioned problem, since fluctuations with lower frequency and larger amplitude may be still observed.

Other researchers have tried to incorporate capacity drop within the LWR framework by accounting for the bounded acceleration of vehicles entering a bottleneck location. These models predict the same trajectories with LWR in some areas of the computational domain, while satisfying kinematic constraints (imposed by bounded acceleration) in other areas. For instance, Lebacque (2003) proposed a two-phase traffic flow model, where the first phase corresponds to the LWR model, while in the second phase the acceleration of vehicles is constant and equal to a maximum value; this model results in a higher-order model which is difficult to analyze and calibrate. Moreover, also a simplified approach that proposes only a modification of the FD is presented in the same study, while similar assumptions have been made by Monamy et al. (2012) for a link-node model that has been also partially validated with real data for a merging scenario. Moreover, Srivastava et al. (2015) proposed a modified CTM with a different demand function, which linearly decreases under over-saturated density conditions. In addition, Khoshyaran and Lebacque (2015) extended the model proposed by Monamy et al. (2012) by considering an internal state node model in order to account with the bounded acceleration and the introduction of a simple mechanism for modeling hysteresis cycles.

The simplified model proposed by Lebacque (2003) and the models proposed by Monamy et al. (2012), Srivastava et al. (2015) and Khoshyaran and Lebacque (2015) utilize a decreasing demand function for densities beyond a critical value to reflect bounded acceleration of vehicles, while the supply function remains the same. However, this model feature is capable to reproduce the capacity drop phenomenon in on-ramp merges only for specific cases; in fact, if the flow that the merge cell can receive is smaller than its capacity (due to the unchanged supply function), the desired effect is cancelled, because the merge cell cannot become congested and continues operating at capacity. Following the above concept, Karafyllis et al. (2016) introduced generalized versions of discrete approximations of the LWR model, allowing for a wide range of demand functions to be taken into account, where the capacity drop can be included through the definition of linearly decreasing or even discontinuous demand functions for overcritical densities; while Roncoli et al. (2015a) included the possibility of capacity drop into a multi-lane first-order traffic flow model in order to define a quadratic programming optimization problem (Roncoli et al., 2015b); a modified FD is used, similar to the one proposed by Lebacque (2003), and the capacity drop is triggered by lateral and on-ramp flows.

Furthermore, Leclercq et al. (2011) proposed to quantify capacity drop as a consequence of bounded acceleration of merging vehicles. The capacity drop is defined as a function of the on-ramp demand considering two possible demand scenarios (low/high demand) and a set of different model parameters. The basis of this work is the Newell-Daganzo (ND) model (Newell, 1982; Daganzo, 1995b), enhanced by the introduction of different formulas and rules to compute the capacity of the merge cell depending on the on-ramp and mainstream demands. However, this approach focuses mainly on determining the value of the capacity drop, without advising on any methodology for its implementation into discretized LWR-type models. Extending the previous concept, Leclercq et al. (2016) developed a methodology for estimating the effective capacity (discharge flow), i.e., the observed flow downstream of freeway merges. The authors focused on the impact of merging process, combined by bounded acceleration, with heterogeneous vehicle characteristics. They concluded that for the estimation of the effective capacity, a proper estimation of vehicles mean characteristics would be sufficient.

Models that count for the uncertainty of the FD have also been considered. Li et al. (2012) analysed the potential influences caused by the inaccurate specification of the FD by developing a numerical procedure to evaluate how the uncertainty of the FD would affect the output of the LWR model. Moreover, Jin and Amin (2017), based on the uncertain nature of capacity, proposed a modified CTM incorporating a Markov chain-based model for the saturation regimes; however, their work is mainly focused on the calibration method rather than analyzing the qualitative behavior of the proposed model.

Last but not least, some researchers have tried to incorporate the capacity drop phenomenon into LWR-type models through the hysteresis phenomenon of traffic flow (see, e.g. Alvarez-Icaza and Islas, 2013; Yuan et al., 2015), which was first reported by Treiterer and Myers (1974) and derives from the fact that the acceleration and the deceleration of vehicles are not symmetric procedures. This means, that,

whenever the traffic is moving from free-flow to congested regime, the observed flow reaches the bottleneck capacity; while, the transition from congested regime to free-flow occurs via a maximum flow which is lower than the capacity. Thus, two different branches may exist in the FD, whereby the deceleration branch lies above the acceleration branch. It is worth noting that second-order models employing a continuous FD may naturally produce the hysteresis behavior of traffic flow and the capacity drop, thanks to the included dynamic speed equation (see e.g. Papageorgiou et al., 1990, Fig. 14).

In this study several of the proposed approaches are tested in order to provide further insights about their implications and their capability to reproduce correctly the desired traffic patterns at an active bottleneck due to on-ramp merging. The selected modeling approaches are those that are simple in terms of the number of parameters and can be formulated within a simple general, discretized LWR (CTM-like) model framework. In particular, the following modeling approaches are selected to be tested in the next section: i) the “inverse lambda” shape FD modeling approach proposed by Muralidharan and Horowitz (2015) and Li et al. (2015) in Section 3.6; ii) the model, which introduces a weaving parameter for lane-changing effects, proposed by Muralidharan and Horowitz (2015) in Section 3.5; iii) the modeling approach proposed by Torné et al. (2014) with appropriate modifications for the case of an on-ramp merge in Section 3.4; iv) a similar modeling approach with the one proposed by Han et al. (2016) (and further utilized by Han et al. (2017) for traffic control purposes) which utilizes linearly decreasing capacity for the downstream supply in Section 3.7; and finally, v) a new modeling approach extending the ones proposed by Lebacque (2003) and Monamy et al. (2012) in Section 3.8.

3. Model formulations

For the subsequent description and testing of different capacity-drop approaches, a simple formulation of a discretized LWR model is utilized. Despite the fact that some of the considered approaches are originally introduced for more sophisticated models, their implementation is here based on a common formulation, which also permits a clearer understanding and a fairer result comparison. Also the notation (see Table 1) is kept consistent throughout the paper for all the described approaches. Moreover, a simple illustrative example is constructed, which demonstrates the qualitative behavior of the different approaches that are tested hereafter.

3.1. Basic discretized-LWR model formulation

A discretized LWR model considers, apart from the discretization of time, the discretization of the network in a finite number of cells and the definition of rules for sending and receiving traffic flow. To this end, a motorway stretch is divided into n cells as shown in Fig. 1. The i^{th} cell (where $i = 1, \dots, n$) is characterized by a single state variable ρ_i , which corresponds to the density of vehicles, namely the number of vehicles divided by the length of the cell, implying that the state of the motorway is entirely described by the n -dimensional vector $\rho = (\rho_1, \dots, \rho_n)$ evolving according to a n -dimensional nonlinear difference equation. The movement of vehicles from one cell to the next is governed by the steady-state relation between flow and density, i.e. the corresponding FD. This relation is characterized by its concave branches, where the demand part and the supply part reflect, respectively, its increasing and decreasing branches (Godunov, 1959; Lebacque, 1996). The density value (unimodal FDs) or values (multimodal FDs), for which the maximum flow is attained, is defined as the critical density of the cell. The general model is entirely described by Eqs. (1), (2), (3), and (4), while the definitions of the variables and parameters are given in Table 1.

$$\rho_1(k+1) = \rho_1(k) + \frac{T}{l_1 L_1} \left(-\max \left(0, \frac{f_1(\rho_1(k), \rho_2(k), r_2(k))}{1 - p_1} \right) + \min \left(r_1(k), f_{S,1}(\rho_1(k)) \right) \right), \quad (1)$$

$$\rho_i(k+1) = \rho_i(k) + \frac{T}{l_i L_i} \left(f_{i-1}(\rho_{i-1}(k), \rho_i(k), r_i(k)) - \max \left(0, \frac{f_i(\rho_i(k), \rho_{i+1}(k), r_{i+1}(k))}{1-p_i} \right) + r_i(k) \right) \text{ for}$$

$$i = 2, \dots, n,$$

$$f_i(\rho_i(k), \rho_{i+1}(k), r_{i+1}(k)) = \min \{ (1-p_i) f_{D,i}(\rho_i(k)), f_{S,i+1}(\rho_{i+1}(k)) - r_{i+1}(k) \} \text{ for } i = 1, \dots, n-1,$$

$$f_n(\rho_n(k)) = f_{D,n}(\rho_n(k)),$$

$$f_{D,i}(\rho_i(k)) = \min \{ g_i(\rho_i(k)), Q_i \} \text{ for } i = 1, \dots, n$$

$$f_{S,i+1}(\rho_{i+1}(k)) = \min \{ Q_{i+1}, w_{i+1} (\rho_{\max,i+1} - \rho_{i+1}(k)) \}_{i+1} \text{ for } i = 1, \dots, n-1.$$

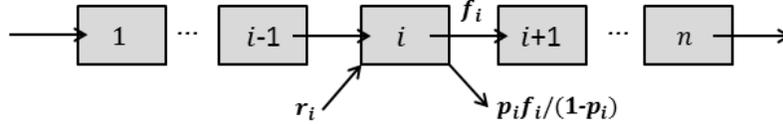


Fig. 1. The space-discretization of a hypothetical motorway stretch.

Notice that, according to this formulation, r_i is the factual on-ramp flow, which may be only a portion of the corresponding on-ramp demand. This portion may reflect any applying priority rules at the junctions of the freeway and can be fully determined using appropriate existing priority models (see, for example, Karafyllis et al., 2016; Daganzo, 1995). However, in this study, we are focusing on capturing the capacity drop at the mainstream region and we dispose of factual (measured) on-ramp flows, hence, there is no need and, in fact, no possibility, in lack of internal ramp data, to explicitly consider a specific priority

Table 1
Models' variables and parameters.

Symbol	Name	Units
T	simulation time step	h
k	discrete time index	0,1,2,...
l_i	number of lanes in the i^{th} cell	dimensionless
L_i	length of the i^{th} cell	km
$\rho_i(k)$	density of the i^{th} cell at time kT	veh/km/lane
$\rho_{\max,i}$	storage capacity of the i^{th} cell	veh/km/lane
$\rho_{cr,i}$	critical density of the i^{th} cell	veh/km/lane
$f_i(\rho_i(k), \rho_{i+1}(k), r_{i+1}(k))$	actual outflow from the i^{th} to the $(i+1)^{\text{th}}$ cell during $(k, k+1]T$	veh/h
$f_{D,i}(\rho_i(k))$	demand part of the FD of the i^{th} cell during $(k, k+1]T$	veh/h
$f_{S,i}(\rho_i(k))$	supply part of the FD of the i^{th} cell during $(k, k+1]T$	veh/h
Q_i	capacity flow of the i^{th} cell	veh/h
$r_i(k)$	factual on-ramp flow at the cell during $(k, k+1]T$	veh/h
p_i	percentage of the actual flow exiting from the off-ramp of the i^{th} cell (exit-rate of the i^{th} cell)	dimensionless
$v_{f,i}$	free-flow speed of the i^{th} cell	km/h
w_i	congestion wave speed of the i^{th} cell	km/h

scheme. Thus, the factual on-ramp flow is not restricted by the amount of vehicles within the cell, while restrictions (the supply part of the FD) apply for the flow coming from the upstream cell. Note also, that for the first cell, the entering mainstream flow is modeled as an on-ramp, thus considered as r_1 ; this implies that any appearing congestion in the stretch should never reach the upstream boundary. Finally, the downstream end of the freeway is assumed uncongested.

Finally, notice that in (3) and (4) the demand and supply functions are completed by assuming capacity flow values Q_i for overcritical and undercritical densities, respectively. Consequently, the model predicts capacity flow even when congestion is created (no capacity drop), in accordance with the non-discretised LWR model. Note also, that the right-hand side of the FD in (4) is described by a linear function (with a negative slope w_i); while the left-hand side of the FD in (3) is assumed to be a non-decreasing function $g_i(\rho_i)$.

3.2. Different shapes for the FD

Different functions $g_i(\rho_i)$ can be used for the demand function in (3). The original CTM formulation by Daganzo (1994) considers a triangular-shaped FD (Fig. 2(a)), where $g_i(\rho_i) = v_{f,i}\rho_i l_i$, $g_i(\rho_{cr,i}) = Q_i$ and $w_i = Q_i / ((\rho_{max,i} - \rho_{cr,i})l_i)$ (with $w_i < v_{f,i}$ and $v_{f,i} < L_i/T$). This formulation has two main drawbacks: first, when using realistic free-flow and congestion-wave speeds, it may lead to high (and sometimes unrealistic) capacity flow; second, only one speed value is considered for all under-critical densities, which is often not

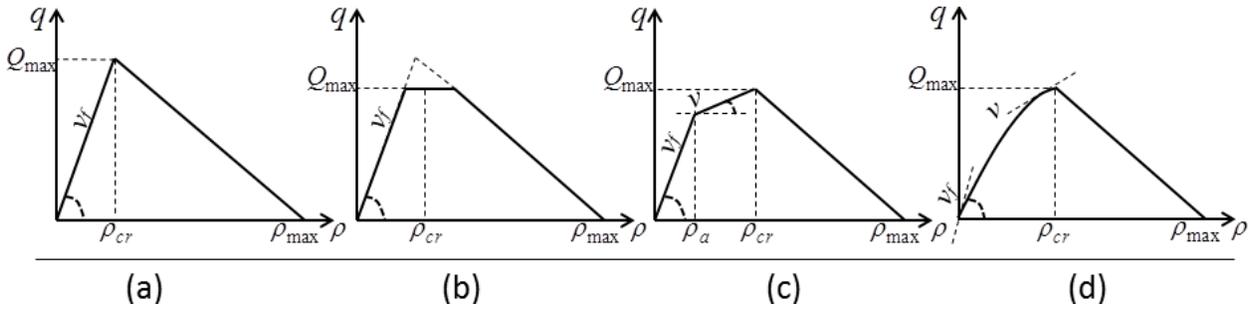


Fig. 2. Different choices for the left-hand side of the fundamental diagram corresponding to: (a) a triangular FD (CTM), (b) a trapezoidal FD, (c) a piecewise linear FD and (d) a nonlinear FD.

compatible with traffic observations. To overcome the first issue, a trapezoidal FD can be used, where $g_i(\rho_i) = v_{f,i}\rho_i l_i$, $g_i(\rho_{cr,i}) \geq Q_i$ and $w_i \geq Q_i / ((\rho_{max,i} - \rho_{cr,i})l_i)$, as illustrated in Fig. 2(b). In this case, the critical density, instead of being unique for both the FD parts, can be selected within an interval of densities, thus increasing also the degree of freedom for model calibration. Nevertheless, in real traffic, the observed speed may be characterized by a decreasing-behavior also for undercritical densities, which can be reflected by using a nonlinear concave function g_i (Fig. 2(d)), where $g_i(\rho_{cr,i}) = Q_i$ and $w_i = Q_i / ((\rho_{max,i} - \rho_{cr,i})l_i)$. An opportune calibration of such function may lead to more realistic results. As an example, a nonlinear exponential function, as proposed by Messmer and Papageorgiou (1990), can be employed (see, Eq. (17)). A similar behavior can also be obtained, with similar accuracy, considering a piecewise-linear approximation of the nonlinear function (Fig. 2(c)), which is helpful, for instance, in case linear constraints are needed for the formulation of an optimization problem (see e.g., Ziliaskopoulos, 2000; Roncoli et al., 2015a).

3.3. Illustrative example

In order to illustrate the behavior of each approach, a simple hypothetical motorway stretch is considered, consisting of a set of $n = 15$ homogeneous cells of equal length (500 m) and common FD parameters. The motorway stretch includes an on-ramp which is located at the upstream boundary of the cell $i = 13$. The parameters that characterize the network are shown in Table 2. For the sake of simplicity, in all the following tests, the function g_i utilized in the Eq. (3) is selected to be $g_i(\rho_i) = v_{f,i} \rho_i l_i$ (for $i = 1, \dots, 15$), leading to triangular-shaped FD (Fig. 2(a)). A hypothetical trapezoidal traffic demand scenario is applied to the network (see Fig. 3), which, for some time period, generates a flow higher than the capacity at the merge area, thus generating congestion that spills back for some cells, however without reaching the network origin. In order to initialize the system from a steady state, the density for every cell at the very first time instant is set $\rho_i(0) = 11.7$ veh/km/lane for $i = 1, \dots, 12$ and $\rho_i(0) = 13.3$ veh/km/lane for $i = 13, 14, 15$. The simulation time horizon is $T^{hor} = 4$ h for all the following tests.

Table 2

Parameters characterizing the utilized hypothetical network (same for all approaches).

τ	n	L_i	l_i	$\rho_{cr,i}$	$v_{f,i}$	$\rho_{max,i}$	Q_i	w_i
5/3600	15	0.5	3	20	100	120	6000	20

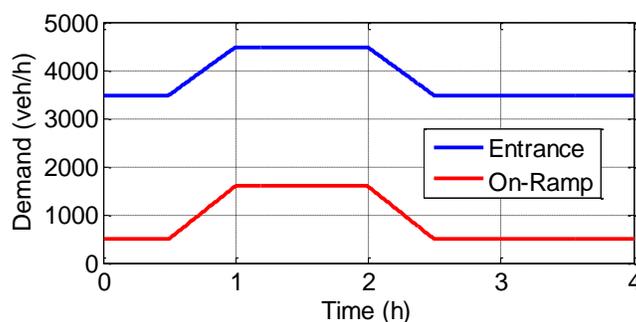


Fig. 3. Traffic demand scenario for the mainstream motorway (blue line) and the on-ramp (red line).

Fig. 4 (a), (b), (c) illustrates some significant characteristics of the CTM in case congestion is created at an on-ramp merge. Once the total demand (in this case, the sum of mainstream and factual ramp flows) exceeds the bottleneck's capacity, only a portion of the available mainstream flow is allowed to access the 13th cell. This causes an increase of density in the upstream cell (the 12th cell) (see Fig. 4(a), red line), which eventually enters into a congested state, generating a congestion wave that propagates to further upstream cells. During this period, the density in the merge cell remains at its critical value (see Fig. 4(a), blue line), allowing an outflow equal to capacity (Fig. 4(c)). As a consequence of the observations above, the speed at the 12th cell decreases (Fig. 4(b), red line), while the speed at the merge cell (13th) remains constant and equal to the free speed. Note that, in contrast to this modeled behavior, the merge cell is typically congested in real traffic; while the exit flow is reduced upon the onset of congestion due to capacity drop.

3.4. Approach 1: Switching logic for maximum flow

One effective way to implement a FD characterized by the inverse-lambda shape is via the definition of an opportune switching logic to define dynamically the current maximum flow. An example can be found in the study proposed by Torné et al. (2014), where a set of rules is proposed to impose capacity drop in case VSL are applied in a certain area of the network. The concept is based on the coexistence of two FDs for the

same location: a triangular-shaped one, active in case VSL are not applied (and, thus, no congestion is present); and a trapezoidal-shaped one, characterized by lower capacity that materializes in case congestion is present. This method can be extended straightforwardly to the case of bottlenecks due to lane drops, tunnels, etc.; in addition, it is shown here that it is also effective in case congestion is generated because of a merging on-ramp. The formulation is described by (1), (2), and:

$$f_{D,i}(\rho_i(k)) = \min \{g_i(\rho_i(k)), R_i(k)\} \text{ for } i=1, \dots, n, \quad (5)$$

$$f_{S,i+1}(\rho_{i+1}(k)) = \min \{R_{i+1}(k), w_{i+1}(\rho_{\max,i+1} - \rho_{i+1}(k))l_{i+1}\} \text{ for } i=1, \dots, n-1, \quad (6)$$

where

$$R_1(k) = Q_1, \quad R_2(k) = Q_2, \quad (7)$$

$$R_{i+1}(k+1) = \begin{cases} \bar{Q}_{i+1} & \text{if } w_i(\rho_{\max,i} - \rho_i(k))l_i < \min(f_{i-1,D}(\rho_{i-1}(k)), R_i(k)) \\ Q_{i+1} & \text{o.w.} \end{cases} \text{ for } i=2, \dots, n-1. \quad (8)$$

R_i are auxiliary variables that define the maximum flow for cell i , and \bar{Q}_i corresponds, for this modeling approach, to the queue discharge flow observed after the congestion onset. \bar{Q}_i can also be viewed as $\bar{Q}_i = \alpha Q_i$, i.e., a portion $\alpha < 1$ of the capacity flow. For this simulation test, this portion is constant and equal to $\alpha = 0.95$. Eq. (7) reflects the assumption that the backspilling congestion does not reach the entrance of the network. Moreover, all cells are initially uncongested, thus $R_i(0) = Q_i$, for every $i=1, \dots, n$.

Fig. 4 (d), (e), (f), illustrates the behavior resulting from the application of this approach. The main idea lies in decreasing the capacity of the cell located immediately downstream of a congested one. More specifically, when the total flow (on-ramp and mainstream) exceeds the capacity of the merge cell, the density of the upstream cell starts increasing (see Fig. 3(d), red line), while at the same time its speed starts decreasing (see Fig. 3(e), red line); consequently, after some time, its supply function becomes smaller than the demand function of the upstream cell; this, according to (8), triggers a reduction of the maximum flow for the downstream cell (see Fig. 3(f)), which persists until the overall demand is sufficiently decreased. As a drawback, the flow reduction appears with some delay after the congestion starts, since this reduction materializes only when both the demand flow of the 11th cell and the maximum flow of the 12th cell become higher than the supply of the 12th cell. Furthermore, it is interesting to point out that, despite the flow-drop, there is no congestion, i.e. no over-critical density (Fig. 3(d), blue line), and therefore also no speed-drop (Fig. 3(e), blue line), at the merge cell.

3.5. Approach 2: Introduction of a weaving parameter

Another option to achieve a reduced outflow at a merge cell is via the introduction of a weaving parameter that essentially affects the supply function at the merge cell, as proposed by Muralidharan and Horowitz (2015). The purpose of this parameter is to take into account the ‘‘intensity’’ of lane changing maneuvers performed by vehicles just entered from the on-ramp, imposing a reduction of the available space for vehicles coming from upstream. The mathematical formulation consists of (1), (9), (3), and (4) where

$$f_i(\rho_i(k), \rho_{i+1}(k), r_{i+1}(k)) = \min \{(1 - p_i)f_{D,i}(\rho_i(k)), f_{S,i+1}(\rho_{i+1}(k)) - \eta_r r_{i+1}(k)\}, \text{ for } i=1, \dots, n-1, \quad (9)$$

where $\eta_r > 1$ is the weaving parameter that aims to reduce the available space at the merge cell, thus limiting farther the mainstream flow entering the cell. For a given η_r , the resulting outflow from the merge cell, namely the queue discharge flow, is a function of the on-ramp flow. It can be seen from Fig. 4(i) (where $\eta_r = 1.2$ is used), that the capacity flow is never reached even in case of low factual on-ramp flow; for this reason, the flow reduction is barely visible. Notice also that, similarly to Approach 1, the merge cell is not congested (Fig. 4(g), (h), blue line).

3.6. Approach 3: Reduction of the demand function

Another formulation, also utilized by Muralidharan and Horowitz (2015) and Li et al. (2015), consists of the definition of a discontinuous demand part of the FD at bottleneck locations. More specifically, a flow lower than the capacity is defined, which materializes once the density of the cell becomes overcritical. In particular, the model can be described by (1), (2), (4), and:

$$f_{D,i}(\rho_i(k)) = \begin{cases} g_i(\rho_i(k)) & \text{if } \rho_i(k) \leq \rho_{cr,i} \\ \bar{Q}_i & \text{o.w.} \end{cases} \quad \text{for } i = 1, \dots, n. \quad (10)$$

This approach produces the normal behavior of LWR model when the density of the upstream of the merge cell (here the 12th cell) is undercritical (Fig. 4(j)), leading the merge cell to reach properly capacity flow; then, whenever the density becomes overcritical, the outflow from the upstream of the merge cell drops to a value corresponding to \bar{Q}_i which in turn leads the outflow from the merge cell (discharge flow) to a value equal to $\bar{Q}_i + r_{i+1}(k)$ (Fig. 4(l)); in this case, $\bar{Q}_i = \alpha Q_i$, and $\alpha = 0.7$ are employed. The flow drop can be observed only if the value of the parameter *alpha* is selected to ensure that $\bar{Q}_i + r_{i+1}(k) < Q_{i+1}$. The main drawback of this approach is that traffic congestion persists longer than in the other cases, because, once formed, its disappearance can only be triggered by a sufficient decrease of the arriving demand (that must become smaller than \bar{Q}_i), irrespectively of any variation of the ramp inflow. Again, no congestion appears at the merge cell (Fig. 4(k), (j), blue line).

3.7. Approach 4: Linear reduction of maximum flow

As previously mentioned, the presence of capacity drop within traffic flow models plays a key role for the design and testing of motorway traffic control strategies. Among others, model-based control problems have been widely exploited in recent years because of the possibility to explicitly consider system dynamics and physical constraints. In some works, the classic formulation of first-order models was implemented via use of integer variables and opportune switching rules (see e.g. Ferrara et al., 2015; Muralidharan and Horowitz, 2015; Sun and Horowitz, 2005). In other works, (see e.g. Roncoli et al., 2015a; Ziliaskopoulos, 2000), linear inequalities (derived from the piecewise linear FD) were considered as constraints in the optimization problem; hereafter, some variants of these models are presented, which allow to define linearly constrained formulations for corresponding optimization problems. Here, a similar formulation as the one proposed by Han et al. (2016) is presented.

A concept similar to Approach 1 is considered, albeit with the introduction of an additional linear term that reduces the supply function of a downstream cell. Specifically, when congestion starts in cell i ($\rho_i > \rho_{cr,i}$), the maximum flow materialized within the supply term of the downstream cell $i+1$ is linearly decreased as a function of ρ_i , according to the following equations:

$$f_{S,i+1}(\rho_{i+1}(k)) = \min \{ F_{i+1}(\rho_i(k)), w_{i+1}(\rho_{\max,i+1} - \rho_{i+1}(k)) \}_{i+1} \quad \text{for } i = 1, \dots, n-1, \quad (11)$$

where $F_{i+1}(\rho_i(k))$ is given by:

$$F_{i+1}(\rho_i(k)) = \begin{cases} Q_{i+1} & \text{if } \rho_i(k) \leq \rho_{cr,i} \\ \bar{Q}_{i+1} + \frac{Q_{i+1} - \bar{Q}_{i+1}}{\rho_{cr,i} - \rho_{\max,i}} (\rho_i(k) - \rho_{\max,i}) & \text{o.w.} \end{cases} \quad \text{for } i = 1, \dots, n-1 \quad (12)$$

with $\bar{Q}_i = \alpha Q_i$; where in the illustrative example $\alpha = 0.9$. The proposed formulation is thus given by (1), (2), (3), (11), and (12). For under-critical densities, F_i is constant and equal to the capacity flow; however,

in case the density of the i^{th} cell increases beyond its critical value (Fig. 4(m), red line), the maximum flow of the supply function of the $(i+1)^{\text{th}}$ cell is reduced linearly (Fig. 4(o)). Therefore, the queue discharge flow from the bottleneck's cell depends on the factual on-ramp flow which imposes higher density values for the upstream cells. It should be mentioned here that in case there is no particular need for linear formulations, other functions can also be considered for the above reduction of the maximum flow. This approach appears to work appropriately also for bottlenecks due to capacity reduction (e.g., lane drops, tunnels). Differently from Approach 2, the model is capable of reaching capacity before congestion starts, and the capacity drop appears with a shorter delay with respect to the one observed in Approach 1. On the other hand, the merge cell remains uncongested, similarly to all previous approaches.

3.8. Approach 5: Increased space for vehicles entering a bottleneck location

Yet another approach may be conceived, which, in contrast to all previous approaches, allows for the bottleneck (e.g. merge) cell to become congested (as in real traffic observations), while producing a reduced outflow from the merge cell as well. Two different mechanisms are employed in order to achieve this behavior. The first one is activated in case of high on-ramp flow and imposes that the merge cell is able to receive more flow than its nominal capacity. To this end, in case a cell contains an external on-ramp, the flux function f_i is modified via the introduction of the parameter $\theta_r < 1$. In the basic discretized LWR formulation, the supply function of a cell becomes active (smaller than the demand function of the previous cell), when the density of this cell is equal or exceeds its critical value. The introduction of θ_r acts as a delay in the activation of the supply function by reducing it, in terms of the on-ramp flow, less than the real on-ramp flow would otherwise impose (in contrast to Approach 2, where the parameter η_r hastens the activation of the supply function). Essentially, this contributes to the increase of the available space for the upstream mainstream flow entering the merge cell. As a result, the density at merge cell is allowed to increase beyond the critical ($\rho_i > \rho_{cr,i}$), and therefore become congested. Notice that the inflow of the merge cell can exceed its nominal capacity, but the same does not hold for its outflow, since the latter is determined by its demand function. However, this cannot produce any reduced outflow (capacity drop) by itself. The second mechanism consists in the introduction of a decreasing demand function for overcritical densities, similar to the one proposed by Lebacque (2003) and Monamy et al. (2012). Due to the fact that the density of the merge cell can now be overcritical (thanks to the application of the first mechanism), the decreasing demand function imposes smaller outflow, i.e. a capacity drop. There is a large range of decreasing functions that can be considered. Here, for simplicity, a linearly decreasing function is selected. Notice that the proposed demand functions are continuous, in contrast with the demand functions employed in Approach 3. The mathematical formulation consists of (1), (13), (14), and (4) where:

$$f_i(\rho_i(k), \rho_{i+1}(k), r_{i+1}(k)) = \min \left\{ (1 - p_i) f_{D,i}(\rho_i(k)), f_{S,i+1}(\rho_{i+1}(k)) - \theta_r r_{i+1}(k) \right\} \text{ for } i = 1, \dots, n-1, \quad (13)$$

$$f_{D,i}(\rho_i(k)) = \begin{cases} g_i(\rho_i(k)) & \text{if } \rho_i(k) \leq \rho_{cr,i} \\ \bar{Q}_i + \bar{Q}_i \frac{\rho_i(k) - \rho_{cr,i}}{\rho_{cr,i} - \rho_{\max,i}} & \text{o.w.} \end{cases} \text{ for } i = 1, \dots, n. \quad (14)$$

In the illustrative example, $\theta_r = 0.7$ and $\bar{Q}_i = \alpha Q_i$ are used, where $\alpha = 0.4$. Fig. 4(p), (q), (r) illustrates the above considerations. Specifically, by the time the merge cell becomes congested (Fig. 4(p), blue line), producing also the corresponding speed drop (Fig. 4(q), blue line), and the corresponding capacity drop is observed for the flow exiting the merge cell (Fig. 4(r)). The magnitude of this drop and the resulting queue discharge flow are determined by the combination of the values of parameters α , θ_r and the on-ramp flow. At the same time, θ_r affects also the magnitude of the increase of the density at the merge cell, i.e., reducing θ_r produces a higher density increase at the merge cell, resulting in a higher speed. Finally, it is interesting to point out that the congestion is first created at the merge cell and the flow drop occurs immediately after the maximum flow is reached, in accordance with real traffic observations.

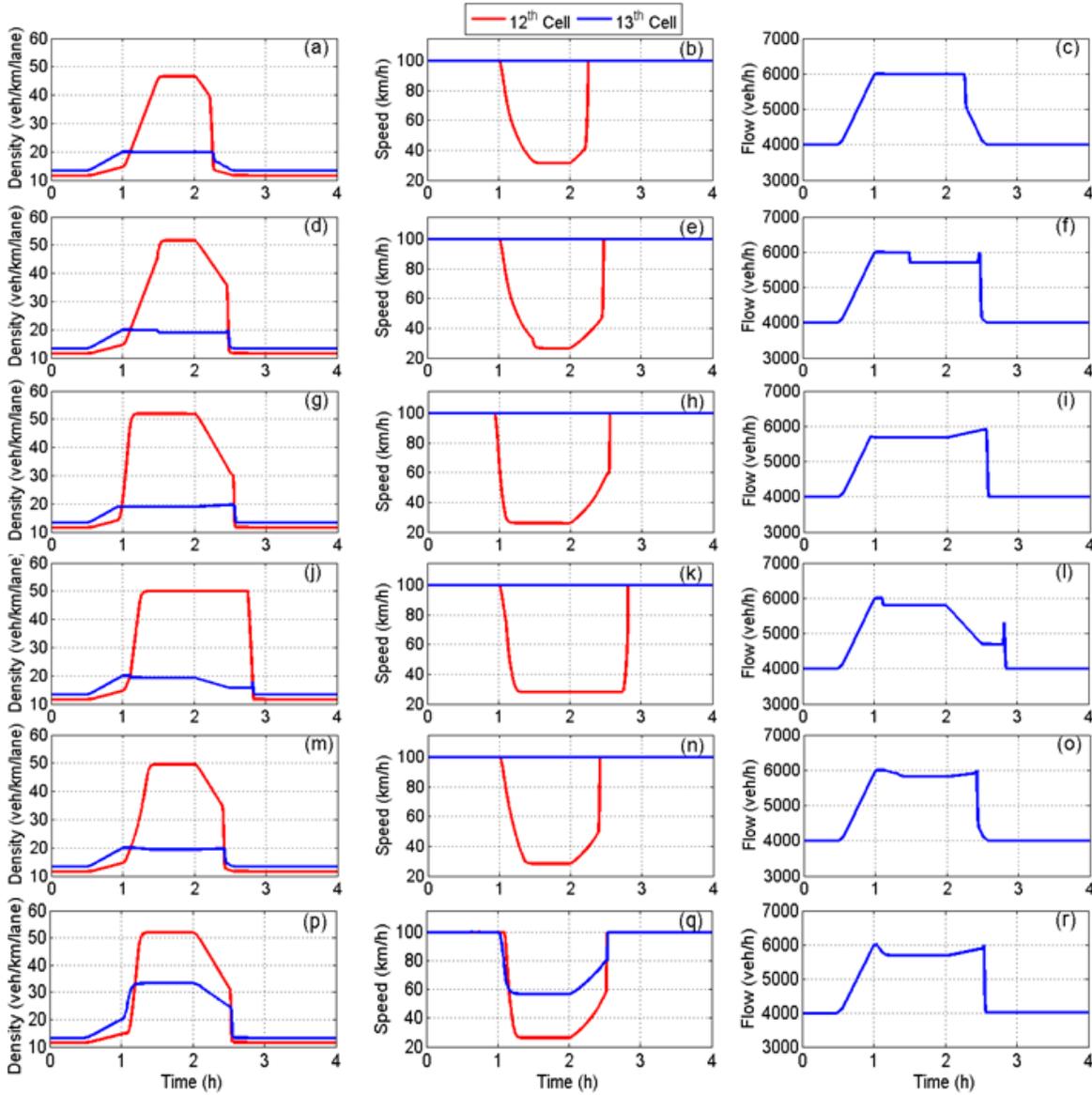


Fig. 4. The time-series of the density of the 12th and 13th cell, the speed of the 12th and 13th cell and the outflow from the 13th cell for the application of (a),(b),(c) CTM, (d),(e),(f) Approach 1, (g),(h),(i) Approach 2, (j),(k),(l) Approach 3, (m),(n),(o) Approach 4, (p),(q),(r) Approach 5.

This behavior can be alternatively obtained with different formulations (instead of using the parameter θ_r). For instance, such an effect can be also obtained by considering an increased upper bound (capacity) in (4) for undercritical densities, e.g. increasing capacity by 5% and increasing accordingly the resulting from the common FD wave speed (e.g. increased similarly by 5%) (Kontorinaki et al., 2016.). Again, a decreasing demand function for over-critical densities (14) has to be considered so as to achieve the desired capacity drop. This alternative modeling approach acts in the same way as parameter θ_r does (delay the activation of the supply function), but it may also be applied for bottlenecks due to lane drops, tunnels etc., and not only for on-ramp merges.

This new modeling approach has also been examined regarding its relation and consistency with the Partial Differential Equation (PDE) of the LWR model (see Appendix). Among others, this analysis enables to assess how the introduced parameters of Approach 5 affect the solution of the discretized model when the discretization parameters (cell length and simulation time step) tend to zero. However, in order to

conduct such an analysis, it is important to determine the two ways the on-ramp flow term may be treated within the discretization: first by assuming that the on-ramp flow term is distributed flow within specific space bounds (Section A.1 of the Appendix); second by assuming the on-ramp flow is a concentrated (Dirac function) flow at a given space point of the freeway (Section A.2 of the Appendix). From the analysis it follows that, when treating the on-ramp via the first way, the proposed discretized model is consistent with the LWR PDE, while simulation results reveal that in some cases the solution of the discretized model converges (as the discretization parameters tend to zero) to the solution of the CTM (which in turn converges to the LWR PDE solution). However, following the second way for treating the on-ramp flow, one can only define the integral form of the PDE at the point where the on-ramp is implemented (due to the apparent singularity introduced by the Dirac function). The analysis is then performed by means of the resulting shock speed of the proposed model (following from the Rankine-Hugoniot condition) and by comparing it with the shock speed of the CTM following the methodology described by LeVeque (2002). In this case, the analysis indicates that the solution of the discretized model converges (as the discretization parameters tend to zero) to a different solution with a different shock speed from the one imposed by CTM. However, the model is still consistent with the PDE of LWR everywhere else except the point where the on-ramp is implemented. Therefore, it can be concluded that Approach 5 can be reduced to a kinematic wave model endowed with an appropriate FD (Zhang, 2001). The reader should be transferred to the Appendix for a more detailed justification of the above analysis.

3.9. First-order model with drivers' anticipation

Other first-order models, which are not of LWR-type and do not utilize the demand-supply method, have been proposed in the past. For comparison purposes in the calibration procedure of the next section, an alternative first-order model, which was first proposed by Lighthill and Whitham (1955) in order to represent the diffusion of kinematic waves, is introduced. The following discretized formulation is a variation of Model E presented and tested by Papageorgiou et al. (1989), with the difference that the model is formulated here in terms of flow, instead of speed. The complete model is described by (1), (15), (16) and (17):

$$f_i(\rho_i(k), \rho_{i+1}(k), \rho_{i+2}(k)) = \beta(1 - p_i)q_i(\rho_i(k), \rho_{i+1}(k)) + (1 - \beta)(q_{i+1}(\rho_{i+1}(k), \rho_{i+2}(k)) - r_{i+1}(k))$$

for $i = 1, \dots, n - 2$ (15)

$$f_{n-1}(\rho_{n-1}(k), \rho_n(k)) = (1 - p_{n-1})q_{n-1}(\rho_{n-1}(k), \rho_n(k)), f_n(\rho_n(k)) = Q_n^e(\rho_n(k))$$

$$q_i(\rho_i(k), \rho_{i+1}(k)) = Q_i^e(\rho_i(k)) - \frac{\tilde{v}_i}{L_i}(\rho_{i+1}(k) - \rho_i(k)) \text{ for } i = 1, \dots, n-1, q_n(\rho_n(k)) = Q_n^e(\rho_n(k)) \quad (16)$$

where $0 < \beta \leq 1$ and \tilde{v} are model parameters, q_i represents the total outflow from the i^{th} cell and Q_i^e represents the FD of the i^{th} cell. This model includes an anticipation term that influences the total outflow from a cell according to the downstream prevailing conditions. This mechanism, included in (16), suggests that drivers adjust their speed, by also taking into account the downstream density. Moreover, in the space-discretised version, the traffic volume from a cell to another is a convex combination of the total traffic volume of the current and the next cells (15).

Thanks to the anticipation term, this model is capable to reproduce the capacity drop phenomenon. The utilized mathematical formula for the FD for the calibration test below is an exponential concave function of density attaining its maximum at the critical density (see (17)). However, it should be noted that in case, for some specific reasons (e.g., the formulation of an optimization problem), linear constraints are needed, this formula may be replaced by a piecewise-linear concave function.

Finally, it is worth highlighting that all the presented approaches have the same computational complexity, virtually equivalent to the basic implementation of CTM. On the other hand, Approach 1 requires a higher amount of memory than the rest of the approaches due to the need to store additional auxiliary variables R_i , which is similar to the case of second-order models.

4. Calibration results

In this section the approaches described in Section 3 are validated and compared regarding the accuracy of reproducing traffic conditions in a real motorway stretch with particular emphasis on the reproduction of the capacity drop phenomenon.

4.1. Motorway network and calibration set-up

The considered motorway stretch of 9.5 km in length is part of the M56 motorway in the United Kingdom, direction from Chester to Manchester. This three-lane motorway stretch includes an off-ramp and a two-lane on-ramp, which, before entering the motorway, is divided into two separate lanes. The corresponding on-ramp flows of each lane enter the motorway at two different locations, as shown in Fig. 5. Fig. 5 displays the locations of the on-ramps and off-ramp and the locations of the available detector stations. In order to apply the selected traffic flow models, the examined motorway stretch is divided into 38 model cells of about 250 m each, as shown in Fig. 5. Using this representation, the motorway cells are well-defined, and the model equations presented in the previous section are directly applicable.

The real traffic data used in this study were obtained from the MIDAS database (Highways Agency, 2007). The traffic data include flow and speed measurements at each detector location, with a time resolution of 60 s. The traffic data analysis showed that, within this motorway stretch, recurrent congestion is created during the morning peak hours due to the high on-ramp flow. In particular, Fig. 6(a) displays the space-time diagram of the real speed measurements for 03/06/2014. It is observed that congestion is created upstream of the second on-ramp during 7–8 a.m. which spills back several kilometers. Moreover, downstream of the second on-ramp, there is an area characterized by low speed, due to the acceleration of vehicles exiting the congestion area. Fig. 7 presents the time-series of the flow measurements (black line) from detector station D 8180 which is located downstream of the congestion creation area (see, also, Fig. 5). It is observed that a capacity drop is present there, as the merge area outflow drops visibly when congestion sets in (between 7:10 a.m. and 8:10 a.m.).

In order to apply the examined models to this motorway stretch and achieve a fair comparison, it is important to first calibrate the models using the available real traffic data. The model calibration procedure aims to specify the model parameter values, so that the representation of the network traffic conditions is as accurate as the model structure allows. This can be achieved by employing a suitable optimization methodology which aims at minimizing the discrepancy between the model estimations and the real traffic data. More details on the utilized model calibration procedure are provided by Spiliopoulou et al. (2014).

In the current study, the Nelder-Mead optimization method is employed for the calibration of the examined traffic flow models. The models are fed with boundary data (inflows at the upstream boundary

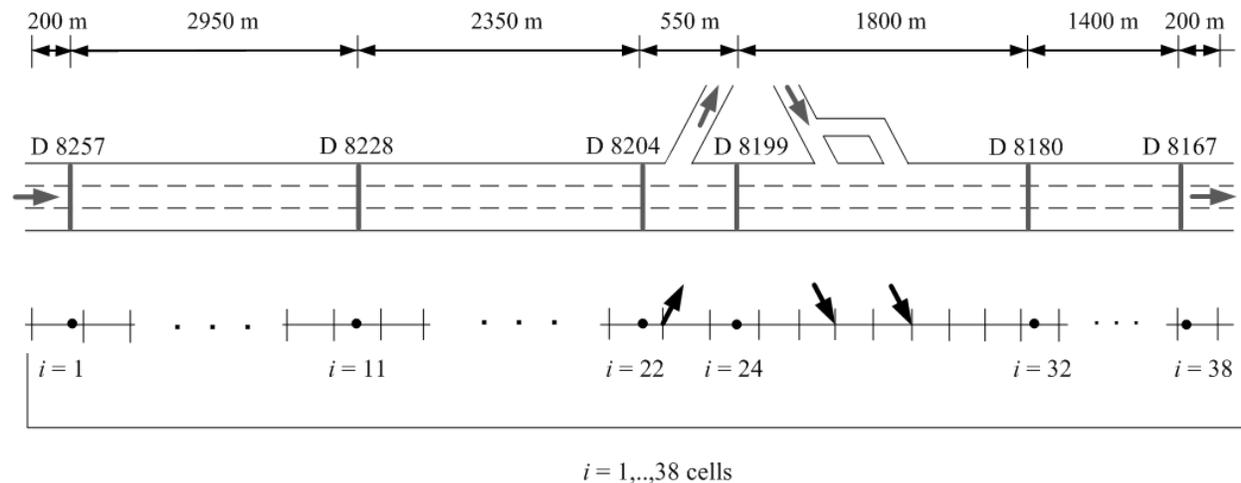


Fig. 5. Representation of the considered freeway stretch.

and the on-ramps and exit rates at the off-ramp) and produce the stretch-internal traffic states according to the respective equations and parameter values. The utilized performance index (PI) under minimization is the Root-Mean-Square Error (RMSE) of the real versus the model-predicted speed values at all detector locations. The models are calibrated using real traffic data from 03/06/2014 and a simulation time step equal to $T = 5$ s. It should be stressed that all cells of the modeled motorway stretch are characterized by the same parameters of the FD for each model. After the calibration procedure, the accuracy and robustness of the resulted models is evaluated by validating the produced models with different traffic data (from the same motorway site) than the data used for their calibration. In this study, the models are validated using real traffic data from 19/06/2014.

4.2. Basic discretized-LWR formulation

As mentioned before, the investigated capacity drop approaches are based on the discretized LWR model. This basic first-order model cannot reflect the capacity drop phenomenon; however, different shapes of the FD may improve the model's accuracy. To investigate this, four different shapes of the FD are examined, all applied to the basic discretized LWR model, i.e. triangular FD, trapezoidal FD, piecewise linear FD, and nonlinear exponential FD (see, Fig. 2).

Table 3 includes the calibrated model parameter values and Table 4 the corresponding PI values for the calibration and the validation datasets. It is interesting to see that all four variations (Triang. FD, Trapez. FD, PWL FD, NL FD) produced a similar value for the Q (capacity) parameter. Moreover, as it was expected, the use of a triangular FD results in a low ρ_{cr} value, lower than in the other formulations. Fig. 6(a)-(e) displays the space-time diagrams of the real speed measurements and the corresponding models' predictions of speed for the calibration date. It is observed that the models using a triangular or a trapezoidal FD predict free flow conditions at all areas outside congestion. In contrast, the use of a piecewise linear or non-linear FD allows for mean speed variations also outside of the congestion area, thus achieving higher accuracy at lower densities, compared to the first two formulations. Considering the above results, first-order LWR-type models with nonlinear FD are used in the subsequent investigations of capacity drop approaches, i.e., the function g , used in the demand function, is an exponential increasing function. More specifically, function g is defined as

$$g(\rho_i) = v_{f,i} \rho_i l_i \exp \left(-\frac{1}{c_i} \left(\frac{\rho_i}{\rho_{cr,i}} \right)^{c_i} \right), \quad (17)$$

where $c_i = -1/\ln(Q_i/(l_i v_{f,i} \rho_{cr,i}))$.

4.3. Capacity drop approaches

Five capacity drop approaches, which were described in Section 3, are implemented for this simple, but typical, motorway stretch. Table 3 includes the estimated model parameter values for all five approaches. It should be mentioned that in all examined approaches the maximum capacity flow Q was fixed at 6900 veh/h, which is close to the highest flows observed in the network. This was done in order to achieve a fair comparison of the models regarding the reproduction of the capacity drop phenomenon.

Table 3 shows that in all five approaches similar values were estimated for the v_f and ρ_{cr} parameters, while quite different values were obtained for the parameters w and ρ_{max} due to the different formulations adopted for the reproduction of the capacity drop phenomenon. Moreover, it should be noted that, although in all approaches the parameter α is related to the magnitude of the capacity drop, the impact on the resulting capacity drop is substantially different, and for this reason the value of α varies in the different approaches. Table 4 presents the PI values for the calibration and the validation datasets. It is observed that the models achieve similar PI values, which implies that they are all able to reproduce the traffic conditions in this network with reasonable and comparable accuracy. More specifically, all approaches, except for Approach 3, improve the PI value compared to the basic LWR model with NL FD.

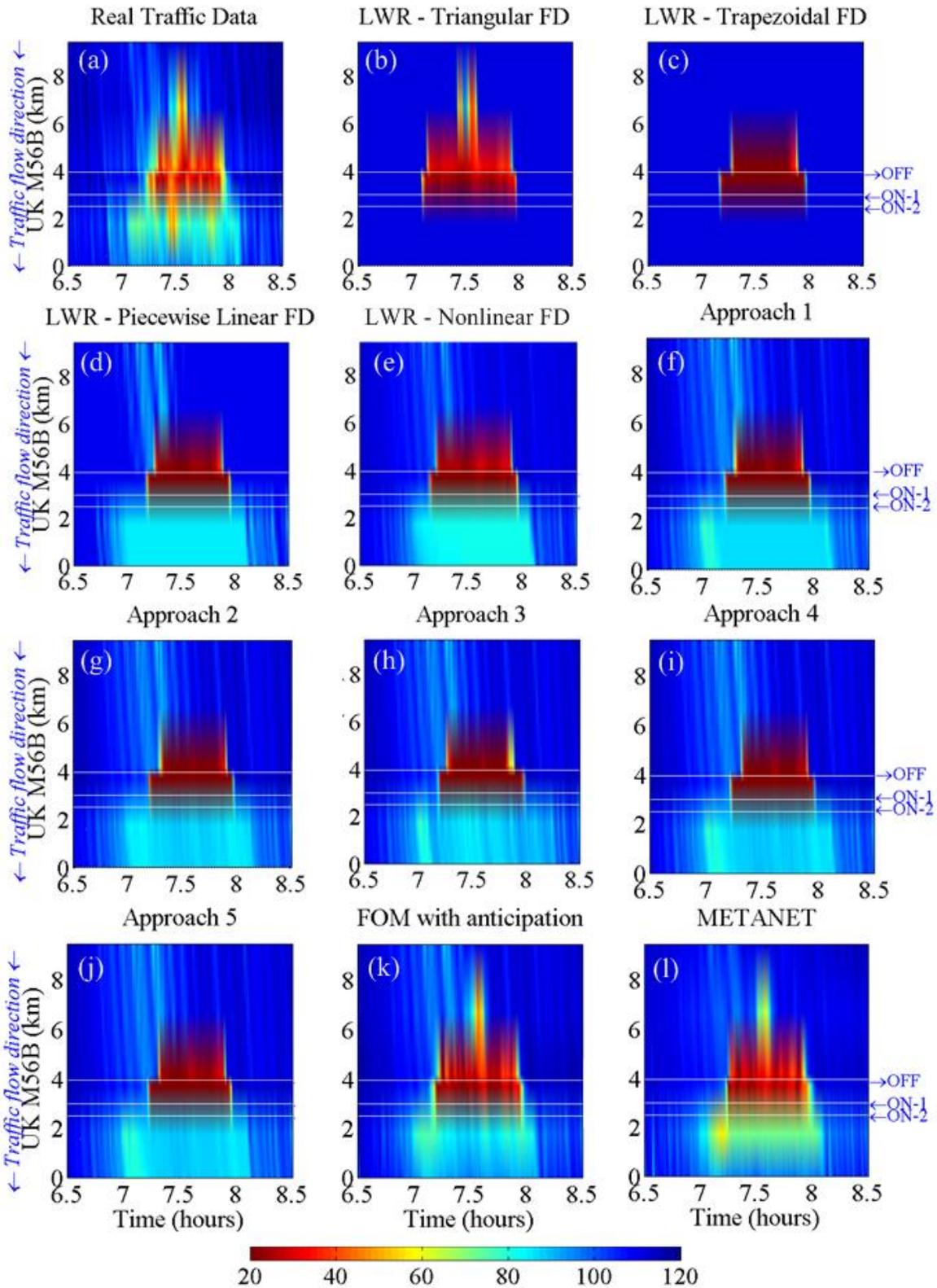


Fig. 6. Space-time diagrams of speed for (a) the real traffic data; (b) the LWR model with triangular FD; (c) the LWR model with trapezoidal FD; (d) the LWR model with piecewise linear FD; and (e) the LWR model with nonlinear FD; (f) Approach 1; (g) Approach 2; (h) Approach 3; (i) Approach 4; (j) Approach 5; (k) FOM with anticipation; and (l) METANET model for 03/06/2014.

Table 3
Calibrated parameter values for all examined models.

<i>Model</i>	v_f (km/h)	ρ_{cr} (veh/km/lane)	w (km/h)	ρ_{max} (veh/km/lane)	Q (veh/h)	ρ_a (veh/km/lane)	α	η_r	θ_r
Triang. FD	.0	18.7	21.2	117.4	6282	-	-	-	-
Trapez. FD	112.0	-	21.8	145.0	6192	-	-	-	-
PWL FD	110.5	24.7	14.8	165.5	6258	4	-	-	-
NL FD	113.5	25.3	12.2	195.4	6225	-	-	-	-
Approach 1	123.2	36.4	25.9	124.9	6900	-	0.89	-	-
Approach 2	123.4	35.9	21.9	139.2	6900	-	-	1.56	-
Approach 3	122.8	33.5	21.4	149.1	6900	-	0.72	-	-
Approach 4	23.0	5.6	25.9	131.2	6900	-	-	-	-
Approach 5	8	5.6	33.4	5	6900	-	34	-	0.71
FOM with anticipation	119.7	29.4	-	-	6402	-	-	-	-
METANET	114.2	28.9	-	-	6525	-	-	-	-

Table 4
Performance indices for the calibration and validation days for all examined models.

<i>Model</i>	<i>03/06/2014</i>		<i>19/06/2014</i>	
	<i>Speed Error (PI)</i> (km/h)	<i>Flow Error</i> (veh/h)	<i>Speed Error (PI)</i> (km/h)	<i>Flow Error</i> (veh/h)
Triang. FD	18.0	625.7	23.4	477.8
Trapez. FD	18.0	651.8	24.2	473.9
PWL FD	12.6	626.6	19.0	479.9
NL FD	12.6	653.4	19.0	487.8
Approach 1	1.6	603.0	19.0	425.8
Approach 2	11.9	606.7	18.9	485.8
Approach 3	12.8	685.1	18.9	439.3
Approach 4	11.3	584.3	18.6	425.9
Approach 5	10.9	576.0	18.4	431.4
FOM with anticipation	9.9	598.6	17.2	446.2
METANET	7.9	471.2	14.8	420.6

Table 3 and Table 4 also include the calibration results for the first-order model with anticipation (presented in Section 3.9) and the second-order model METANET (Messmer and Papageorgiou, 1990), which are applied to this motorway stretch for comparison purposes. Note that Table 3 presents only some of the

estimated parameters of these two models while the rest parameters were estimated equal to: $\tilde{v} = 6.4 \text{ km}^2/\text{h}$ and $\beta = 0.76$ for the first-order model with anticipation and $\tau = 26.8 \text{ s}$, $v = 45.6 \text{ km}^2/\text{h}$, $\delta = 0.1 \text{ h/km}$, $\kappa = 10 \text{ veh/km/lane}$, $v_{min} = 7 \text{ km/h}$ for METANET model.

Fig. 6(f)-(j) presents the space-time diagrams of the corresponding speed estimations for all five capacity-drop approaches for the calibration date. It is observed that the estimations of all five approaches are close to the real speed data and are actually similar to each other. Fig. 6(k)-(l) displays the corresponding speed estimations of the first-order model with anticipation and the second-order model METANET. It is shown here that the first-order model with anticipation achieves a remarkably high accuracy in representing the prevailing traffic conditions thanks to the included anticipation term (Eq. (16)) On the other hand, METANET model produces, as expected, the most realistic representation of the traffic characteristics, thanks to the fact that it accounts also the vehicle acceleration capabilities and the driver reaction time.

Regarding the reproduction of the capacity drop phenomenon, Fig. 7 displays the time-series of the real flow measurements and the corresponding models' estimations at the location of detector station D 8180 (see, Fig. 5), which is placed about 800 m downstream of the merge area. It is observed that, except for the basic LWR model with nonlinear FD, all five approaches produce a reduced merge area outflow during the congestion period. Moreover, all approaches, except for Approach 2, are capable to estimate a high merge area outflow (close to capacity) just before the onset of congestion, which is in accordance with the real flow values observed. Fig. 7 (g) shows that the first-order model with anticipation also produces a

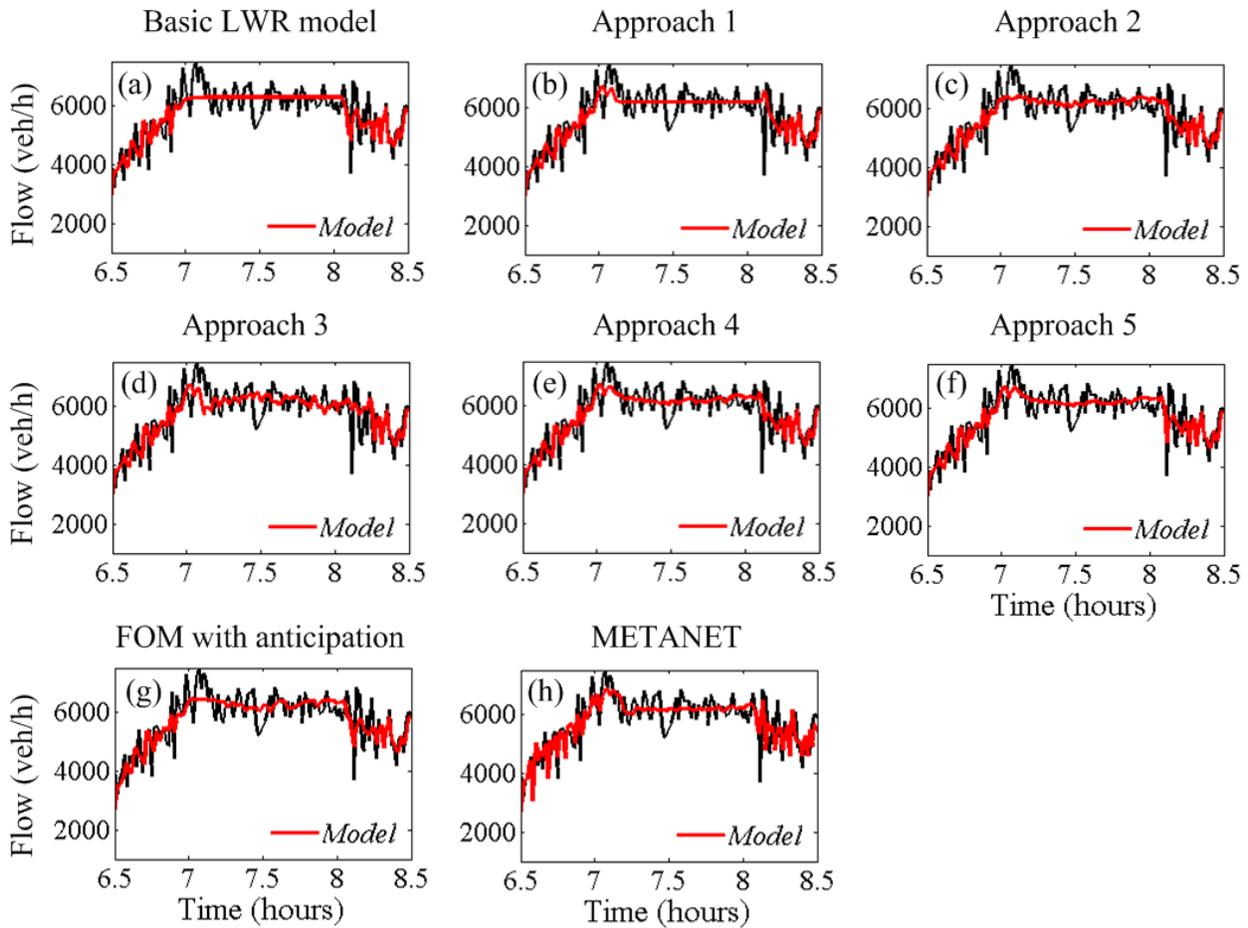


Fig. 7. Time-series of the real flow measurements at the location of detector station D 8180 and the corresponding flow estimations of (a) the basic LWR model; (b) Approach 1; (c) Approach 2; (d) Approach 3; (e) Approach 4; (f) Approach 5; (g) FOM with anticipation; and (h) METANET model; for 03/06/2014.

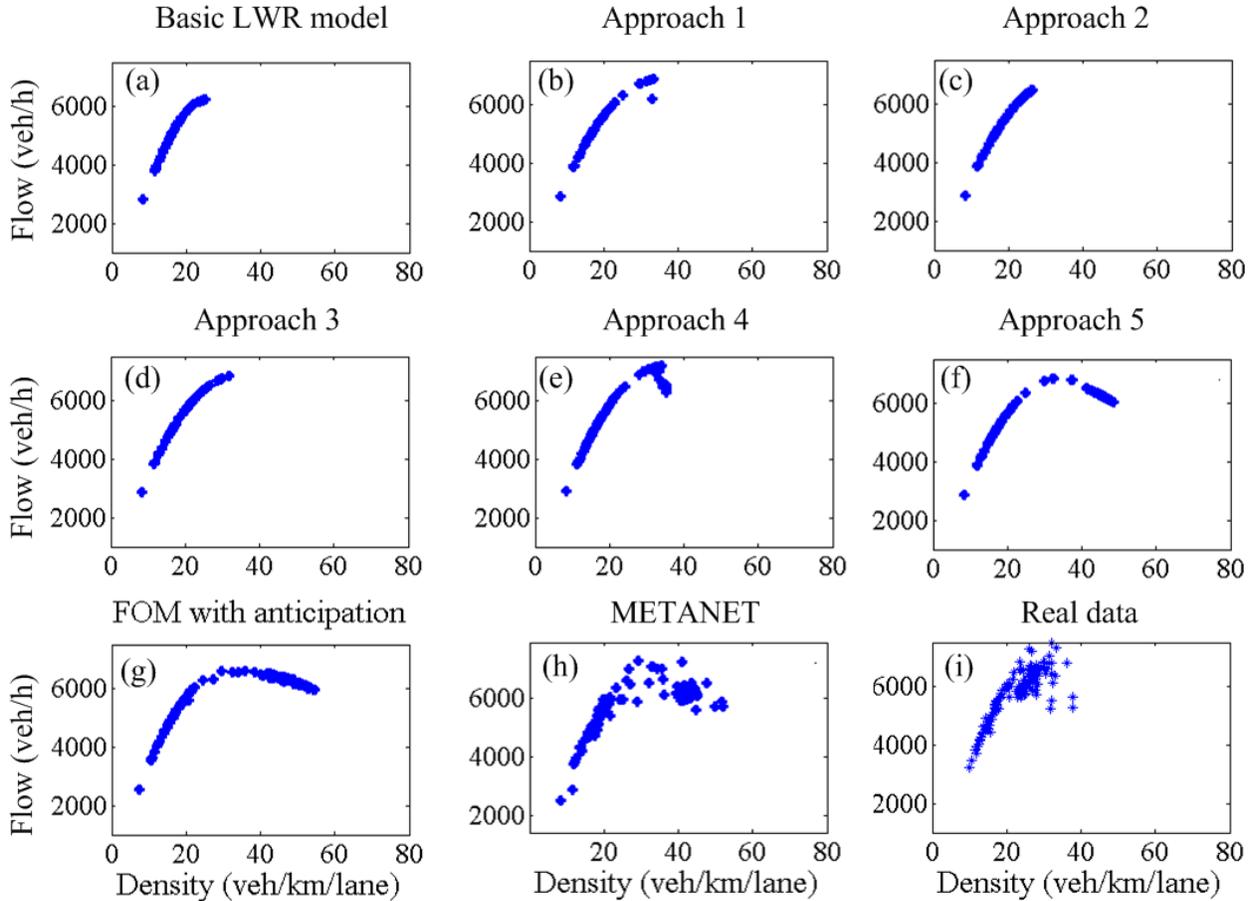


Fig. 8. Flow-density diagram at the merge cell using (a) the basic LWR model; (b) Approach 1; (c) Approach 2; (d) Approach 3; (e) Approach 4; (f) Approach 5; (g) FOM with anticipation; (h) METANET model; and (i) real data at detector D 8180, for 03/06/2014.

reduced merge area outflow, however with a smaller flow drop than the real observed flow drop. Finally, Fig. 7(h) presents the flow estimations of the second-order model METANET which are very close to the real traffic measurements.

In order to evaluate quantitatively the accuracy of the five approaches in reproducing the capacity drop phenomenon, the RMSE of the real flow measurements and the corresponding model estimations of flow is calculated, for the motorway cell where the detector station D 8180 is located and for a time window around the time when capacity drop appears (i.e. between 7–7:15 a.m.). Table 4 includes the corresponding flow error for all examined models for the calibration and the validation datasets. It is observed that all approaches, except for Approach 3, for the calibration day, and Approach 2, for the validation day, achieve a lower error compared to the basic LWR formulation without capacity drop. Furthermore, the first-order model with anticipation, although reproducing correctly the propagation of congestion (as it can be also deduced from the low PI value), seems less capable to create a satisfactory capacity drop. As Table 4 shows, the flow error in Approach 4 and Approach 5 is noticeably smaller than the flow error of the first-order model with anticipation.

Finally, it is noted that the second-order model METANET achieves the highest accuracy in reproducing the capacity drop compared to all employed first-order models. This confirms previous consistent findings on the higher accuracy of second-order models compared to first-order ones. Specifically, Cremer and Papageorgiou (1981) used traffic data from a German Autobahn and found that, after calibration, a discretised form of the LWR model had a 226% higher standard deviation of the

modelling error compared to a second-order model; similarly, Papageorgiou et al. (1989) used data from the Boulevard Périphérique in Paris and found the standard deviation of the mean speed error to be 17.5 km/h for the discretised LWR model versus 8.3 km/h for a second-order model; it should be noted that these works were conducted prior to the introduction of the Godunov scheme for LWR model discretization (Lebacque, 1996) and used a simple discretization equation for calculating the flow between cells. Michalopoulos et al. (1992) used a Lax discretisation of the LWR model and concluded that “When there is downstream congestion, all high-order models performed substantially more accurately than the simple continuum model”. Spiliopoulou et al. (2014) compared CTM and the METANET second-order model used here while modelling a congested off-ramp area on a Greek motorway; Frejo (2015) compared the same models using data from freeway I-210 West in Southern California; both studies reported a better match of METANET to real data. Finally, Fan and Seibold (2013) concluded that the second-order ARZ model (Aw and Rascle, 2000; Zhang, 2002) is significantly more accurate than the LWR model. Clearly, this increased accuracy of second-order models comes at the expense of higher computational effort and complexity, as discussed in Section 1. Therefore, the development of LWR-type models with more realistic behaviour is a valid research endeavour.

Fig. 8 displays the flow versus density diagram (i.e., the FD) at the merge cell (cell 29th, Fig. 5) for the basic LWR model (again, using a nonlinear demand part of the FD) and all five approaches. It is observed that, as expected, the basic LWR model is not able to reproduce the capacity drop phenomenon (see Fig. 7(a) and Fig. 8(a)). Comparing the five examined approaches, it is observed that actually only Approach 1, Approach 4 and Approach 5 produce a capacity drop at the merge cell, resulting though in different FD shapes (due to their different formulations).

In particular, in Approach 1 the merge area discharge flow corresponds exactly to the pre-specified value $\bar{Q} = aQ = 6141$ veh/h (see Fig. 8(b)). In Approach 4 (see Fig. 8(e)), the magnitude of the observed capacity drop varies according to the density of the upstream cell; i.e. in case of stronger congestion, characterized by a lower speed, a stronger capacity drop is observed, which is in accordance with some traffic observations. Finally, in Approach 5, the capacity drop observed follows the shape of the demand function of the merge cell while the magnitude of the observed capacity drop also depends on the magnitude of the applied on-ramp volume (see Fig. 8(f)). On the other hand, Approach 2 and Approach 3 do not produce a capacity drop at the merge cell (see Fig. 8(c) and Fig. 8(d)) although they manage to produce a reduced outflow from the merge area during congestion period (see Fig. 7(c) and Fig. 7(d)). In particular, in Approach 2 the observed merge area outflow never reaches capacity, even before the onset of congestion, in accordance with the behavior described in Section 3. Moreover, regarding Approach 3, the discharge flow that materializes is also dependent on the on-ramp flow entering the merge cell (which causes the fluctuations that can be observed in the corresponding plot), whereas the mainstream flow exiting the cell upstream of the merge cell, during the congestion period, is constantly equal to $\bar{Q} = aQ = 4968$ veh/h. Fig. 8 also includes, for comparison, the flow versus density diagram at the merge cell for the first-order model with anticipation (Fig. 8(g)) and the second-order model METANET (Fig. 8(h)); as well as the corresponding real-data diagram at the location of the detector station D 8180, which is about 800 m downstream of the merge area (Fig. 8(i)). As a result, Fig. 8(i) cannot be directly compared to the other plots, but it is included here as it corresponds to the closest measurement point downstream of the bottleneck location. Notice that the density in Fig. 8(i) has been estimated from flow and speed measurements by $\rho(t) = q(t)/v(t)$.

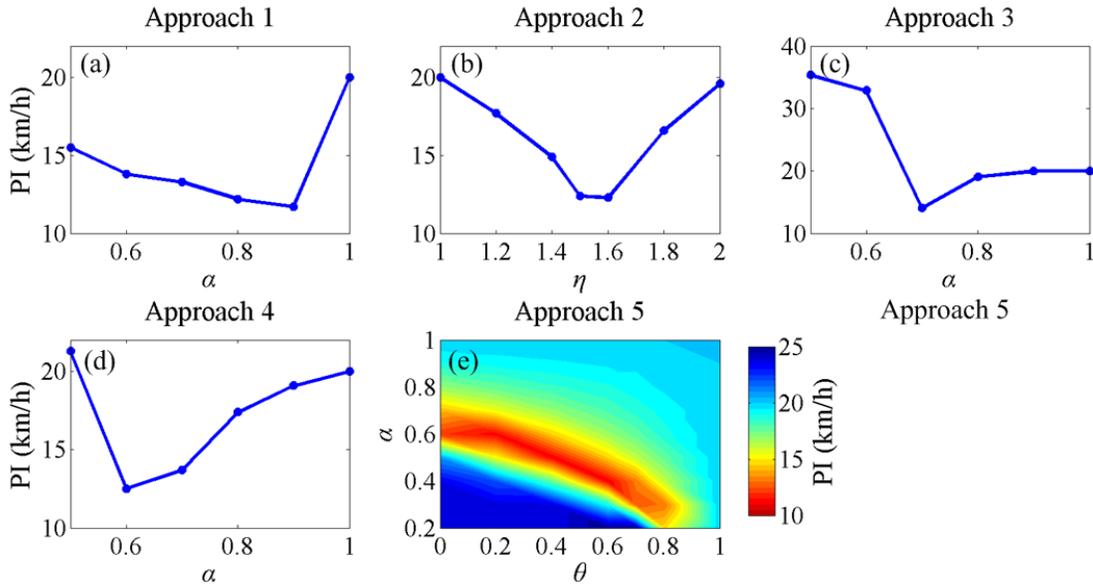


Fig. 9. Sensitivity investigations in changes to the models' parameters related to the capacity drop, in terms of PI value for (a) Approach 1; (b) Approach 2; (c) Approach 3; (d) Approach 4; and (e) Approach 5 for 03/06/2014.

While examining the presented five approaches, a question was raised regarding the sensitivity of the models to variations of the parameters α or η_r and/or θ_r (depending on the approach) which are involved in the reproduction of the capacity drop phenomenon. To investigate this issue, different values were fixed for these parameters and the models were calibrated again (with respect to the rest parameters) for each examined value. Fig. 9 displays all the related results. As an example, Fig. 9(a) presents the best obtained PI values after calibrating Approach 1 for different fixed values of the parameter *alpha* within the range [0.5, 1]. It is observed that the model achieves lowest PI values for α close to 0.9. Similarly, in Approach 2 the PI is minimized for η_r in the range [1.4, 1.5] (Fig. 9(b)), in Approach 3 for α close to 0.7 (Fig. 9(c)) and in Approach 4 for α close to 0.6 (Fig. 9(d)). Regarding Approach 5, which includes two parameters related to the capacity drop, α and θ_r , the investigations include different coupled values of these two parameters and the model was calibrated again for each one of these couples. Fig. 9(e) presents the best obtained PI value for each investigated couple. It may be seen that different coupled values of the two parameters lead to equally low PI values. This means that these two parameters are strongly correlated. Considering the above investigations, it is concluded that the models are sensitive to the value of the parameters related to the capacity drop. Finally, note that for $\alpha = 1$, in Approach 1, 3 and 4, and for $\eta_r = 1$ and $\theta_r = 1$, in Approach 2 and Approach 5, respectively, the basic LWR model is obtained.

5. Conclusions

This study presents an overview of modeling approaches to include capacity drop into LWR traffic flow models. The presented approaches are first described and tested for a hypothetical network and traffic demand scenario to highlight their principal behavior and qualitative properties; eventually, the models were rigorously calibrated and validated using real data from a motorway in the U.K. The obtained results show that, although the tested models employ different mechanisms to reflect the capacity drop phenomenon, they all manage to produce a flow reduction at the merge area whenever traffic congestion is present. The obtained results were found to be quantitatively similar with respect to the achieved PI values, which is mainly attributed to a traffic situation with limited complexity and few internal comparison data. Moreover, some of the approaches that incorporate the capacity drop into LWR-type models produced a lower flow error (better reproduction of capacity drop) than the one obtained by the FOM with anticipation.

Regarding the proposed model (Approach 5), the performed theoretical analysis (presented in the Appendix) indicates that the formal underlying PDE is the LWR model, but with a source term present (i.e. a non-homogeneous PDE). More specifically, by considering the on-ramp flow as a distributed flow within specific bounds of the freeway stretch, we have shown that the discretization scheme is consistent with the PDE of LWR. This implies that, as the discretization grid becomes denser, the impact of the parameters of Approach 5 (α and θ_r) tends to fade out. On the other hand, by treating the on-ramp flow as a singular source, i.e., a Dirac function, at a specific point of the freeway stretch, we have shown that the solution of the discretized model converges (as the discretization parameters tend to zero) to a different solution with a different shock speed from the one imposed by CTM. However, in this case, the model is still consistent with the PDE of LWR everywhere else except the point where the on-ramp is implemented.

Future investigations involving more complex traffic situations and richer data might shed more light on the comparative quantitative accuracy of different approaches. Moreover, it would be interesting to test and evaluate the behavior of the described models in case traffic control strategies are applied.

Acknowledgements

The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement n. [321132], project TRAMAN21. We are grateful to Highways England for providing access to the MIDAS data. We thank our colleagues Professors A.I. Delis, I. Karafyllis and I.K. Nikolos for their contribution to the analysis of the Appendix.

References

- Alecsandru, C., Quddus, A., Huang, K.C., Rouhieh, B., Khan, A.R., Zeng, Q., 2011. An assessment of the cell-transmission traffic flow paradigm: development and applications. Presented at the 90th Annual Meeting of Transportation Research Board, p. no.11-1152.
- Alvarez-Icaza, L., Islas, G.J., 2013. Hysteretic cell transmission model. Presented at the 16th International IEEE Conference on Intelligent Transportation Systems - (ITSC), pp. 578–583. doi:10.1109/ITSC.2013.6728293
- Aw, A., Rascle, M., 2000. Resurrection of “second order” models of traffic flow. *SIAM Journal on Applied Mathematics* 60, 916–938. doi:10.1137/S0036139997332099
- Banks, J.H., 1991. The two-capacity phenomenon: Some theoretical issues. *Transportation Research Record* 234–241.
- Cassidy, M., Ahn, S., 2005. Driver turn-taking behavior in congested freeway merges. *Transportation Research Record* 1934, 140–147. doi:10.3141/1934-15
- Cassidy, M.J., 1998. Bivariate relations in nearly stationary highway traffic. *Transportation Research Part B: Methodological* 32, 49–59. doi:10.1016/S0191-2615(97)00012-X
- Cassidy, M.J., Rudjanakanoknad, J., 2005. Increasing the capacity of an isolated merge by metering its on-ramp. *Transportation Research Part B: Methodological* 39, 896–913. doi:10.1016/j.trb.2004.12.001
- Chung, K., Rudjanakanoknad, J., Cassidy, M.J., 2007. Relation between traffic density and capacity drop at three freeway bottlenecks. *Transportation Research Part B: Methodological* 41, 82–95. doi:10.1016/j.trb.2006.02.011
- Cremer, M., Papageorgiou, M., 1981. Parameter identification for a traffic flow model. *Automatica* 17, pp. 837–843. doi: [10.1016/0005-1098\(81\)90071-6](https://doi.org/10.1016/0005-1098(81)90071-6)
- Daganzo, C.F., 1995a. Requiem for second-order fluid approximations of traffic flow. *Transportation Research Part B: Methodological* 29, 277–286. doi:10.1016/0191-2615(95)00007-Z
- Daganzo, C.F., 1995b. The cell transmission model, part II: Network traffic. *Transportation Research Part B: Methodological* 29, 79–93.
- 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, 269–287. doi:10.1016/0191-2615(94)90002-7
- Delis, A.I., Nikolos, I.K., Papageorgiou, M., 2014. High-resolution numerical relaxation approximations to second-order macroscopic traffic flow models. *Transportation Research Part C: Emerging Technologies* 44, 318–349. doi:10.1016/j.trc.2014.04.004
- Dixon, K., Hummer, J., Lorscheider, A., 1996. Capacity for north Carolina freeway work zones. *Transportation Research Record* 1529, 27–34. doi:10.3141/1529-04
- Edie, L.C., 1961. Car-following and steady-state theory for noncongested traffic. *Operations Research* 9, 66–76. doi:10.1287/opre.9.1.66
- Fan, S., Seibold, B., 2012. A comparison of data-fitted first order traffic models and their second order generalizations via trajectory and sensor data. arXiv:1208.0382[physics.soc-ph].
- Fan, S., Seibold, B., 2013. Data-fitted first-order traffic models and their second-order generalizations: comparison by trajectory and sensor data. *Transportation Research Record: Journal of the Transportation Research Board*, 2391, pp. 32-43. doi: [10.3141/2391-04](https://doi.org/10.3141/2391-04)
- Ferrara, A., Sacone, S., Siri, S., 2015. Event-triggered model predictive schemes for freeway traffic control. *Transportation Research Part C: Emerging Technologies, Special Issue: Advanced Road Traffic Control* 58, 554–567. doi:10.1016/j.trc.2015.01.020
- Frejo, J. R. D., 2015. Model predictive control for freeway traffic networks. PhD Thesis, Universidad de Sevilla, Spain.
- Godunov, S.K., 1959. A difference method for numerical calculation of discontinuous solutions of hydrodynamic equations. *Matematicheskii Sbornik* 47, 271–306.
- Gomes, G., Horowitz, R., 2006. Optimal freeway ramp metering using the asymmetric cell transmission model. *Transportation Research Part C: Emerging Technologies* 14, 244–262. doi:10.1016/j.trc.2006.08.001
- Hall, F.L., Agyemang-Duah, K., 1991. Freeway capacity drop and the definition of capacity. *Transportation Research Record* 91–98.
- Han, Y., Hegyi, A., Yuan, Y., Hoogendoorn, S., Papageorgiou, M., Roncoli, C., 2017. Resolving freeway jam waves by discrete first-order model-based predictive control of variable speed limits. *Transportation Research Part C: Emerging Technologies* 77, 405-420.
- Han, Y., Yuan, Y., Hegyi, A., Hoogendoorn, S., 2016. A new extension of discrete first-order model to reproduce the propagation of jam waves. Presented at the 95th Annual Meeting of Transportation Research Board, p. no.16-3482.
- Helbing, D., Johansson, A.F., 2009. On the controversy around Daganzo’s requiem for and Aw-Rasche’s resurrection of second-order traffic flow models. *The European Physical Journal B* 69, 549–562. doi:10.1140/epjb/e2009-00182-7
- Highways Agency, 2007. *Motorway Incident Detection and Automatic Signalling (MIDAS) Design Standard*. (No. 1st ed.). Bristol, UK.
- Jin, L., Amin, S., 2017. Calibration of a macroscopic traffic flow model with stochastic saturation rates. Presented at the 96th Annual Meeting of Transportation Research Board, p. no. 17-02496.
- Jin, W.-L., Gan, Q.-J., Lebacque, J.-P., 2015a. A kinematic wave theory of capacity drop. *Transportation Research Part B: Methodological* 81, 316–329. doi:10.1016/j.trb.2015.07.020
- Jin, W. L. 2015b. Continuous formulations and analytical properties of the link transmission model. *Transportation Research Part B: Methodological* 74, pp. 88-103. doi: [10.1016/j.trb.2014.12.006](https://doi.org/10.1016/j.trb.2014.12.006)
- Jin, W.-L., 2010. A kinematic wave theory of lane-changing traffic flow. *Transportation Research Part B: Methodological* 44, 1001–1021. doi:10.1016/j.trb.2009.12.014
- Jin, W.-L., Chen, L., Puckett, E.G., 2009. Supply-demand Diagrams and a New Framework for Analyzing the Inhomogeneous Lighthill-Whitham-Richards Model, in: Lam, W.H.K., Wong, S.C., Lo, H.K. (Eds.), *Transportation and Traffic Theory 2009: Golden Jubilee*. Springer US, pp. 603–635. doi:10.1007/978-1-4419-0820-9_30
- Karafyllis, I., Kontorinaki, M., Papageorgiou, M., 2016. Global exponential stabilization of freeway models. *International Journal of Robust and Nonlinear Control* 62, 1184–1210. doi:10.1002/rnc.3412

- Khoshyaran, M.M., Lebacque, J.-P., 2015. Capacity drop and traffic hysteresis as a consequence of bounded acceleration. *IFAC-PapersOnLine*, 48, 766-771. doi: [10.1016/j.ifacol.2015.05.160](https://doi.org/10.1016/j.ifacol.2015.05.160)
- Kontorinaki, M., Spiliopoulou, A., Roncoli, C., Papageorgiou, M., 2016. Capacity drop in first-order traffic flow models: overview and real-data validation. Presented at the 95th Annual Meeting of Transportation Research Board, p. no. 16-3541.
- Koshi, M., Iwasaki, M., Ohkura, I., 1983. Some findings and an overview of vehicular flow characteristics. Presented at the International Symposium on Transportation and Traffic Theory.
- Kotsialos, A., Papageorgiou, M., 2004. Nonlinear optimal control applied to coordinated ramp metering. *IEEE Transactions on Control Systems Technology* 12, 920–933. doi:10.1109/TCST.2004.833406
- Kuwahara, M., Akamatsu, T., 1997. Decomposition of the reactive dynamic assignment with queues for a many-to-many origin-destination pattern. *Transportation Research Part B: Methodological* 31, pp. 1-10.
- Landman, R.L., Hegyi, A., Hoogendoorn, S.P., 2015. Coordinated ramp metering based on on-ramp saturation time synchronisation. Presented at the 94th Annual Meeting of Transportation Research Board, p. no.15-4037.
- Laval, J.A., Daganzo, C.F., 2006. Lane-changing in traffic streams. *Transportation Research Part B: Methodological* 40, 251–264. doi:10.1016/j.trb.2005.04.003
- Lebacque, J., 2003. Two-phase bounded-acceleration traffic flow model: Analytical solutions and applications. *Transportation Research Record* 1852, 220–230. doi:10.3141/1852-27
- Lebacque, J.P., 1996. The Godunov scheme and what it means for first order traffic flow models. Presented at the International Symposium on Transportation and Traffic Theory, pp. 647–677.
- Lebacque, J.P., Buisson, C., Lesort, J.B., Mongeot, H., 1996. The Strada model for dynamic assignment. Presented at the Intelligent Transportation: Realizing the Future. Abstracts of the Third World Congress on Intelligent Transport Systems.
- Leclercq, L., Laval, J.A., Chiabaut, N., 2011. Capacity drops at merges: an endogenous model. *Transportation Research Part B: Methodological* 45, 1302–1313. doi:10.1016/j.trb.2011.05.007
- Leclercq, L., Knoop, V.L., Marczak, F., Hoogerndorn, S.P., 2016. Capacity drop at merges: New analytical investigations. *Transportation Research Part C: Emerging Technologies* 62, 171-187. doi: [10.1016/j.trc.2015.06.025](https://doi.org/10.1016/j.trc.2015.06.025)
- LeVeque, R.J., 2002. *Finite Volume Methods for Hyperbolic Problems*. Cambridge University Press.
- Li, Z., Liu, P., Xu, C., Wang, W., 2015. Optimal mainline variable speed limit control to improve safety on large-scale freeway segments. *Computer-Aided Civil and Infrastructure Engineering*. doi:10.1111/mice.12164
- Li, J., Chen, Q.Y., Wnag, H., Ni, D. 2012. Analysis of LWR model with fundamental diagram subject to uncertainties. *Transportmetrica* 8, 387-405. doi: [10.1080/18128602.2010.521532](https://doi.org/10.1080/18128602.2010.521532)
- Lighthill, M.J., Whitham, G.B., 1955. On kinematic waves II: A theory of traffic flow on long crowded roads. *Royal Society of London A: Mathematical, Physical and Engineering Sciences* 229, 317–345. doi:10.1098/rspa.1955.0089
- van Lint, J.W.C., Hoogendoorn, S., Schreuder, M., 2008. Fastlane: new multiclass first-order traffic flow model. *Transportation Research Record* 2088, 177–187. doi:10.3141/2088-19
- Lo, H.K., 2001. A cell-based traffic control formulation: strategies and benefits of dynamic timing plans. *Transportation Science* 35, 148–164. doi:10.1287/trsc.35.2.148.10136
- Messmer, A., Papageorgiou, M., 1990. METANET: a macroscopic simulation program for motorway networks. *Traffic Engineering and Control* 31, 466–470.
- Michalopoulos, P.G., Yi, P., Lyrintzis, A.S., 1992. Development of an improved high-order continuum traffic flow model. *Transportation Research Record* 1365, pp. 125–132.
- Monamy, T., Haj-Salem, H., Lebacque, J.-P., 2012. A macroscopic node model related to capacity drop. *Procedia - Social and Behavioral Sciences* 54, 1388–1396. doi:10.1016/j.sbspro.2012.09.853
- Munoz, L., Sun, X., Sun, D., Gomes, G., Horowitz, R., 2004. Methodological calibration of the cell transmission model. Presented at the American Control Conference, pp. 798–803.
- Muralidharan, A., Horowitz, R., 2015. Computationally efficient model predictive control of freeway networks. *Transportation Research Part C: Emerging Technologies, Special Issue: Advanced Road*

- Traffic Control 58, 532–553. doi:10.1016/j.trc.2015.03.029
- Newell, G.F., 1982. Applications of queueing theory, Vol. 733. Chapman & Hall, New York.
- Newell, G.F., 1993. A simplified theory of kinematic waves in highway traffic, part I: general theory, part II: queuing at freeway bottlenecks, part III: multi-destination flows. *Transportation Research Part B* 27, pp. 281–313. doi: [10.1016/0191-2615\(93\)90038-C](https://doi.org/10.1016/0191-2615(93)90038-C)
- Nie, X., Zhang, H.M., 2005. A comparative study of some macroscopic link models used in dynamic traffic assignment. *Networks and Spatial Economics* 5, 89–115. doi: 10.1007/s11067-005-6663-6
- Papageorgiou, M., Blosseville, J.-M., Hadj-Salem, H., 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, 345–359. doi:10.1016/0191-2607(90)90047-A
- Papageorgiou, M., Blosseville, J.-M., Hadj-Salem, H., 1989. Macroscopic modelling of traffic flow on the Boulevard Périphérique in Paris. *Transportation Research Part B: Methodological* 23, 29–47. doi:10.1016/0191-2615(89)90021-0
- Papageorgiou, M., Diakaki, C., Dinopoulou, V., Kotsialos, A., Wang, Y., 2003. Review of road traffic control strategies. *Proceedings of the IEEE* 91, 2043–2067. doi:10.1109/JPROC.2003.819610
- Papageorgiou, M., Hadj-Salem, H., Blosseville, J.M., 1991. ALINEA: A local feedback control law for on-ramp metering. *Transportation Research Record* 58–64.
- Payne, H.J., 1971. Models of freeway traffic and control. *Mathematical Models of Public Systems*.
- Richards, P.I., 1956. Shock waves on the highway. *Operations Research* 4, 42–51. doi:10.1287/opre.4.1.42
- Roncoli, C., Papageorgiou, M., Papamichail, I., 2015a. Traffic flow optimisation in presence of vehicle automation and communication systems – Part II: Optimal control for multi-lane motorways. *Transportation Research Part C: Emerging Technologies* 57, 260–275. doi:10.1016/j.trc.2015.05.011
- Roncoli, C., Papageorgiou, M., Papamichail, I., 2015b. Traffic flow optimisation in presence of vehicle automation and communication systems – Part I: A first-order multi-lane model for motorway traffic. *Transportation Research Part C: Emerging Technologies* 57, 241–259. doi:10.1016/j.trc.2015.06.014
- Smith, M.J., 1984. The existence of a time-dependent equilibrium distribution of arrivals at a single bottleneck. *Transportation Science* 18, pp. 385–394.
- Smith, B., Qin, L., Venkatanarayana, R., 2003. Characterization of freeway capacity reduction resulting from traffic accidents. *Journal of Transportation Engineering* 129, 362–368. doi:10.1061/(ASCE)0733-947X
- Spiliopoulou, A., Kontorinaki, M., Papageorgiou, M., Kopelias, P., 2014. Macroscopic traffic flow model validation at congested freeway off-ramp areas. *Transportation Research Part C: Emerging Technologies* 41, 18–29. doi:10.1016/j.trc.2014.01.009
- Srivastava, A., Geroliminis, N., 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. doi:10.1016/j.trc.2013.02.006
- Srivastava, A., Jin, W. L., Lebacque, J.-P., 2015. A modified cell transmission model with realistic queue discharge features at signalized intersections. *Transportation Research Part B: Methodological*, 81, pp. 302–315. doi: [10.1016/j.trb.2015.05.013](https://doi.org/10.1016/j.trb.2015.05.013)
- Sun, X., Horowitz, R., 2005. A localized switching ramp-metering controller with a queue length regulator for congested freeways. Presented at the Proceedings of the American Control Conference, pp. 2141–2146. doi:10.1109/ACC.2005.1470287
- Torné, J.M., Soriguera, F., Geroliminis, N., 2014. Coordinated active traffic management freeway strategies using capacity-lagged cell transmission model. Presented at the 93rd Annual Meeting of Transportation Research Board. p. no. 14-3941.
- Treiber, M., Kesting, A., Helbing, D., 2006. Understanding widely scattered traffic flows, the capacity drop, and platoons as effects of variance-driven time gaps. *Physical Review E* 74, 16123. doi:10.1103/PhysRevE.74.016123
- Treiterer, J., Myers, J., 1974. The hysteresis phenomenon in traffic flow. *Transportation and Traffic Theory* 6, 13–38.

- Whitham, G.B., 2011. Linear and nonlinear waves. John Wiley & Sons.
- Wong, G.C.K., Wong, S.C. 2002. A multi-class traffic flow model—an extension of LWR model with heterogeneous drivers. *Transportation Research Part A: Policy and Practice* 36, pp. 827-841. doi: [10.1016/S0965-8564\(01\)00042-8](https://doi.org/10.1016/S0965-8564(01)00042-8)
- Yperman, I., Logghe, S., Tampere, C., Immers, B., 2006. The multi-commodity link transmission model for dynamic network loading. Presented at the 85th Annual Meeting of Transportation Research Board. p. no. 06-1062.
- Yuan, K., Knoop, V., Schreiter, T., Hoogerendoorn, S., 2015. A hybrid kinematic wave model incorporating capacity drops. Presented at the Proceedings of the IEEE 18th International Conference on Intelligent Transportation Systems, pp. 471–476. doi:10.1109/ITSC.2015.84
- Zhang, H., Ritchie, S.G., Recker, W.W., 1996. Some general results on the optimal ramp control problem. *Transportation Research Part C: Emerging Technologies* 4, pp. 51–69. doi:10.1016/0968-090X(96)00002-2
- Zhang, H.M., 2001. New perspectives on continuum traffic flow models. *Networks and Spatial Economics* 1, 9–33. doi: 10.1023/A:1011539112438
- Zhang, H.M., 2002. A non-equilibrium traffic model devoid of gas-like behavior. *Transportation Research Part B* 36, 275–290. doi: [10.1016/S0191-2615\(00\)00050-3](https://doi.org/10.1016/S0191-2615(00)00050-3)
- Ziliaskopoulos, A.K., 2000. A linear programming model for the single destination system optimum dynamic traffic assignment problem. *Transportation Science* 34, pp. 37–49. doi:10.1287/trsc.34.1.37.12281

Appendix

The analysis of the proposed new discretized model (Approach 5) with respect to its relation with the continuous-space and continuous-time PDE of the LWR is presented here. More specifically, we investigate how the introduced parameters of Approach 5 affect the solution of the discretized model when the discretization parameters (cell lengths and simulation time step) tend to zero. However, in order to conduct such an analysis, it is important to determine the way the on-ramp flow term is treated. There are two main ways to treat the on-ramp: first by assuming that the on-ramp flow term is distributed flow within specific space bounds; second by assuming the on-ramp flow is a concentrated (Dirac function) flow at a given space point of the freeway. Sections A.1 and A.2 present the analysis for the two respective cases.

In any case, we consider the difference-equation state-space model described by equations (1), (13), (14) and (4), which describe the proposed model (Approach 5) for which this analysis is performed. Notice that, in case $\theta_r = 1$ (in (13)) and $f_{D,i}(\rho_i) = Q_i$ for $\rho_i \geq \rho_{cr}$ (in (14)), the basic discretized LWR model is obtained. Additionally, if $g(\rho_i) = v_{f,i} \rho_i l_i$ (in (14)), then we end up with the CTM. The discretization time step is denoted as previously with T h. In order to simplify the present analysis, we consider that the above equations describe a single-lane homogeneous freeway cell with no off-ramps (which are not considered important in the present analysis). Thus, we assume that $p_i = 0$, $L_i = L$, $f_{D,i} = f_D$, $f_{S,i} = f_S$ and $f_i = f$ for $i = 1, \dots, n$. Let A km and B km be the starting and ending points of the freeway stretch, respectively. Then the total freeway length is $B - A$ km, and the proposed model is described by

$$\rho_i(k+1) = \rho_i(k) + \frac{T}{L} (-f(\rho_i(k), \rho_{i+1}(k), r_{i+1}(k)) + f(\rho_{i-1}(k), \rho_i(k), r_i(k)) + r_i(k)) \quad (\text{A.1})$$

where

$$f(\rho_i(k), \rho_{i+1}(k), r_{i+1}(k)) = \min \{f_D(\rho_i(k)), f_S(\rho_{i+1}(k)) - \theta_r r_{i+1}(k)\}, \quad (\text{A.2})$$

and f_D , f_S are given by (14), (4) and $0 < \theta \leq 1$.

A.1. Distributed on-ramp flow

In this case we assume that each on-ramp has an acceleration lane of length equal to L_{rmp} km. We consider the following definitions

$$\rho_i(k) := \frac{1}{L} \int_{iL}^{iL+L} \rho(kT, z) dz \quad (\text{A.3})$$

$$r_i(k) = r_i(kT) := \int_{iL}^{iL+L} u(kT, z) dz \quad (\text{A.4})$$

where $\rho \in C^1$ ([veh/km]) corresponds to the density and $u \in C^0$ ([veh/time/space]) corresponds to the inflow per unit space at a given time instant. Notice that, by definitions (A.3) and (A.4), we have considered average density and inflow values, respectively, for every cell. Since a uniform distribution of the entering flow from the on-ramp to a number of cells is assumed, we have for each such cell

$$u(kT, z) = \frac{r_{rmp}(kT)}{L_{rmp}}, \quad (\text{A.5})$$

where $r_{rmp}(kT)$ is the total on-ramp inflow. Moreover, we consider arbitrary C^1 functions $f_D(\rho)$ and $f_S(\rho)$ that satisfy $f_D(\rho) = f_S(\rho)$ for $\rho = \rho_{cr}$, $f_D(\rho) < f_S(\rho)$ for $\rho < \rho_{cr}$, $f_D(\rho) > f_S(\rho)$ for $\rho > \rho_{cr}$, $f_D(0) = 0$ and $f_S(\rho_{max}) = 0$. Notice that the above requirements are satisfied for the demand and supply functions considered in (14) and (4) respectively. Substituting (A.3) and (A.4) in (A.1) we obtain

$$\begin{aligned} & \frac{1}{T} \left(\int_{iL}^{iL+L} (\rho(kT, z) - \rho(kT+T, z)) dz \right) = \\ & -f \left(\frac{1}{L} \int_{iL}^{iL+L} \rho(kT, z) dz, \frac{1}{L} \int_{iL+L}^{iL+2L} \rho(kT, z) dz, \int_{iL+L}^{iL+2L} u(kT, z) dz \right) + f \left(\frac{1}{L} \int_{iL-L}^{iL} \rho(kT, z) dz, \frac{1}{L} \int_{iL}^{iL+L} \rho(kT, z) dz, \int_{iL}^{iL+L} u(kT, z) dz \right) \\ & + \int_{iL}^{iL+L} u(kT, z) dz \end{aligned} \quad (\text{A.6})$$

Set $t = kT$ and $iL = x$. Then, using the Mean Value Theorem (MVT) for $T \rightarrow 0$, it follows from (A.6) that

$$\begin{aligned} & \int_x^{x+L} \frac{\partial \rho}{\partial t}(t, z) dz = \\ & -f \left(\frac{1}{L} \int_x^{x+L} \rho(t, z) dz, \frac{1}{L} \int_{x+L}^{x+2L} \rho(t, z) dz, \int_{x+L}^{x+2L} u(t, z) dz \right) + f \left(\frac{1}{L} \int_{x-L}^x \rho(t, z) dz, \frac{1}{L} \int_x^{x+L} \rho(t, z) dz, \int_x^{x+L} u(t, z) dz \right) + \int_x^{x+L} u(t, z) dz \end{aligned} \quad (\text{A.7})$$

We use the following definition

$$F(t, x, L) := f \left(\frac{1}{L} \int_x^{x+L} \rho(t, z) dz, \frac{1}{L} \int_{x+L}^{x+2L} \rho(t, z) dz, \int_{x+L}^{x+2L} u(t, z) dz \right) = f(a_1, a_2, a_3), \quad (\text{A.8})$$

where a_1, a_2, a_3 have been introduced as auxiliary variables to ease the analysis corresponding to the arguments of the flux function f . Combining (A.7) and (A.8) we obtain

$$\int_x^{x+L} \left(\frac{\partial \rho}{\partial t}(t, z) - u(t, z) \right) dz = -F(t, x, L) + F(t, x - L, L). \quad (\text{A.9})$$

Then, we make the following substitution

$$-F(t, x, L) + F(t, x - L, L) = - \int_{x-L}^x \frac{\partial F}{\partial z}(t, z, L) dz,$$

which combining with (A.9) yields the following relation, which holds for every $L > 0$

$$\int_x^{x+L} \left(\frac{\partial \rho}{\partial t}(t, z) - u(t, z) + \frac{\partial F}{\partial z}(t, z - L, L) \right) dz = 0 \quad (\text{A.10})$$

Then, using the fact that (A.10) holds for every $L > 0$ and the MVT for $L \rightarrow 0$ we have that

$$\frac{\partial \rho}{\partial t}(t, x) = - \frac{\partial F}{\partial x}(t, x, 0) + u(t, x). \quad (\text{A.11})$$

Now, in order to obtain the flux function F in terms of the demand and supply functions, we follow the next steps. For constant $L > 0$ we have that

$$\begin{aligned} \frac{\partial F}{\partial x}(t, x, L) &= \frac{\partial f}{\partial a_1} \left(\frac{1}{L} \int_x^{x+L} \rho(kT, z) dz, \frac{1}{L} \int_{x+L}^{x+2L} \rho(kT, z) dz, \int_{x+L}^{x+2L} u(kT, z) dz \right) \left(\frac{1}{L} (\rho(t, x+L) - \rho(t, x)) \right) + \\ &\frac{\partial f}{\partial a_2} \left(\frac{1}{L} \int_x^{x+L} \rho(kT, z) dz, \frac{1}{L} \int_{x+L}^{x+2L} \rho(kT, z) dz, \int_{x+L}^{x+2L} u(kT, z) dz \right) \left(\frac{1}{L} (\rho(t, x+2L) - \rho(t, x+L)) \right) + \\ &\frac{\partial f}{\partial a_3} \left(\frac{1}{L} \int_x^{x+L} \rho(kT, z) dz, \frac{1}{L} \int_{x+L}^{x+2L} \rho(kT, z) dz, \int_{x+L}^{x+2L} u(kT, z) dz \right) (u(t, x+2L) - u(t, x+L)). \end{aligned} \quad (\text{A.12})$$

Thus, using the Mean Value Theorem (MVT) for $L \rightarrow 0$ and using (A.2) we obtain

$$\frac{\partial F}{\partial x}(t, x, 0) = \left\{ \begin{array}{l} f_D'(\rho(t, x)) \text{ if } \rho(t, x) \leq \rho_{cr} \\ 0 \text{ if } \rho(t, x) > \rho_{cr} \end{array} \right\} + \left\{ \begin{array}{l} 0 \text{ if } \rho(t, x) \leq \rho_{cr} \\ f_S'(\rho(t, x)) \text{ if } \rho(t, x) > \rho_{cr} \end{array} \right\} \frac{\partial \rho}{\partial x}(t, x) \quad (\text{A.13})$$

which in turn corresponds to the following PDE

$$\frac{\partial \rho}{\partial t}(t, x) + \frac{\partial}{\partial x} q(\rho(t, x)) = u(t, x) \text{ for } t > 0 \text{ and } A < x < B \quad (\text{A.14})$$

with

$$q(\rho) = \int_0^\rho \left(\frac{\partial f}{\partial a_1}(z, z, 0) + \frac{\partial f}{\partial a_2}(z, z, 0) \right) dz = \begin{cases} f_D(\rho) & \text{if } \rho \leq \rho_{cr} \\ f_S(\rho) & \text{if } \rho > \rho_{cr} \end{cases}. \quad (\text{A.15})$$

The above analysis proves the consistency of the discretized model (A.1), (A.2) with the LWR model (A.14). The result has been obtained by considering smooth functions f_D, f_S, f , as well as the fact that $\rho(t, x)$ lies within appropriate intervals. However, a similar analysis as above can be made even for piecewise differentiable functions. In such cases the solution converges to generalized (weak) solutions of (A.14). Therefore, the discrete-time model is consistent to the LWR model (A.14).

This result is supported by some conducted numerical experiments (Fig. 10 (a), (b), (c)). The simulation results depicted in these figures have been obtained using the same illustrative scenario of Section 3.3. More specifically, density and flow for the complete freeway stretch at the same time instant (i.e., $t = 1.9$ h) are depicted for the two different schemes (CTM and A5) and for three different spatial discretizations ($L = 0.5, L = 0.1, L = 0.05$ km). The numerical results indicate that, as L tends to zero, both schemes converge to the same solution.

A.2. Concentrated on-ramp flow as a singular source

In this section, we show that, when the on-ramp is incorporated as a singular source (for both CTM and Approach 5 model), the effect of the on-ramp contribution is more pronounced, independently of the spatial discretization, and a modified shock speed is derived for the model described by Approach 5 in comparison with CTM. We point out here that, in the case where the on-ramp contribution is incorporated as a singular source (i.e., as a Dirac function), the differential form of the conservation law is not valid in the position where the on-ramp is implemented; only the integral form of the conservation law is valid there. Consequently, the comparison between the CTM and Approach 5 is materialized on this base.

Referring again to the model (A.1) - (A.2), we use the compact forms $\rho_i^k = \rho_i(k), r_i^k = r_i(k), f_D(\rho_i(k)) = f_{D,i}^k$ and $f_S(\rho_i(k)) = f_{S,i}^k$. Assuming that the on-ramp is implemented only at a distinct point of

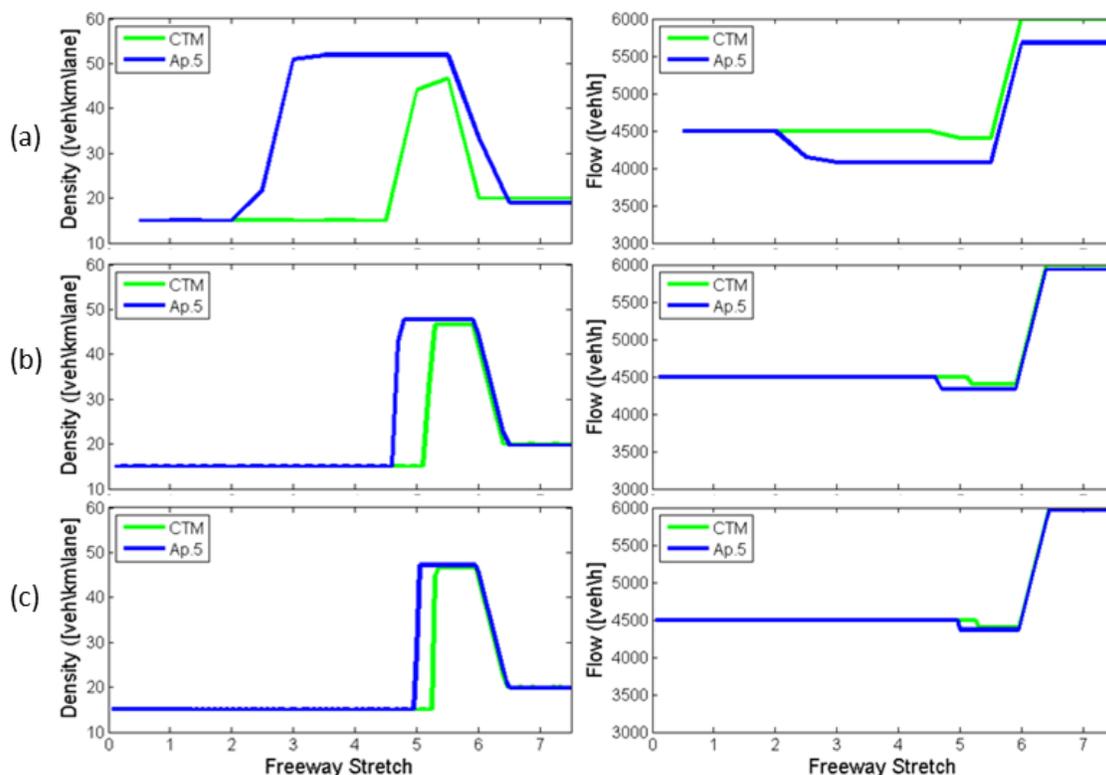


Fig. 10. (Distributed on-ramp flow) The density (left) and the flow (right) at $t = 1.9$ h in the freeway stretch for (a) $L = 0.5$ km, (b) $L = 0.1$ km and (c) $L = 0.05$ km.

the i^{th} computational cell, and there are no other similar sources within the freeway (i.e., $r_j = 0$ for $j \neq i$), we obtain from (A.1) with (A.2) that

$$\rho_i^{k+1} = \rho_i^k + \frac{T}{L} \left(-\min(f_{D,i}^k, f_{S,i+1}^k - 0) + \min(f_{D,i-1}^k, f_{S,i}^k - \theta_r r_i^k) + r_i^k \right) \quad (\text{A.16})$$

where $r_i^k = r_{rmp}(t) = r_{rmp}(kT)$. We recall again here that the differences between the CTM and Approach 5 lie on the parameter θ_r , and the right part of the demand function. Thus, referring to Fig. 11(i) and by denoting with s the speed of the left moving shock, we have that $L = -sT$. Then, we obtain from (A.16) that

$$s(\rho_i^{k+1} - \rho_i^k) = \min(f_{D,i}^k, f_{S,i+1}^k) - \min(f_{D,i-1}^k, f_{S,i}^k - \theta_r r_i^k) - r_i^k. \quad (\text{A.17})$$

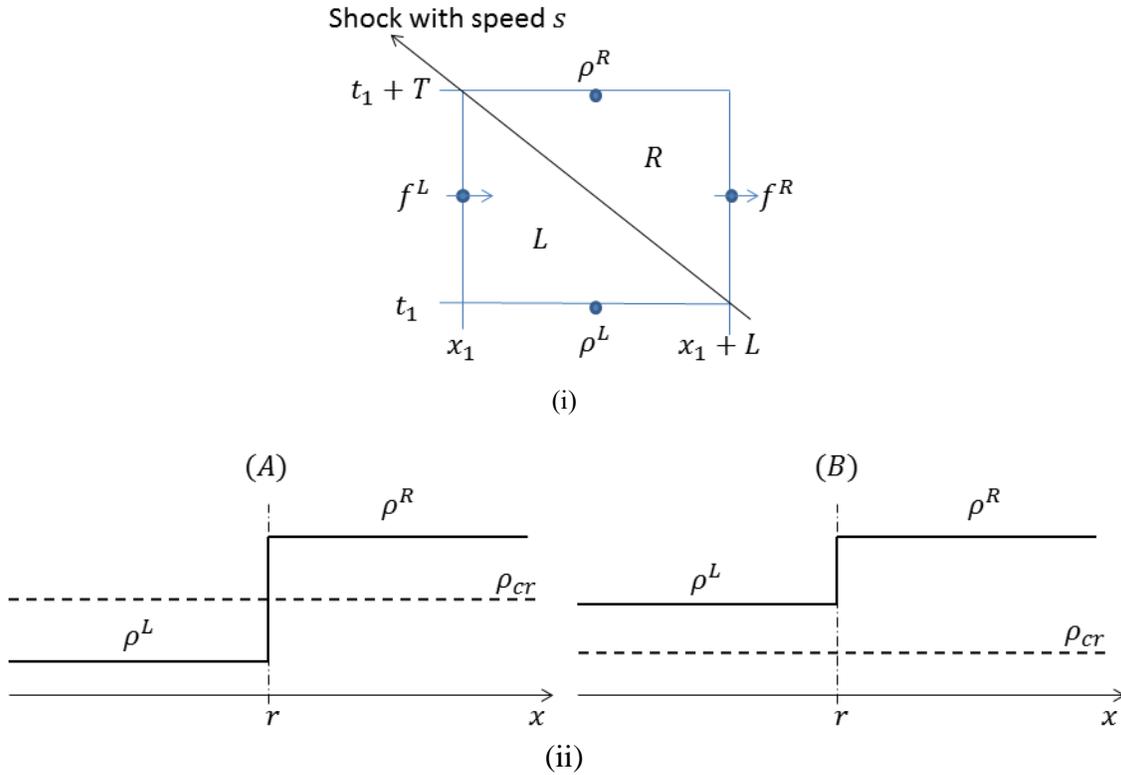


Fig. 11. (i) The Rankine-Hugoniot jump conditions are determined by integrating over an infinitesimal rectangular region on the $x-t$ plane, (ii) Characteristic Riemann problems.

We will compute jump conditions at the location of the on-ramp using the procedure proposed by (LeVeque, 2002). Fig. 11(i) depicts a small rectangular region in which a shock wave is present. The Rankine-Hugoniot jump conditions will be derived using this illustrative example. In this case, (A.17) becomes

$$s(\rho^R - \rho^L) = \min(f_D^R, f_S^R) - \min(f_D^L, f_S^L - \theta_r r_i) - r_i. \quad (\text{A.18})$$

We assume that the on-ramp flow cannot exceed the capacity of the network and therefore $Q > r_i$. Moreover, in order to simplify the present analysis, we consider $g(\rho_i) = v_f \rho_i$. Thus, we have that $\rho_{cr} = Q/v_f$, $Q = w(\rho_{\max} - \rho_{cr})$ and, therefore, $\rho_{\max} = (v_f/w + 1)\rho_{cr}$. The two considered illustrative cases, which we intent to examine here, are shown in Fig. 11(ii). Then, (A.18) imposes that the resulting speed of the shock wave depends on the intervals in which ρ^R and ρ^L lie. Therefore, we distinguish points $\rho_1 = (Q - r_i)/v_f$, $\rho_2 = (Q - \theta_r r_i)/v_f$ and $\rho_3 = \rho_{\max} - r_i/w$. Notice that $0 < \rho_1 < \rho_2 < \rho_{cr} < \rho_3 < \rho_{\max}$.

Table 5 shows the resulting speed and sign of the shock for case (A), for both CTM and Approach 5. As it can be clearly seen, the intensity of the shock clearly depends on the magnitude of the on-ramp flow. For that reason, there are cases ($\rho_2 < \rho^L < \rho_{cr}$) where the intensity of the shock is larger for Approach 5 due to the effect of the parameter θ_r . For the case (B) the shock speed is negative (the shock wave moves to the left), no matter what are the values of ρ^R and ρ^L . However, as it is shown in Table 6, the intensity of the shock is higher for Approach 5, depending again on the selection of the parameter θ .

Table 5
Speed and sign of the shock for CTM and Approach 5 for case (A).

(A)		CTM		Approach 5	
$0 < \rho^L < \rho_1$	$\rho_{cr} < \rho^R < \rho_3$	+ or -	$s = \frac{w(\rho_{\max} - \rho^R) - v_f \rho^L - r_i}{\rho^R - \rho^L}$	+ or -	$s = \frac{w(\rho_{\max} - \rho^R) - v_f \rho^L - r_i}{\rho^R - \rho^L}$
	$\rho_3 < \rho^R < \rho_{\max}$	-		-	
$\rho_1 < \rho^L < \rho_2$	$\rho_{cr} < \rho^R < \rho_3$	-	$s = \frac{w(\rho_{\max} - \rho^R) - Q}{\rho^R - \rho^L}$	+ or -	$s = \frac{w(\rho_{\max} - \rho^R) - v_f \rho^L - r_i}{\rho^R - \rho^L}$
	$\rho_3 < \rho^R < \rho_{\max}$	-		-	
$\rho_2 < \rho^L < \rho_{cr}$	$\rho_{cr} < \rho^R < \rho_3$	-	$s = \frac{w(\rho_{\max} - \rho^R) - Q}{\rho^R - \rho^L}$	-	$s = \frac{w(\rho_{\max} - \rho^R) - Q - (1 - \theta_r)r_i}{\rho^R - \rho^L}$
	$\rho_3 < \rho^R < \rho_{\max}$	-		-	

Table 6
Sign and intensity of the shock for CTM and Approach 5 for case (B).

(B)	CTM	Approach 5
$\rho_{cr} < \rho^L < \rho^R < \rho_{\max}$	$-w$	$-w - \frac{(1 - \theta_r)r_i}{\rho^R - \rho^L}$

The previous theoretical observations are illustrated by the numerical experiments shown in Fig. 12 using the same illustrative scenario of Section 3.3. More specifically, density and flow for the complete freeway stretch at the same time instant (i.e., $t = 1.9$ h) are depicted for the two different schemes (CTM and Approach 5) and for three different spatial discretizations ($L = 0.5$, $L = 0.1$, $L = 0.05$ km). The numerical results indicate that, as L tends to zero, the two schemes converge to a different solution; the scheme of

Approach 5 results in a higher (negative) shock speed and a higher density at the congested region compared to the shock generated by CTM.

Following from the previous analysis, it is clear that the proposed scheme (Approach 5) with its modified supply and demand functions introduces a modified jump condition in the Riemann problems depicted previously, when the on-ramp flow is considered as a concentrated singular source term in a single cell. In this case, the proposed modifications to the supply and demand functions are always present as L tends to zero. These modifications can be viewed as imposed inhomogeneities in the on-ramp cell which result in a generalized Riemann problem solved by a modified Godunov scheme similar to those presented by (Jin et al., 2009; Lebacque, 1996). On the contrary, when a distributed on-ramp entering flow is considered (as in Section A.1), the corresponding modifications to the supply and demand functions vanish as L tends to zero.

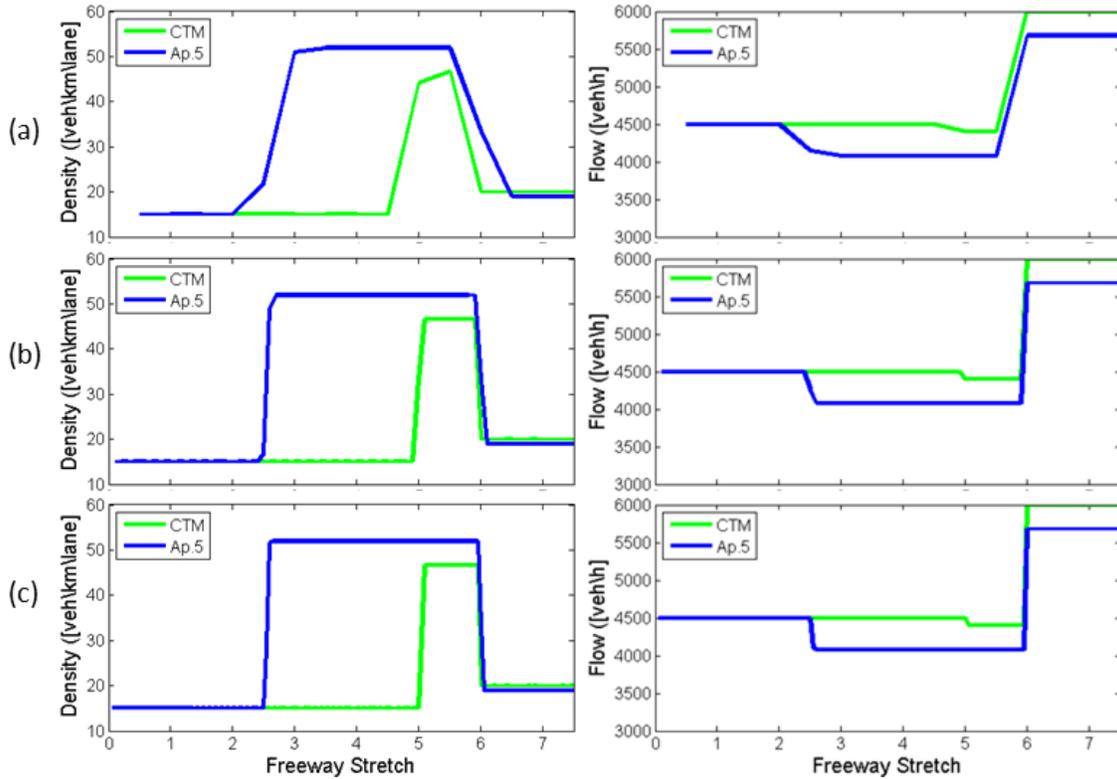


Fig. 12. (Concentrated on-ramp flow) The density (left) and the flow (right) at $t = 1.9h$ in the freeway stretch for (a) $L = 0.5$ km, (b) $L = 0.1$ km and (c) $L = 0.05$ km.