



HAL
open science

Macroscopic Traffic Dynamics Under Fast-Varying Demand

Ludovic Leclercq, Mahendra Paipuri

► **To cite this version:**

Ludovic Leclercq, Mahendra Paipuri. Macroscopic Traffic Dynamics Under Fast-Varying Demand. Transportation Science, 2019, 53 (6), pp1-20. 10.1287/trsc.2019.0908 . hal-02497472v1

HAL Id: hal-02497472

<https://hal.science/hal-02497472v1>

Submitted on 3 Mar 2020 (v1), last revised 24 Mar 2020 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Macroscopic traffic dynamics in reservoirs under fast-varying demand profiles

Ludovic Leclercq (Corresponding author)

Professor

Univ. Lyon, IFSTTAR, ENTPE, LICIT, F-69518, Lyon, France

Ludovic.leclercq@ifsttar.fr

Mahendra Paipuri

Post-doc

Univ. Lyon, IFSTTAR, ENTPE, LICIT, F-69518, Lyon, France

mahendra.paipuri@ifsttar.fr

Paper submitted for publication in Transportation Science

Original submission date: September, 6th, 2018

Revision 1 date: December, 10th, 2018

Revision 2 (Accepted version): February, 6th, 2018

Abstract

Aggregated dynamic traffic models based on the Network Macroscopic Fundamental Diagram (MFD) have been founded on stationary or steady-state approximations. Theoretically, they are only qualified with slow-varying demand profiles as inputs. In practice, MFD models are often fed with fast-varying demand patterns classically observed during peak hours. This paper investigates the accuracy of existing MFD model outputs in such a case. It also provides a fresh look at hysteresis loops that may arise on MFD shapes and show how they are connected to the demand profile.

The exact solutions of the LWR model for a simple, but meaningful test case corresponding to the loading and the recovery of an arterial with a single bottleneck are first derived. The spatial integration of these solutions permits to derive the average flow- and the outflow-MFD patterns. The average flow-MFD exhibits a clockwise hysteresis loop with higher flows during the network loading than during the recovery. The outflow MFD shows an opposite trend with a counter-clockwise hysteresis loop. The detailed analysis of the LWR solution highlights that these patterns are fully driven by the congestion dynamics and different wave propagation during both periods. This complements existing studies that tend to explain hysteresis by heterogeneous network loading due to unbalanced flows at intersections. It should be noted that the size of the production-MFD hysteresis loop is influenced by the demand loading profile, which means that the MFD is more sensitive to the demand than initially expected. The second part of the paper is about the implications of these observations on the MFD simulation. The main conclusion is that the accumulation-based formulation captures properly the aggregate dynamics of the LWR model during saturation but not in free-flow. It should be calibrated with an average flow-MFD that does not account for the hysteresis pattern to properly estimate the outflow. On the contrary, the original formulation of the trip-based model is only valid in free-flow. During saturation, it should be patched to modify the outflow values. The best approach is a hybrid model that combines the trip-based model in free-flow and the accumulation-based model during saturation. Note that in free-flow the trip-based model is better fed by the average flow-MFD with hysteresis as it gives better estimation for the mean speed. Finally, some evidences are presented and tend to confirm that all the conclusions remain valid for more complex urban settings.

Introduction

The concept of (Network) Macroscopic Fundamental Diagram (MFD) is currently very popular in traffic flow theory as it permits to revisit traffic simulation, management and control at large-city scale. Basically, the MFD corresponds to a well-defined relationship between average flow and density over an urban or freeway network included within a given perimeter. The idea of an MFD belongs to Godfrey (1969) and similar approaches were introduced later by Herman and Prigogine (1979), Mahmassani *et al.* (1984, 1987) and Daganzo (2007). The verification of its existence with dynamic features has been firstly realized using loop data from downtown Yokohama (Geroliminis and Daganzo, 2008).

Later, numerous studies based on experimental or simulation data have then confirmed the existence of the MFD with low scatter. However, other studies have also highlighted the limitations of the MFD as a unimodal and invariant curve able to perfectly describe all mean traffic conditions within a region. (Buisson and Ladier, 2009) first showed that a bimodal curve corresponding to a clockwise hysteresis pattern has been observed for a particular day in the city of Toulouse (France). It appears there that the network is performing better during the loading (onset of congestion) than during the recovery (offset of congestion) leading to a clockwise hysteresis-like loop when plotting all day observations. (Daganzo, 2011) provides both analytical and numerical insights to explain such a phenomenon. Based on a closed ring-road setup, the former paper shows that the transition from free-flow to congestion is stable but that congestion state is then unstable. At the end of the loading process, congestion is unevenly distributed over the network leading to more instability during recovery and lower flow rate. So, as the network is unstable in congestion, pockets of heavy congestion locally appear and prevent the system from an efficient recovery. The clockwise MFD hysteresis pattern is then related to unevenly distributed traffic conditions that are more likely to happen during the offset than the onset of congestion. This has been confirmed by several further studies based on freeway data (Geroliminis and Sun, 2011; Saberi and Mahmassani, 2012; 2013). It is also mentioned that the hysteresis loop is more likely to be large when drivers are not able to adapt their trips to avoid congested areas by lack of information or routing options, which is more often the case on freeways. More generally, other studies have highlighted the correlation between the network spatial heterogeneity in terms of traffic conditions and the MFD shape, e.g. (Mazlounian *et al.*, 2010; Knoop and Hoogendoorn, 2013). Another factor that appears to influence the MFD shape is the demand pattern. (Leclercq *et al.*, 2015) shows based on numerical simulation that changes in the OD matrix modify the internal routing patterns. (Daganzo *et al.*, 2011; Mahmassani *et al.*, 2013) both focus on urban network gridlocks. (Daganzo, 2011) shows that faster network loadings lead to gridlocks with higher accumulation compared to slower ones. (Mahmassani *et al.* 2013) shows that higher demand level during the congestion peak results in larger clockwise hysteresis loop.

Defining the right shape of the MFD is crucial for simulation applications as it will obviously influence the accuracy of the simulation results. First, when it comes to MFD simulation, it is important to distinguish the average flow-MFD (p-MFD), which relates the mean internal flow to the mean density, from the outflow-MFD (o-MFD), which relates the exiting flow from a region to the mean density. The exiting flow is also called the trip completion rate as it includes both vehicles that are crossing the perimeter and those that are ending their trips inside the area. Also defined in the literature, the production-MFD, i.e. the travel production vs. the vehicle accumulation in an area, corresponds to the average flow-MFD multiplied by a constant scaling factor equal to the inverse of the network total length. These two diagrams are then very similar and this explains why, while focusing on the average flow-MFD, we choose the abbreviation p-MFD, which is the most common in the literature. The o-MFD is also known as the network exit function (Daganzo, 2007). It is hardly observable in practice except from simulation data. Note that o-MFD and p-MFD have been proven strictly proportional in steady-state conditions only (Daganzo, 2007). This relationship is however often applied in MFD simulators outside its strict validity domain. With regards to MFD models, two different expressions can be found in the literature. The accumulation-based formulation corresponds to a conservation equation where the outflow is given by the o-MFD, see (Daganzo, 2007; Geroliminis, 2007) for the foundations. The trip-based formulation (Arnott, 2013; Fosgerau, 2015; Lamotte and Geroliminis, 2017) is focusing on internal vehicle trips, whose speed is related to the p-MFD. (Mariotte *et al.*, 2017) and (Laval *et al.*, 2018) propose a complete review of mathematical and physical properties of both formulations.

For now, most implementations of accumulation- and trip-based models for simulation purposes are using unimodal MFD shapes. One rationale is that unimodal MFD curves represent stationary traffic states while observed deviations and hysteresis patterns are related to transition periods like congestion onset and offset. However, this means that MFD models are only qualified for slow-varying demand patterns, which fulfill the quasi-static approximation. In practice, designing urban models for reproducing congestion propagation during peak hours require to apply fast-varying demand profiles. Such loadings are very common in MFD simulation, while little attention has been paid to verify whether the outputs are still accurate. This paper aims to fill this gap by benchmarking existing MFD models against the aggregation of the local traffic dynamics for a typical peak hour. In particular, we will investigate the clockwise hysteresis pattern that appears during congestion peaks when aggregating local observations and to look if and how MFD model outputs are reproducing this phenomenon.

First, we will propose a refined description of the congestion mechanisms that are triggering the clockwise hysteresis loop when a recovery period follows a loading one and emphasize the central role played by internal bottlenecks. We are going to demonstrate that local traffic dynamics (congestion wave propagation) at a bottleneck can explain by itself the hysteresis pattern even in the very simple and stable case of an urban arterial. This is not contradictory with previous explanations that consider the distribution of the congestion over the network but clearly highlights the importance of the congestion dynamics. This can be considered as an additional factor, which is also related to the demand pattern. As our analysis is based on analytical and numerical calculations, we will be able to study not only the p-MFD but also the o-MFD shapes. In particular, we are going to show that the o-MFD is exhibiting a counter-clockwise hysteresis pattern during congestion peak. To our best knowledge, such a pattern has never been highlighted with local observations and only appears on trip-based MFD simulation results. Second, we will study in detail the capabilities of existing MFD models to reproduce the hysteresis phenomena. This will provide significant insights about the best calibration and implementation methods for both MFD models in order to improve traffic dynamics consistency between the aggregate and local scales.

Our study is focused on the simplest case of an urban reservoir that can be found in the literature, i.e. an arterial with no internal turns (or well-balanced lateral flows). On one hand, the existence of a well-defined unimodal MFD has already been proven for arterials and stationary traffic conditions and its exact expression can analytically be derived using the cuts theory (Daganzo and Geroliminis, 2008). On the other hand, the computation of the exact solutions of the LWR model is also possible. This permits to investigate the deviations of aggregated local traffic states from the theoretical stationary MFD during a dynamic loading and discharge. The arterial setting includes an internal bottleneck, which triggers congestion. This setting may look very simplistic, but we are convinced that it stands as a good proxy for much more complex urban networks. Most of them exhibit internal bottlenecks (bridges, tunnels, major intersections...), which activate congestion patterns when the demand exceeds the capacity. So, the loading of such networks looks like a demand peak that is saturating the internal bottlenecks and lead to backward propagation of congestion, just like in the arterial case. A complete saturation of the network is achieved only when the congestion has reached the perimeter boundaries. So, we find important to first propose a complete description of this simple case to really understand the underlying physical mechanisms. Then, we will provide numerical evidences that show that our findings can be generalized for more complex urban situations.

The paper is organized as follows. Section 1 presents the integration of local traffic dynamics in the case of an arterial with a bottleneck. It explains the clockwise hysteresis loops and shows that its magnitude is highly related to the demand pattern (intensity of the peak). Section 2 investigates how existing MFD models reproduce the arterial dynamics compared to the integrated local solutions. In particular, we will show that if the trip-based performs better in free-flow conditions than the accumulation-based model, it may exhibit inaccurate patterns when saturation arises and have to be patched. Section 3 not only provides more evidence of the hysteresis loops (clockwise in p-MFD and counterclockwise in o-MFD) but also presents further comparisons of MFD models with micro-simulation for more complex urban settings. They fully confirm the conclusions we draw on the arterial case. Finally, section 4 presents the conclusions and a short discussion.

1. Scaling-up local traffic dynamics to a single reservoir: the arterial case with a bottleneck

In this section, we are considering an arterial with multiple traffic signals as an idealized representation of an urban network. Among all signals, one has a reduced green time and then acts as a bottleneck. This is a very typical situation of an intersection with similar flow values for conflicting movements where green times have to be shared. Local traffic flow dynamics is calculated based on exact methods (analytical and numerical) and then aggregated to determine the resulting time-dependent MFD patterns.

1.1. Description of the arterial case and the network state variables

Let us consider an arterial with 8 links of 135 m length, see Figure 1a. All internal connections are controlled by traffic signals with green time $G=40$ s, cycle time $C=60$ s and offset $\delta=3$ s. The fifth signal is acting as a bottleneck as its green time is reduced to 30 s. Traffic dynamics is reproduced by the Lighthill, Whitham and Richards (LWR) model (Lighthill and Whitham, 1955; Richards, 1956) with a triangular fundamental diagram: free-flow speed $u=15$ m/s, wave speed $w=5$ m/s and jam density $\kappa=0.19$ veh/m. The maximal capacity $S=uw\kappa/(u+w)$ is equal to 0.7125 veh/s. The effective capacity at all traffic signals but the fifth is $\mu_s=0.475$ veh/s. At the bottleneck, the effective capacity is $\mu=0.3563$ veh/s. Note that this basic arterial setting is good proxy even for heterogeneous signal timings for each intersection. In that latter case, the intersection with the lowest capacity, or a bunch of intersections when short-block effect happens (Laval and Castrillon, 2015), acts as the bottleneck and drives the major congestion pattern.

The arterial is empty at the time $t=0$ and then loaded with a trapezoidal-like demand profile $\lambda(t)$ representing a classical peak period, see Figure 1b. Note that $\lambda_1 > \mu$ in order congestion to appear and spillback in the arterial. Depending on the values of λ_0 and λ_1 and the duration between T_3 and T_4 , the peak hour can be more or less pronounced. The duration between T_2 and T_3 (respectively T_4 and T_5) describes how sharp is the loading (respectively recovery) phase. The reference demand scenario is: $\lambda_0=0.5\mu$, $\lambda_1=1.2\mu$, $T_1=100$ s, $T_2=400$ s, $T_3=700$ s, $T_4=1600$ s, $T_5=1900$ s. No turning movements are considered along the arterial.

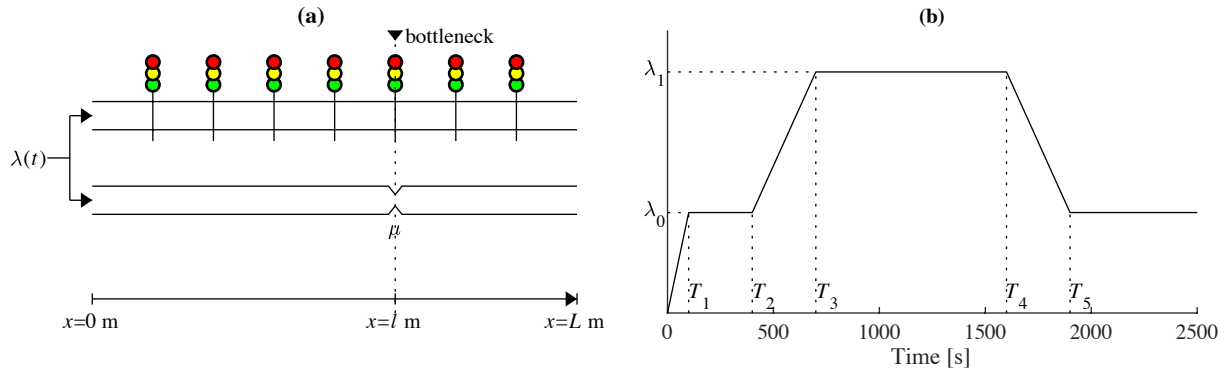


Figure 1: Description of the case study - (a) sketch of the arterial, (b) demand profile.

Let us set $x=0$ m at the arterial entry. The bottleneck location is $x=l=675$ m and the arterial exit location is $x=L=1080$ m, see Figure 1a. The solutions of the LWR model are characterized by the local density $k(x,t)$ and flow $q(x,t)$ values. The aggregated variables $K(t)$ and $Q(t)$ represent the network mean density and flow in the arterial when the arterial is considered as a single reservoir:

$$K(t) = \frac{1}{L} \int_0^L k(x,t) dx; \quad Q(t) = \frac{1}{L} \int_0^L q(x,t) dx \quad \text{Eq. (1)}$$

The relation between Q and K corresponds to the p-MFD. The outflow $Q_{out}(t)$ of the arterial is simply given by $q(L,t)$ as there are no internal trip endings. The relation between Q_{out} and K defines the o-MFD. Note that we are not resorting here to the classical steady-state relationship between the outflow and the mean flow (Daganzo, 2007), i.e. $Q_{out}=(L_{net}/L_{trip})Q$ where L_{net} is the network total length and L_{trip} is the

mean trip length of vehicles inside the reservoir. For an arterial, as these two lengths are identical, Q_{out} should have then been equal to Q . We will show later that this condition does not hold during transitional phases when aggregating the local traffic conditions over the arterial.

1.2. Analytical results without traffic signals

As a first step, the arterial is further simplified by removing all traffic signals and substituting the triangular fundamental diagram by a trapezoidal one with the capacity μ_s . The bottleneck at $x=l$ corresponds to a capacity reduction equal to μ , see Figure 1a. This ensures full analytical tractability of the LWR solutions and permits to study the impacts of the demand peak. To simplify the calculation, we are using the following Newell's transformations (Newell, 1993): $x'=x$; $t'=t-(x/u)$; $k'=k-(q/u)$; $q'=q$. These transformations lead to an asynchronous time coordinate system (Daganzo, 2003), as the time coordinate at any location is measured relative to the passage time of a reference vehicle travelling at the free-flow speed. They are also known as the "shearing transformation" in analytical geometry. Such transformations take advantage of symmetries in the kinematic wave model (Laval and Chilukuri, 2016) to guarantee that the problem expressed in (x',t',k',q') is fully equivalent to the original one while being much simpler to solve as (i) all free-flow densities are equal to 0 and (ii) $u'=+\infty$. This means that calculating the solutions in the asynchronous time frame is equivalent to calculating the solutions in a regular frame with a free-flow set to $+\infty$.

The fundamental diagram in the asynchronous time coordinates is presented in Figure 2b. The flow values at $x'=0$ are defined by the demand $q'(0,t')=\lambda(t')$. Let t'_0 be the time when the demand is exceeding μ . A shockwave is starting at $x'=l$ at time t'_0 . Let $X(t')$ be the shockwave position with respect to time t' . The Rankine-Hugoniot condition defines its derivative $X_i(t')$ at any time when a congestion exists:

$$X_i(t') = \frac{\mu - \lambda(t')}{k'_\mu} \quad \text{Eq. (2)}$$

Where k'_μ is the congested density related to a flow μ , see Figure 2b. The shockwave is going to propagate backwards during the congestion onset and then forward during the congestion offset. It vanishes as soon $X(t')=l$ again. $X(t')$ is arbitrary set to $+\infty$ when no shockwave is present on the arterial, i.e. when $t' < t'_0$ and after the congestion disappears. The solutions of the LWR model in the asynchronous time frame are:

$$\begin{cases} k'(x',t') = 0 \text{ and } q'(x',t') = \lambda(t') \text{ if } 0 \leq x' \leq X(t') \\ k'(x',t') = k'_\mu \text{ and } q'(x',t') = \mu \text{ if } X(t') \leq x' \leq l \\ k'(x',t') = 0 \text{ and } q'(x',t') = \mu \text{ if } l \leq x' \leq L \end{cases} \quad \text{Eq. (3)}$$

Figure 2a presents the analytical solution in the asynchronous time frame (x',t') . The density is equal to 0 everywhere but inside the congestion where it is equal to k'_μ . The flow is equal to $\lambda(t')$ everywhere in space when there is no shockwave or upstream of the shockwave and equal to μ downstream of the shockwave. The derivation of the spatial integration of the local solutions in the asynchronous time frame is then very simple. The aggregate variables $K'(t')$ and $Q'(t')$ are equal to:

$$\begin{cases} K'(t') = \frac{(l - X(t'))k'_\mu}{L} \\ Q'(t') = \frac{1}{L}(X(t')\lambda(t') + (L - X(t'))\mu) = \mu + (\lambda(t') - \mu)\frac{X(t')}{L} \end{cases} \quad \text{Eq. (4)}$$

Interestingly, it comes the following direct relationship between $K'(t')$ and $Q'(t')$ that describes the MFD pattern in the asynchronous time coordinates:

$$Q'(t') = \frac{(L-l)\mu + l\lambda(t')}{L} + \frac{(\mu - \lambda(t'))}{k'_\mu} K'(t') \quad \text{Eq. (5)}$$

During the arterial loading, this relation is linear as soon as $\lambda(t)$ reaches λ_1 . The values of $K'(t')$ vary between 0 and $(l/L)k'_\mu$ and the values of $Q'(t')$ are between μ and $(L-l)\mu/L+l\lambda_1/L$. The relation (5) is also linear during the arterial recovery as soon as $\lambda(t)$ reaches λ_0 . In the latter case, the values of $K'(t')$ have the same range as during the loading while the values of $Q'(t')$ are now between $(L-l)\mu/L+l\lambda_0/L$ and μ . Figure 2b presents the MFD pattern in the asynchronous time coordinates. This pattern corresponds to a clockwise hysteresis loop.

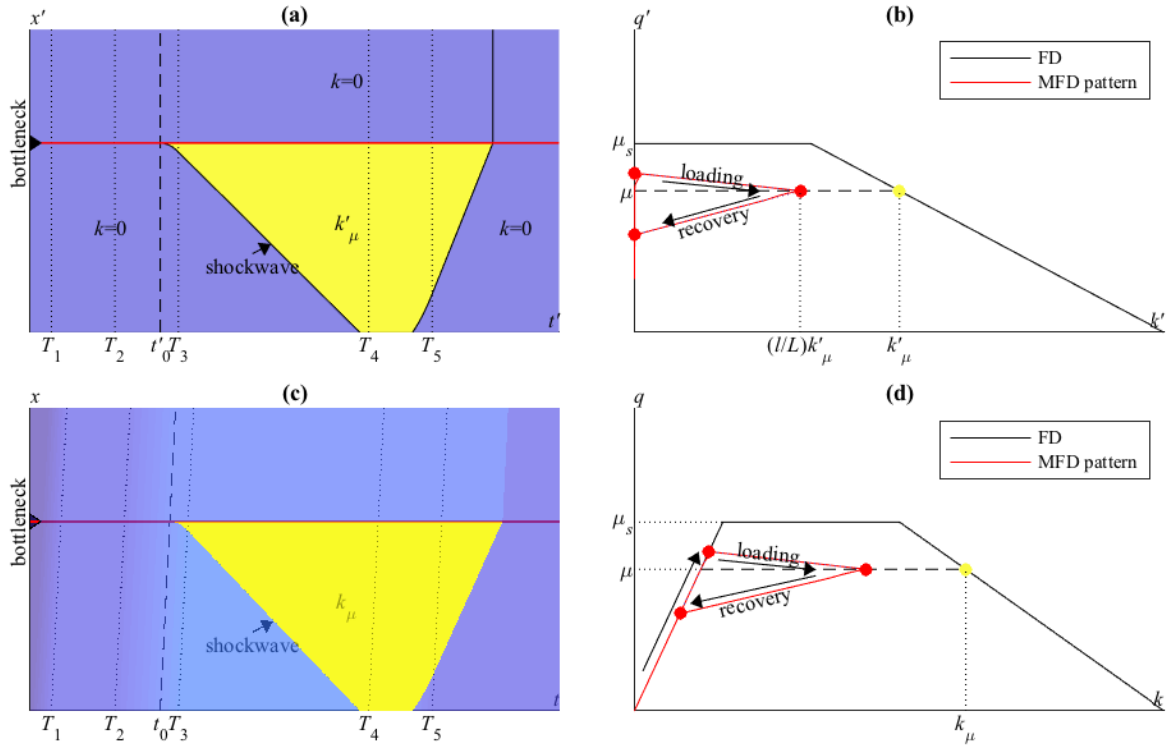


Figure 2: Analytical solution for the arterial without signals and a bottleneck - (a) LWR solution in Newell's moving coordinates, (b) FD and p-MFD pattern for the moving coordinates, (c) LWR solution in the regular coordinates, (d) FD and p-MFD pattern for the regular coordinates.

By applying the inverse Newell's transformations, it is possible to calculate the LWR solutions in the original coordinates (x, t) , see Figure 2c. Based on these solutions, we can also calculate the MFD pattern. It has the same shape as in the asynchronous time frame because the Newell's transformations are linear (Laval and Chilukuri, 2016), see Figure 2d. In particular, the relation between $Q(t)$ and $K(t)$ is linear when $\lambda(t)$ reaches λ_1 during the congestion onset and when $\lambda(t)$ reaches λ_0 during the offset. Note that in Figure 2d the relation looks almost always linear during the offset of congestion. It is because the congestion spillbacks upstream $x=0$ and the demand is then limited to μ for a while. In such a case, the MFD pattern during recovery starts with a linear trend and has only a slight deviation when the inflow is between μ and λ_0 .

This first result is very important as it shows that the clockwise hysteresis pattern can be simply driven by the local traffic dynamics at bottlenecks. Queue formation and discharge are not symmetrical processes as local wave propagation is different during the arterial loading and recovery. Here, the clockwise hysteresis loop arises from congestion dynamics without implying heterogeneous network loading due to unbalanced flows at intersections. Furthermore, it clearly appears here that the hysteresis shape (loop amplitude) is highly related to the demand profile, see eq. (5). To further illustrate this last point, Figure 3a presents different hysteresis loops corresponding to different demand scenarios with identical time intervals but different λ_0 and λ_1 values. It is not surprising from eq. (5) that the higher λ_1 (and/or the lower λ_0) results in the bigger hysteresis loop. Note that when $\lambda_0/\mu=0.5$ and $\lambda_0/\mu=1.1$, the congestion does not go up to the arterial entry. This is why the hysteresis pattern does not touch the right-end red dot, which corresponds to the situation where the arterial is fully congested upstream of the bottleneck.

Figure 3b shows the hysteresis loop when λ_0 and λ_1 are set again to the reference values but when the duration ΔT of the demand increase (T_3-T_2) and decrease (T_5-T_4) now varies from 50 s to 2100 s. The main conclusion here is that the changes in the MFD pattern are rather minimal. Larger ΔT values are smoothing the left-hand corner of the hysteresis loop during loading because the congestion starts as soon as $\lambda(t) > \mu$ and reduces the flow upstream of the bottleneck when the inflow is increasing slower. The highest mean flow values are observed when the demand increase is the sharpest. In that case, high flow values can enter the arterials before reaching the congested part where the flow is limited to μ . Similar conclusions can be drawn during the recovery phase but with an even lower amplitude.

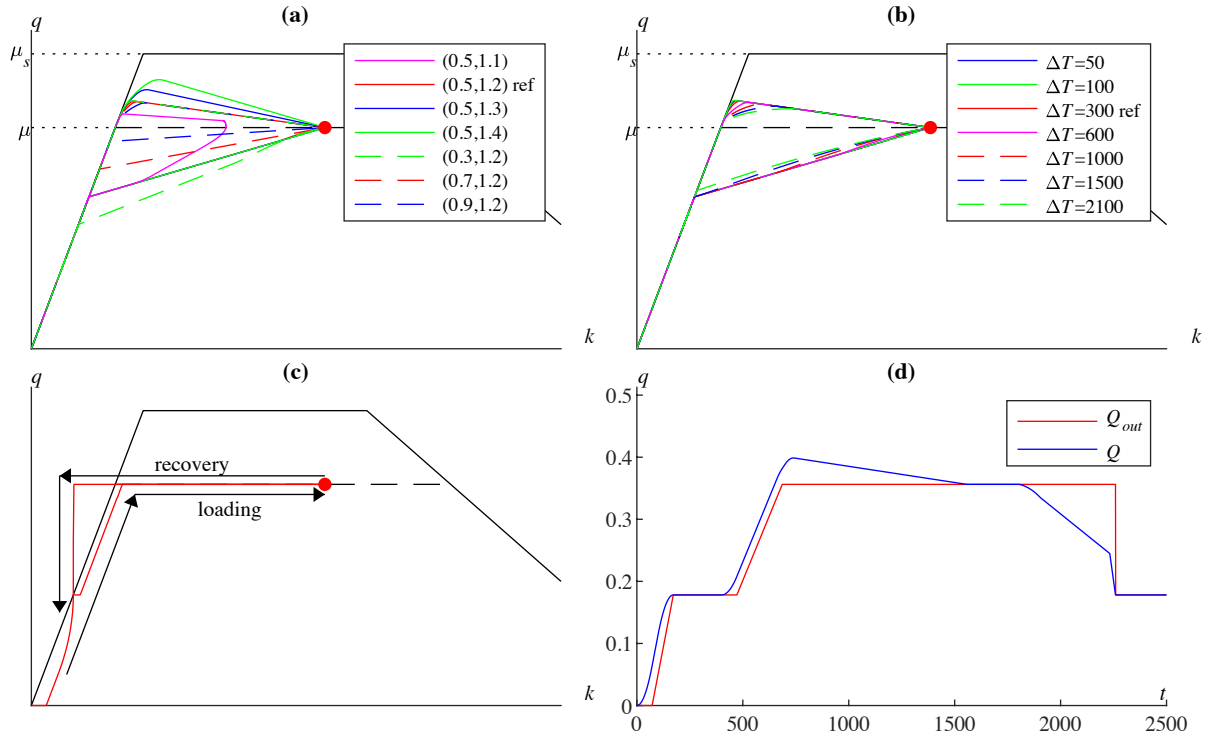


Figure 3: Hysteresis loops - (a) p-MFD clockwise hysteresis loops for different $(\lambda_0/\mu, \lambda_1/\mu)$ values (b) p-MFD clockwise hysteresis loops depending on the demand profile sharpness ΔT (c) o-MFD counter-clockwise hysteresis loops (d) Time-evolution of Q and Q_{out} for the reference scenario

Figure 3c shows now the o-MFD pattern for the reference demand scenario (relationship between Q_{out} and K). This pattern clearly corresponds to a counterclockwise hysteresis loop. This is a significant finding as such a pattern has never been highlighted for o-MFD to the authors best knowledge. The primary reason is that during the network recovery, the maximal capacity μ is maintained at the arterial exit not only until the congestion disappears upstream of the bottleneck, but even longer until all vehicles that have experienced the congestion leave the reservoir. This includes all vehicles that are driving downstream of the bottleneck when the traffic state corresponds to a flow value μ , see Figure 2b. This means that at the very last moment when the maximal capacity μ is observed at the exit, the density inside the reservoir is fully free-flow and equal to λ_0/μ . In short, because the exit is downstream of the bottleneck, the outflow sustains to the discharge rate μ during recovery until all internal effects of congestion have vanished. Note that the o-MFD pattern looks more stable than the p-MFD because it does not depend on λ_1 (the maximal outflow is limited to μ) and only depends on λ_0 during congestion offset to define the density where the outflow is dropping to the free-flow curve after being maintained stable and equal to μ . Of course, we are working here in the idealized case of an urban network reduced to a simple arterial but it makes sense that such a pattern should also appears for more complex network situations where the demand exceeds the internal network capacity during peak hour because the physical underlying mechanisms are very similar. More evidences will be provided in section 3.

Finally, we can test here the validity of the steady-state relationship between the mean flow Q and the outflow Q_{out} at the reservoir level, i.e. $Q=Q_{out}$ in the arterial case (Daganzo, 2007). Figure 3d presents the time-evolution of Q and Q_{out} for the reference demand scenario. The equality between the two clearly

holds as soon as the arterial is in a steady-state, i.e. when $170 \leq t \leq 400$ s, $1560 \leq t \leq 1820$ s and $t > 2280$ s. However, several discrepancies appear during transitional states. During the initial loading of the network ($\lambda(t) \rightarrow \lambda_0$), the mean flow starts to increase sooner while the outflow increases only after the first vehicle has fully crossed the arterial. Such a lag can be observed whenever the demand is changing at the entry. During the second phase of the loading ($\lambda(t) \rightarrow \lambda_1$) the mean flow value can exceed the bottleneck capacity μ during the time period when the congestion is spreading backwards up to the arterial entry, while the outflow is immediately limited to μ because the exit is downstream of the bottleneck. On the contrary, during the network recovery, the mean flow starts decreasing sooner as it takes benefit from congestion vanishing upstream of the bottleneck while the outflow is maintained to maximum. All these observations about outflow evolution will be very useful in section 2 when benchmarking the MFD models and discussing their specifications.

1.3. Exact numerical results with traffic signals

In this section, we are now considering explicitly the effect of traffic signals. Deriving the analytical solution of the LWR model becomes too cumbersome in this case. Fortunately, we can resort to exact numerical methods based on the variational theory (Daganzo, 2005a; 2005b; Laval and Leclercq, 2013) because the fundamental diagram is triangular. Here, we define a numerical grid with $\Delta x = 15$ m and $\Delta t = 1$ s. The value of the cumulative number of vehicles $N(x, t)$ at each grid node is given by:

$$N(x, t) = \min(N(x - \Delta x, t - \Delta t), N(x - \Delta x, t - k\Delta t) + \kappa\Delta x, N(x, t - \Delta t) + \theta\Delta t) \quad \text{Eq. (6)}$$

where $k = u/w$ and $\theta = 0$ if a signal is red during $t - \Delta t$ and t at location x or $\theta = S$ (the maximal capacity) otherwise. The solutions are exact when k is an integer (Daganzo, 2005a), which is true here. Local flow and density values are equal to $q(x, t) = (N(x, t) - N(x, t - \Delta t)) / \Delta t$ and $k(x, t) = (N(x, t) - N(x - \Delta x, t)) / \Delta x$. Macroscopic variables K , Q and Q_{out} are further averaged over a rolling time-horizon of 60 s, i.e. a cycle signal, to filter traffic oscillations generated by traffic signals and the rotation between green and red phases.

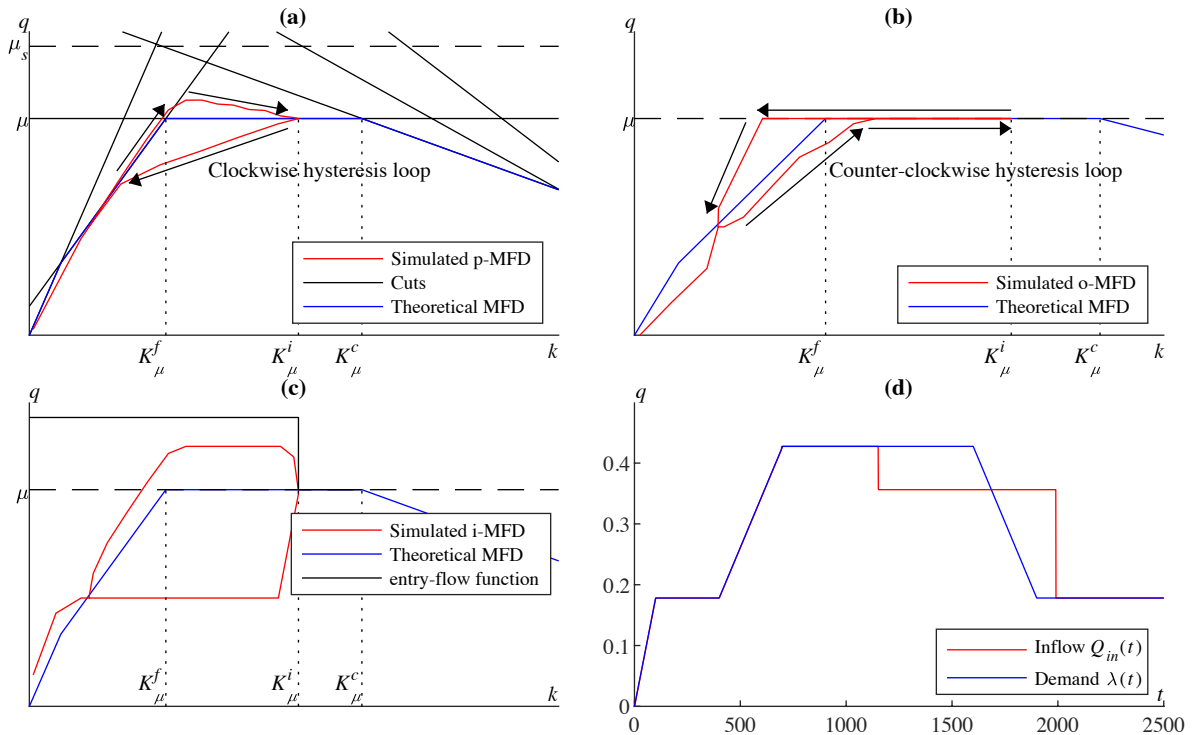


Figure 4: Arterial with traffic signals - (a) p-MFD pattern compared to the theoretical MFD (cuts method) (b) o-MFD pattern (c) inflow-MFD pattern and related entry flow function (d) inflow vs. demand time-profile.

Figure 4 presents the simulation results for the reference demand scenario. Figure 4a exhibits a clockwise hysteresis pattern for the p-MFD, while Figure 4b show a counter-clockwise loop for the o-

MFD. This confirms the results obtained in the previous section without considering the traffic signals. In fact, the MFD patterns both in mean-flow and outflow are very similar to the solutions obtained in Figure 2d and Figure 3c. This is not surprising as the effects of traffic signals are mostly visible on a short time scale, i.e. within a signal cycle, while traffic dynamics in the arterial at larger time scale is driven by bottlenecks capacity and congestion spillbacks. Such effects have been already considered in the previous section in accordance with the signal capacities, i.e. μ_s and μ .

As we are studying an arterial with no turning movements at intersections, it is possible to analytically calculate the MFD corresponding to all stationary conditions using the cuts method (Daganzo and Geroliminis, 2008). This method is based on multiple moving observers that are travelling along the arterial with positive and negative mean speeds v_r while monitoring the mean overtaking flow q_r . Each line $Q=q_r+Kv_r$ defines a cut in the (K,Q) plane, which represents an upper-bound for the MFD. (Leclercq *et al.*, 2013) proposes a solution method based on sufficient variational graphs to determine the cuts that define the lower hull of all possible cuts and demonstrates that this hull is a tight estimation of the MFD. This method has been applied here and the results are shown in Figure 4a, see the cuts in black and the hull in blue. It appears that the p-MFD pattern fits very well with the theoretical MFD in free-flow during the onset of congestion. However, when traffic conditions are saturated (density values between K_{μ}^f and K_{μ}^c in Figure 4a), the p-MFD exhibits a clockwise hysteresis loop that is not materialized on the theoretical MFD. This result can be perceived as surprising because the cut method is also providing the exact average of the LWR solutions. The explanation is all about how network loading is performed. The cut-method assumes that every MFD state (K,Q) is stationary with an uniform loading and a constant capacity at the exit. By definition, this method does not consider transitional states during when the congestion or more generally traffic waves are propagating inside the arterial. The simulation results are obtained based on a dynamic network loading with a trapezoidal shape and thus the hysteresis pattern is clearly driven by transitional phases. This comparison highlights a key assumption of the MFD theory: stationary states inside the reservoir. Such an assumption can only be matched with slow varying demand profile. However, Figure 3b shows that in the case of an arterial even very large ΔT values like 2100 s (time duration of the switch from λ_0 to λ_1) lead to p-MFD patterns far from the stationary cut with a flow value μ that is observed in the theoretical MFD. This means that when considering a network loading and recovery during a peak hour, the observed MFD pattern can be quite far from the theoretical one, in particular, when the network reaches saturation. Furthermore, the effective pattern depends on the network loading, i.e. the demand, which contradicts one of the seminal conjecture of the MFD theory. We will later see the impacts of such observations on MFD models and simulation.

Finally, Figure 4c presents the evolution of the inflow $Q_{in}(t)=q(0,t)$ with respect to the mean density in the arterial. This inflow-MFD function has received little attention in the MFD literature as the effective inflow is usually defined by the demand. This is true here, i.e. $Q_{in}(t)=\lambda(t)$, except during the time period when the congestion has reached the entry, see Figure 4d. In this latter case, we have $Q_{in}(t)=\mu$. These observations permit to define the entry-flow function $I(K)$ in this case, i.e. the maximal flow that can enter the reservoir depending on its current traffic conditions K . This corresponds to the reservoir supply. (Hajiahmadi *et al.*, 2013; Knoop and Hoogendoorn, 2014; Lentzakis *et al.*, 2016) consider that this function should be limited to the MFD capacity (here μ) for free-flow and saturated conditions, i.e. $K < K_{\mu}^c$. It is clear that the inflow can be higher than μ , when $K < K_{\mu}^i$. K_{μ}^i is the mean density value when the flow is equal to μ all along the arterials with congestion upstream of the bottleneck and free-flow downstream. The only limitation to the inflow when $K < K_{\mu}^i$ is the capacity of the first signal, i.e. μ_s . Figure 4c summarizes the $I(K)$ shape for the arterial case. This is consistent with previous observations from (Geroliminis and Daganzo, 2007) for a more complex network setting (microscopic simulation of downtown San Francisco area). Further investigations are clearly needed to better characterized entry-flow functions but for now we will use this setting for MFD simulation in the next section.

2. Capabilities of MFD models to reproduce aggregate arterial dynamics

In this section, we are going to investigate how well the accumulation- and trip-based MFD models reproduce traffic dynamics for the arterial case (with active signals) when compared to the LWR solutions. This will permit to clarify the limitations of the different approaches and to study how they

can be addressed. A special attention will be put on the consequences of the previous observations about hysteresis loops. In the next section, more complex urban settings will be presented to (i) confirm our conclusions and (ii) demonstrate that the arterial case was indeed a relevant proxy to study the discrepancies between MFD models and the aggregation of local traffic dynamics under fast-varying demand profiles.

2.1. Expressions of the different existing MFD models

The seminal and most familiar MFD model considers that traffic dynamics, i.e. the time-evolution of the number of vehicles $n(t)=K(t)L$, inside a region or reservoir is described by the following conservation equation (Daganzo, 2007):

$$\frac{dn(t)}{dt} = f_{in}(t) - f_{out}(t) \quad \text{Eq. (7)}$$

where $f_{in}(t)$ and $f_{out}(t)$ are the inflow and the outflow of the reservoir, respectively. This model is referred in the literature as the accumulation-based, the bathtub or the PL (production over trip length) model. The last name comes from the existence of the p-MFD Q and the steady-state approximation that defines the outflow as the ratio of the travel production P by the mean trip length L_{trip} , i.e. $f_{out}(t)=P/L_{trip}$. In the arterial case, P is equal to the product of the mean flow Q and the network length L , also equal to the L_{trip} . So, it comes that $f_{out}(t)=Q(K(t))$. Note that the accumulation-based model would be better defined using the o-MFD instead of the p-MFD, i.e. $f_{out}(t)=Q_{out}(K(t))$ because the steady-state condition is no longer required as long as the o-MFD is well-defined. The main reason why the p-MFD is usually preferred in the literature is because the o-MFD is hardly observable in practice.

We are considering here that no capacity restriction happens at the exit of the arterial, so the outflow is only defined by the internal traffic state $K(t)$ or $n(t)$. The inflow is, most of the time, equal to the demand but sometimes should be limited by the entry-flow function as explained in the previous section: $f_{in}(t)=\min(\lambda(t),I(K(t)))$. Numerical resolution of eq. (7) is straightforward when considering a time-step Δt :

$$n(t + \Delta t) = n(t) + \left(\min(\lambda(t), I(K(t))) - Q(K(t)) \right) \Delta t \quad \text{Eq. (8)}$$

The trip-based MFD formulation has been first introduced by (Arnott, 2013) and then further investigated in (Lamotte and Geroliminis, 2017; Mariotte *et al.*, 2017). The key principle is that all vehicles inside the reservoir share the same speed $V(t)=Q(t)/K(t)$ at the same time t . Let $N(t)$ be the vehicle that enters the reservoir at time t . $N(t)$ is defined continuously as the integral of the total inflow (cumulative vehicle function). The travel time $T(N(t))$ of vehicle $N(t)$ is determined by integrating the mean reservoir speed V until the vehicle has reached its targeted travel distance L :

$$\int_t^{t+T(N(t))} V(u) du = L \quad \text{Eq. (9)}$$

The mean speed V is directly defined by the p-MFD as $V=Q(K)/K$. Note that $n(t)$ is no longer the main variable of the model but a by-product. It can be calculated by the difference between $N(t)$ and $N_{out}(t)$, where $N_{out}(t)$ is the cumulative outflow function. Again, there is no restriction at the reservoir exit but we may delay the entry of some vehicles to fulfill the limitation given by the entry-flow function $I(K)$.

(Lamotte and Geroliminis, 2017; Mariotte *et al.*, 2017) propose a simple numerical method to solve eq. (9) when the inflow is discretized into vehicles. Each vehicle entry or exit generates an event that triggers an update of the number of vehicle inside the reservoir. Between two events, this number is constant and so is the speed of all vehicles. It suffices to track the remaining distance to travel inside the reservoir for all vehicles to know the next vehicle to exit and when it should do so.

While being both based on the MFD concept, the accumulation- and trip-based formulations are fundamentally different. In particular, the trip-based explicitly considers the delay (travel-time) between the inflow and outflow boundaries, while the accumulation-based does not. Both models do not produce the same solutions as we will see in the following sections.

2.2. The accumulation-based model

Let us first compare the results from the accumulation-based model and the LWR solutions. To determine the MFD shape, two main settings are considered. The first is the theoretical p-MFD and the second is the observed p-MFD through the LWR simulation, see Figure 4a. Note that the first one has a well-defined unimodal shape while the second exhibits a clockwise hysteresis loop. The upper-branch Q_{up} characterizes network loading while the lower-branch Q_{down} is for recovery. We have to define a criterion to identify one case from the other. We choose to track the number of vehicles in the reservoir and to determine if it is currently increasing or decreasing based on the following hysteresis trigger:

$$\begin{cases} \text{if } n(t) > n_{ref} + \Delta n \text{ then } Q = Q_{up} \ \& \ n_{ref} = n(t) \\ \text{if } n(t) < n_{ref} + \Delta n \text{ then } Q = Q_{down} \ \& \ n_{ref} = n(t) \end{cases} \quad \text{Eq. (10)}$$

In practice, $n_{ref}=0$ when $t=0$ and we set Δn to 2. Figure 5 shows all results from the comparisons between accumulation-based MFD and LWR solutions. Let us first focus on the results produced with the theoretical p-MFD (red vs. black curves). In general, the MFD and LWR results look very close for all variables under consideration: numbers of vehicle n (Figure 5a), mean reservoir speed V (Figure 5b), outflow f_{out} (Figure 5c) and experienced travel time TT (Figure 5d). This last variable is derived from the cumulative curves at entry and exit. Note that the LWR solution is calculated every second without filtering for n and TT and with a 60 s symmetric rolling averaging filtering for V and f_{out} . A closer look at the MFD results confirms previous analysis about the accumulation-based formulation in the literature, e.g. (Mariotte *et al.*, 2017; Mariotte and Leclercq, 2018). First, in free-flow (here $t < 1000$ s), we observe immediate reactions in outflow to inflow (demand) variations (no-propagation delay). This is very noticeable at the beginning of the simulation, where the LWR outflow only starts increasing after 70 s, while the MFD outflow increases immediately for the accumulation-based model. This induces wrong estimation of the experienced travel times. When the demand starts increasing again at time $t=T_2$, see Figure 1b, the first reaction of the MFD model is a drop in experienced travel times. This should not happen as the network is currently loading. All these phenomena have been well-documented in (Mariotte *et al.*, 2017) and hence, details are omitted here. What is more important is to highlight that as soon as saturation starts, i.e. when $K > K_{\mu}^f$ (see Figure 5e), the MFD and LWR results are matching very well for all four variables. During the network recovery, the results are also very close except that the MFD model stops the outflow being equal to μ too soon and then exhibits smooth dropping profiles on the contrary to the LWR sharp ones (in particular for the outflow and the number of vehicles). The first point is because the outflow is following the unimodal pattern given by the theoretical p-MFD instead of the counter-clockwise pattern that corresponds to the LWR solution, see Figure 4b. The second point is related to propagation issues between inflow and outflow like during free-flow situations. These two points are not creating inconsistent experienced travel-time predictions but lead to inaccurate estimation of the mean speed during the arterial recovery, see $1900 < t < 2200$ s in Figure 5b.

The main conclusion when comparing the time-evolution of the MFD and LWR results for n , V , f_{out} and TT is that the accumulation-based is performing well with a theoretical p-MFD setting except in free-flow conditions. However, it should be noted that the p-MFD (Figure 5e) and o-MFD (Figure 5f) patterns resulting from the MFD simulation are not in accordance with the aggregated observations coming from the LWR solution, see section 1.3. It is then interesting to consider running the MFD simulation using directly the p-MFD with the hysteresis loop. A modeler would use such a setting if he/she calibrates the MFD shape based on daily empirical observations of the p-MFD. Now, we are comparing the blue and the black curves in Figure 5. It immediately appears that if the p-MFD pattern now matches perfectly the LWR solution, see Figure 5e, the MFD results for n , V , f_{out} and TT are significantly deteriorated. This may look strange at first glance but can be explained. As mentioned earlier, the accumulation-based equation is based on the o-MFD and not the p-MFD. During transitional phases, the approximation that relates the p-MFD and the o-MFD by a constant factor (the inverse of the mean trip length) does not hold as it requires steady state. In the arterial case, the theoretical p-MFD leads to a closer approximation of the o-MFD corresponding to the aggregation of the LWR simulation, see Figure 5e. Even if the theoretical o-MFD is missing the counter-clockwise hysteresis part, it properly adjusts the outflow to μ during the saturation (end of the arterial loading) and the beginning of the arterial recovery. This explains why the MFD results are so good during this period. On the contrary,

the p-MFD with hysteresis leads to a simulated o-MFD with exactly the same hysteresis pattern as $Q_{out}=Q$. This is completely unrealistic for the arterial case, see Figure 5e. The MFD simulation results are then not good in the second case.

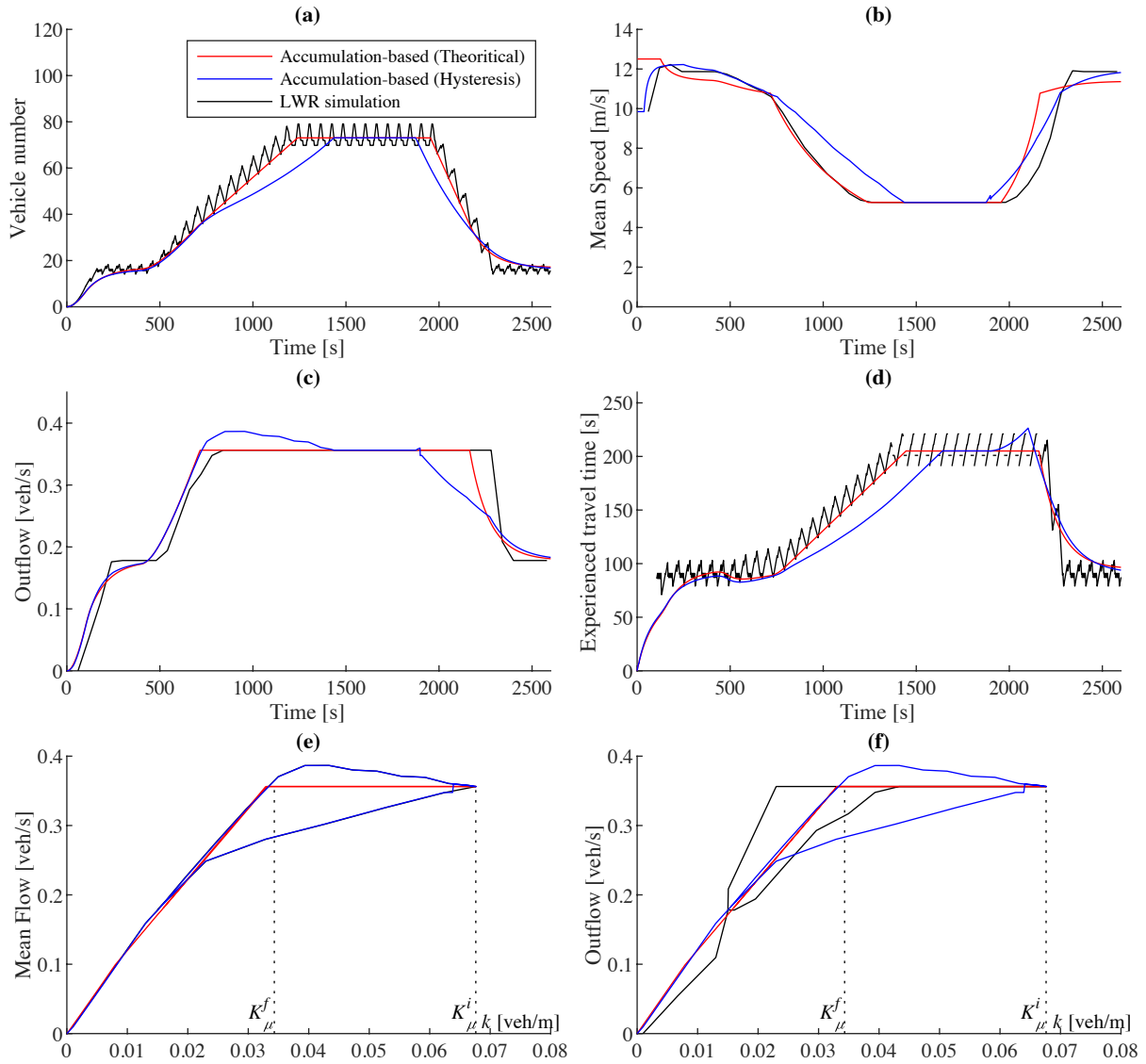


Figure 5: Comparison between accumulation-based MFD and LWR solutions - (a) Time-evolution of the number of vehicles inside the reservoir (b) Time-evolution of the mean speed (c) Time-evolution of the outflow (d) Time-evolution of the experienced travel-time (e) p-MFD (f) o-MFD.

In conclusion here, the accumulation-based model with a theoretical p-MFD provides good results except in free-flow but this is because of the combination of two significant flaws that compensate each other: (i) the theoretical p-MFD is not a good approximation of the observed p-MFD based on the LWR simulation and (ii) the steady-state relation between mean flow and outflow is not valid during transitional phases. The most important conclusion here is that the accumulation-based model should definitely better be fed with a good estimation of the o-MFD. In particular it is very important to have a proper estimation of the outflow capacity. We could have further improved the results of the MFD simulation by testing a last case where the o-MFD corresponds to the counter-clockwise hysteresis pattern in Figure 4b. This would have improved the time-evolution of n , V , f_{out} and TT at the end of the recovery period because the outflow capacity would have been sustained longer as observed in the LWR solution. We are not showing the results here because (i) the results the theoretical p-MFD are already quite good and (ii) we have no means in practice to properly estimate the full o-MFD pattern except in this very simple case.

2.3. The trip-based model

Figure 6 compares the trip-based results with the LWR solution based on exactly the same outputs as for the accumulation-based model. In this section, we are going to test both the theoretical p-MFD shape and the p-MFD with hysteresis. This latter setting makes much more sense here as the trip-based model is directly fed by the p-MFD in speed. The green curves in Figure 6 correspond to the theoretical p-MFD case. It appears that the steady state corresponding to the complete saturation of the arterial at the end of the loading phase is well reproduced by the trip-based model. This is because both outflow and inflow are well set to μ when $K > K^i_\mu$, which means that (i) the entry-flow function is acting properly and (ii) the steady-state predicted by the trip-based model is in accordance with the LWR solution. On the contrary, we observe large discrepancies between the trip-based and the LWR models for all four variables n , V , f_{out} and TT during the network loading and recovery, i.e. when the demand is fast varying. Let us first focus on the recovery phase when $t > 2000$ s. The most noticeable wrong doing of the trip-based model is the jump in the outflow between 2050 and 2200 s. It largely exceeds the reservoir capacity. The explanation comes from the *causality effect* that has been well-documented for exit-flow function models (Merchant and Nemhauser, 1978; Friesz, 1989) and classical link flow theory. When the inflow is decreasing at the entry of the reservoir during recovery, this reduces the number of vehicles inside the reservoir and then immediately increases the speed of all internal vehicles by construction of the trip-based model, see eq. (9). At the beginning of the network recovery a lot of vehicles are in the reservoir and the increase of speed applies to a large number of vehicles. This suddenly highly increases the trip-completion rate and then generates the temporary jump in outflow observed in Figure 6c. It should be noticed that even if the outflow values are clearly inconsistent during this period, this has less impacts on the other variables. The travel time is dropping too soon compared to the LWR solution but has a regular and consistent pattern. The mean speed is increasing too soon as well and is significantly overestimated between 2050 s and 2200 s but again, the trend is consistent. Finally, the accumulation is following the trend of the LWR solution.

The fact that the trip-based model suffers from the causality effect starts being discussed in the MFD literature but no clear evaluation of the impacts on the simulation has been proposed yet. The arterial case presented here is a first attempt in this direction. It should be acknowledged that it magnifies the causality effect because it corresponds to a unidirectional flow where entering (upstream) vehicles should clearly not influence exiting (downstream) ones because of traffic anisotropy. For more general reservoir settings, it is no longer possible to clearly identify upstream and downstream locations as vehicles can enter and exit at the same place along the perimeter. However, if one imagines a reservoir with internal bottlenecks upstream can be defined as the part of the trips upstream of the bottlenecks and downstream the other part. Thus, the arterial setting looks like a proxy for more complex situations as far as fast demand loading and unloading are concerned. The jump in outflow at the beginning of the unloading phase is clearly non-admissible as the perimeter flow (outflow) cannot exceed by such a large extend the reservoir capacity μ defined by its internal bottleneck. Fortunately, a patch is quite easy to propose and implement. It suffices to bound the outflow to μ . When the bound is active, the travel distance is no longer consistent with the integral of the mean speed as we are delaying vehicle exits to fulfill the outflow constraint. The updated results considering this bound are presented in Figure 6, see green lines with circles. The outflow values are now much more consistent with the LWR solution during the recovery phase even if we can still see that the drop is delayed compared to the LWR solution. Such a delay is also noticeable on the accumulation and the travel time results. The mean speed values do not exhibit the same trend and are not much close to the LWR solution.

The patch on the maximal outflow values has improved the trip-based outputs during the recovery phase but has no influence during the loading where discrepancies with the LWR solutions are also visible. When looking at the time-evolution of the accumulation, it appears that the discrepancy starts around $t=700$ s when traffic conditions switch from free-flow to early saturation, i.e. when the density exceeds K^f_μ , see Figure 6e. Until the steady-state saturation state is reached ($K=K^i_\mu$), Figure 6b shows that the mean speed is not accurately estimated by the trip-based model, which explains the discrepancies. The inaccurate estimation of the mean speed comes from the theoretical p-MFD that is not considering the hysteresis loop, see Figure 6e. To improve the trip-based MFD accuracy, it seems then relevant to replace the theoretical p-MFD by the p-MFD with the hysteresis loop. The results are produced in blue

in Figure 6. The p-MFD with hysteresis immediately improves the trip-based simulation results during the loading but also to a lesser extent during the recovery. The mean speed estimations are much more accurate leading to a better time-dependent accumulation profile and a better estimation of the experienced travel times. However, some discrepancies remain between $900 < t < 1300$ s. They are very noticeable for travel time and outflow values. In particular, the outflow curve experiences a drop during this time period on the contrary to the LWR solution, which remains stable and equal to μ .

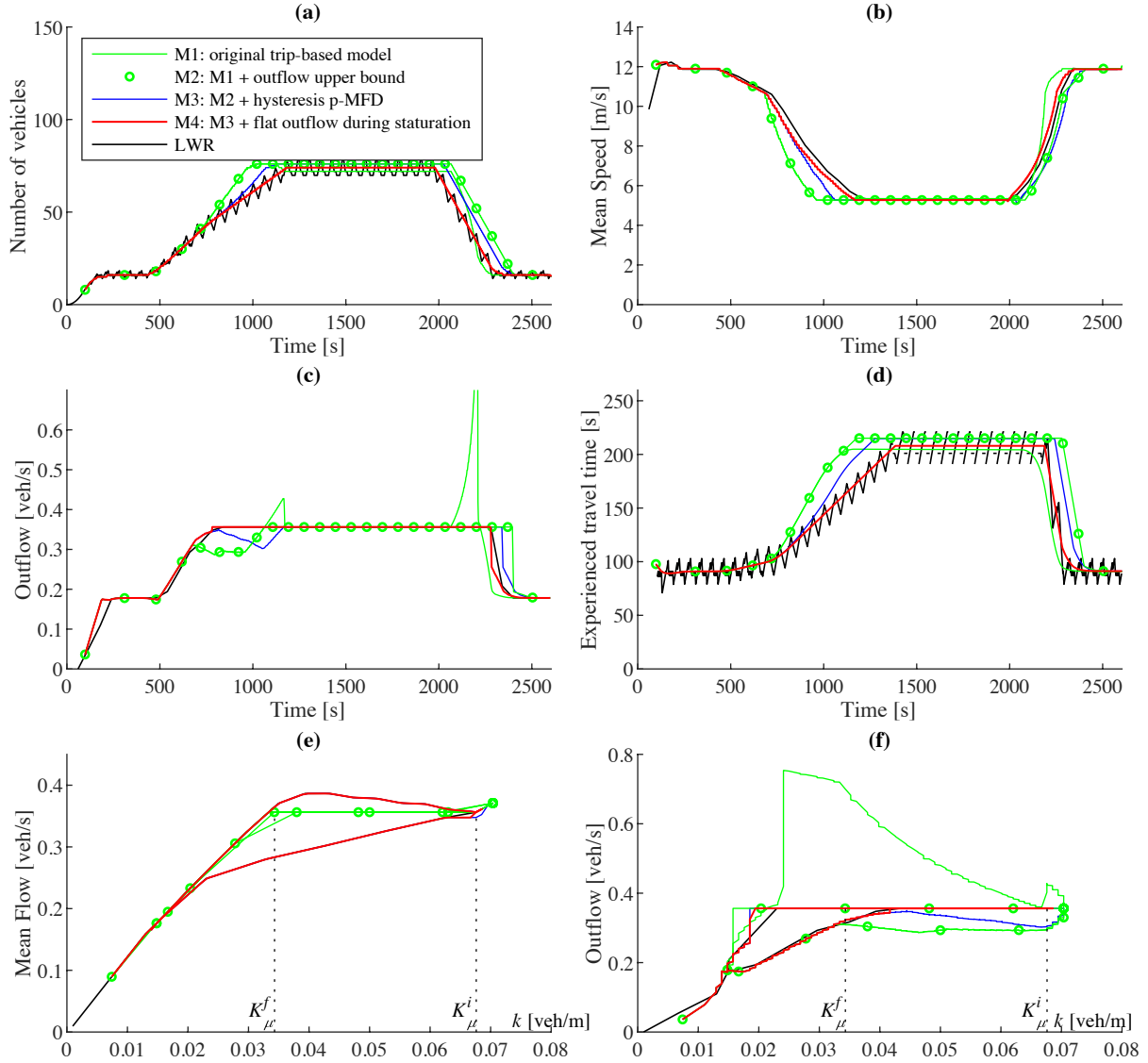


Figure 6: Comparison between trip-based MFD models and LWR solutions - (a) Time-evolution of the number of vehicles inside the reservoir (b) Time-evolution of the mean speed (c) Time-evolution of the outflow (d) Time-evolution of the experimented travel-time (e) p-MFD (f) o-MFD.

The explanation for the outflow drop at the end of the loading phase is again the causality effect. Here, the inflow is increasing and then the accumulation. The arrival of more vehicles into the reservoir leads to a speed reduction for all vehicles including those which are close to exit. When the accumulation is important, i.e. at the end of the loading phase when the reservoir is reaching saturation ($K > K_{\mu}^f$), lots of vehicles see such a speed decrease, which prevents them from exiting. This is exactly the reverse phenomenon compared to what we previously observe during the recovery phase but the concept is similar: traffic anisotropy is not fulfilled. A patch can be proposed here too. As soon as the reservoir reaches saturation ($K > K_{\mu}^f$), we know that the outflow should remain steady and equal to μ , so we can simply implement this condition. When this patch is active, again the integration of the mean speed over vehicle trips is no longer consistent with the travel distance as we are now speeding up the exit of some vehicles. The red curves in Figure 6 present the results of the trip-based model when the two patches

have been implemented and the p-MFD with hysteresis is used. The results match perfectly with the LWR solution. They outperform the results of the accumulation-based model because (i) they are accurate in free-flow in particular when considering outflows and travel times and (ii) the outflow decreases during network recovery is sharp just like the LWR solution.

Figure 6e and Figure 6f show the p-MFD and the o-MFD patterns resulting from the simulation, respectively. Regarding the p-MFD, it is clear that using the p-MFD with hysteresis makes the trip-based outputs match perfectly with the LWR ones. More interestingly, the simulated o-MFD pattern appears completely inconsistent when the maximal bound on the outflow is not implemented (green curve). The second patch helps the outflow to be accurate during the full saturation period ($K_{\mu}^f \leq K \leq K_{\mu}^f$). Finally, when both patches are implemented and the p-MFD is with hysteresis (red curve), the simulated o-MFD pattern is fully in accordance with the LWR simulation. In particular, it exhibits the same counter-clockwise hysteresis loop.

In conclusion, it appears here that the regular formulation of the trip-based model only provides valid results in free-flow conditions. For such traffic states, it outperforms the accumulation-based model as it better reproduces the experienced travel times and the time-lag between outflow responses and inflow variations. The first analysis about the trip-based model in the literature (Lamotte and Geroliminis, 2017; Mariotte *et al.*, 2017) mention the counter-clockwise hysteresis loop for the o-MFD as a very noticeable result. This study shows that a part of this pattern is due to unrealistic outflow values during the network recovery related to the causality effect, see the green curve in Figure 6f. This phenomenon is magnified by the arterial case, but the rational should apply for other network configurations because all vehicles have to travel a certain distance in the reservoir and can be speed-up by a sudden decrease of the total accumulation. In this section, we proposed two patches on the outflow to balance the causality effect during both network loading and recovery: (i) the outflow should be bound to the maximal capacity and (ii) the outflow should be sustained to the maximal capacity when saturation is reached, i.e. $K > K_{\mu}^f$. In the arterial case, these two patches solve most of the problems except the inaccurate speed estimations that are noticeable mainly during the network loading. It should be noted that adding these two patches changes the behavior of the trip-based model during saturation significantly. Basically, it is no longer a trip-based model per se as we are closely monitoring the outflow in such a situation, which means that the model is behaving just like the accumulation-based model. So, the extended trip-based model is more by design an hybrid MFD model that keeps the superiority of the trip-based formulation in free-flow while switching to an accumulation-based like formulation during saturation. Note that a similar switching process has been proposed in (Mariotte and Leclercq, 2018) to better account for congestion spillbacks between multiple reservoirs.

To obtain a perfect match between the trip-based and the LWR models, the p-MFD with hysteresis should be favored. On the contrary to what we observed with the accumulation-based model, this setting improves the simulation results because the trip-based formulation directly resorts to the p-MFD in speed. However, if accurate results cannot be obtained without the two patches on the maximal outflow capacity, a calibration based on the theoretical p-MFD instead of the p-MFD with hysteresis only leads to small discrepancies at the end of the loading and at the beginning of the unloading for n , V , f_{out} and TT . The final formulation of the trip-based model (p-MFD with hysteresis and the two patches) is able to perfectly reproduce the counter-clockwise pattern that is observed on the LWR solution for the o-MFD, see the red curve in Figure 6f. This pattern is less pronounced than the one produced by the original trip-based model but still it exists. It is much more realistic with the extended trip-based model because it always fulfills the reservoir maximal capacity. This pattern is directly triggered by the congestion propagation inside the arterial during the onset and offset of the reservoir as presented in section 1.

3. Further evidences for more complex urban network configurations

In this section, we propose further evidences about the clockwise p-MFD and counter-clockwise o-MFD hysteresis patterns by processing micro-simulation outputs for an urban district during the morning peak hour. Additional benchmarks are also proposed for the different MFD models and this more complex urban configuration.

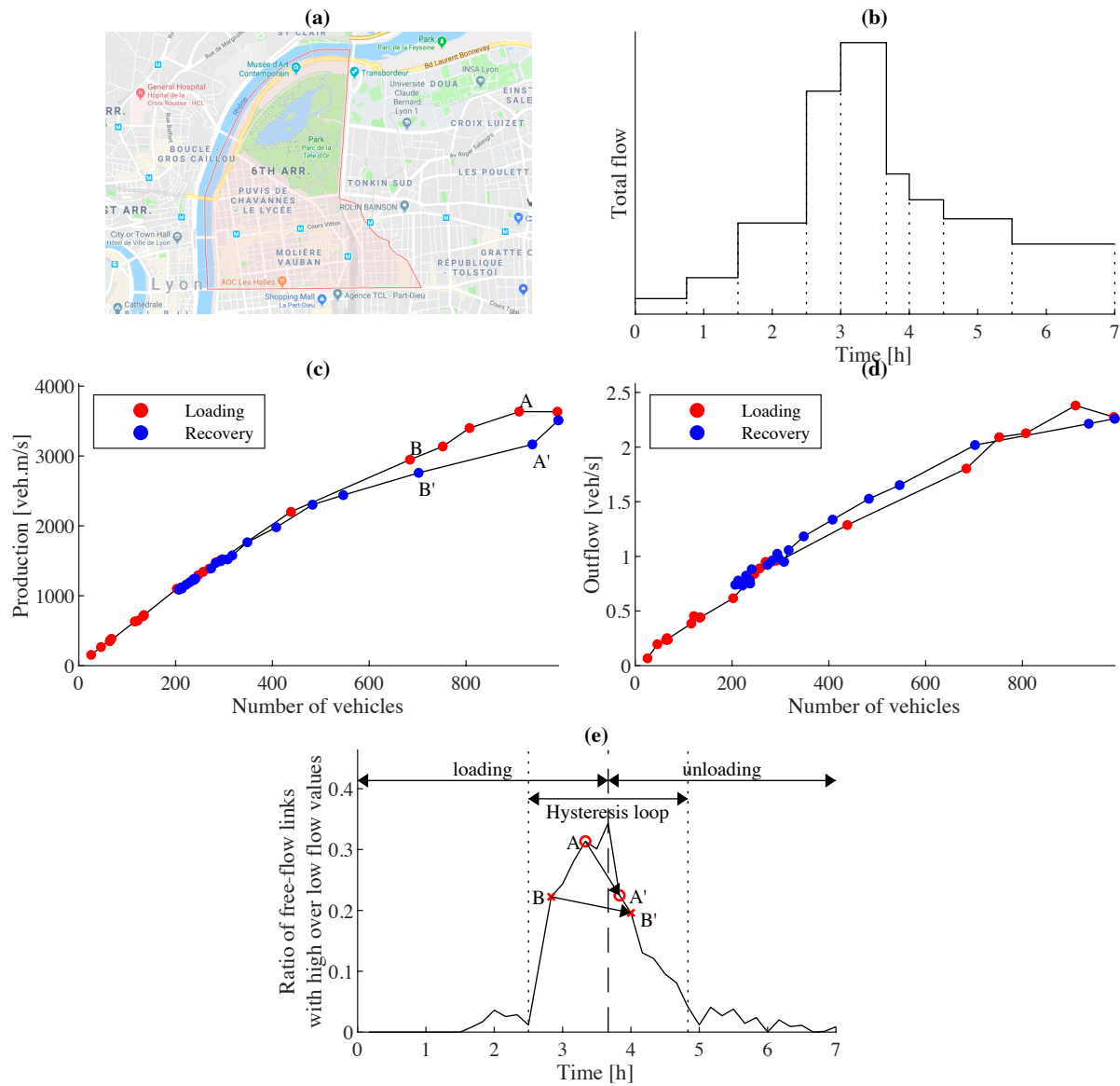


Figure 7: Microscopic simulation results for an urban district (Lyon 6th, France) – (a) Map of the studied area © Google Map 2018 (b) Demand profiles for the simulation (c) simulated p-MFD (aggregation time 10 min) (d) simulated o-MFD (e) time-evolution of the ratio of links that are experimenting (i) free-flow conditions and (ii) flow values higher (over lower) than the maximal network outflow.

The microscopic simulation results come from the Symuvia platform¹. This software has been developed by the authors' research laboratory and is founded on the Newell's car-following law (Newell, 2002; Leclercq et al., 2007). It now includes all the classical features of urban traffic (multi-class, multi-lanes, signalized and un-signalized intersections...). Here, the simulation has been set-up for the 6th district of Lyon for the total time period that mimicking the morning peak hour with a realistic time-dependent OD matrix, see Figure 7a for the network map and Figure 7b for the global time evolution of OD flows. Traffic assignment is based on the Logit's model (Dial, 1971) with a steady path flow distribution for all OD pairs. The simulation results match the loop detector data we have at 15 locations inside the area with a mean error below 9.5% over the total simulation period (7h).

The simulation results are aggregated over 10 min periods to determine the mean flow, outflow and number of vehicles. Note that some trips start and end inside the studied area and are properly considered during aggregation. Figure 7c shows the resulting p-MFD while Figure 7d presents the o-MFD. The p-MFD exhibits a clockwise hysteresis loop while the o-MFD loop is counter-clockwise. These results

¹ The Symuvia platform will be made available open-source during spring 2019.

confirm the patterns observed in the simpler arterial case. In particular, it shows that the counter-clockwise hysteresis loop on the o-MFD can be observed for more complex urban networks. This is an important result because o-MFD are hardly observable in reality because it is very difficult to monitor the trip-completion rate for a given area.

The simulation results presented here are not made to be a universal proof. They show that the aggregate dynamics during the network loading and recovery are following different paths during and close to saturation. The congestion mechanisms related to bottlenecks that we unravel with the arterial case may play a role as the patterns are identical. One confirmation, while not being a formal proof, can be found in Figure 7e. We first select links in the network that are not yet reaching saturation or congestion, i.e. links with a mean spatial speed, which is higher than 70% of the free-flow speed for a given 10 min time period. These ‘free-flow’ links are located upstream and downstream of the already active bottlenecks. If our conjecture in terms of traffic dynamics is correct, the upstream ones should statistically experience more often flow values higher than the maximal outflow during the loading and flow values lower than the maximal outflow during the unloading. To test this conjecture, we plot the time-evolution of the ratio of the total length of free-flow links with flow values higher (over lower) than the maximal outflow observed on the o-MFD. Figure 7e clearly shows a sudden increase of this ratio during the loading phase when the hysteresis loop starts being observed. This ratio then drops as soon as the unloading phase begins. Furthermore, we have identified in Figure 7c two points with close accumulation values but on different sides of the hysteresis loop, i.e. $A \rightarrow A'$ and $B \rightarrow B'$. It appears in Figure 7e that the ratio is lower for points A' and B' , which are on the unloading hysteresis curve. Note that the bigger the difference in flow in the hysteresis, the bigger the drop in the ratio. All these observations reinforce our conjecture that the local traffic dynamics (congestion wave propagation) at a bottleneck during loading and unloading contribute to the hysteresis phenomenon. Note that this is not contradicting other explanations in the literature, which are more related to uneven distribution of congestion and/or unbalanced flows at intersections, as all phenomenon can combine at the network level.

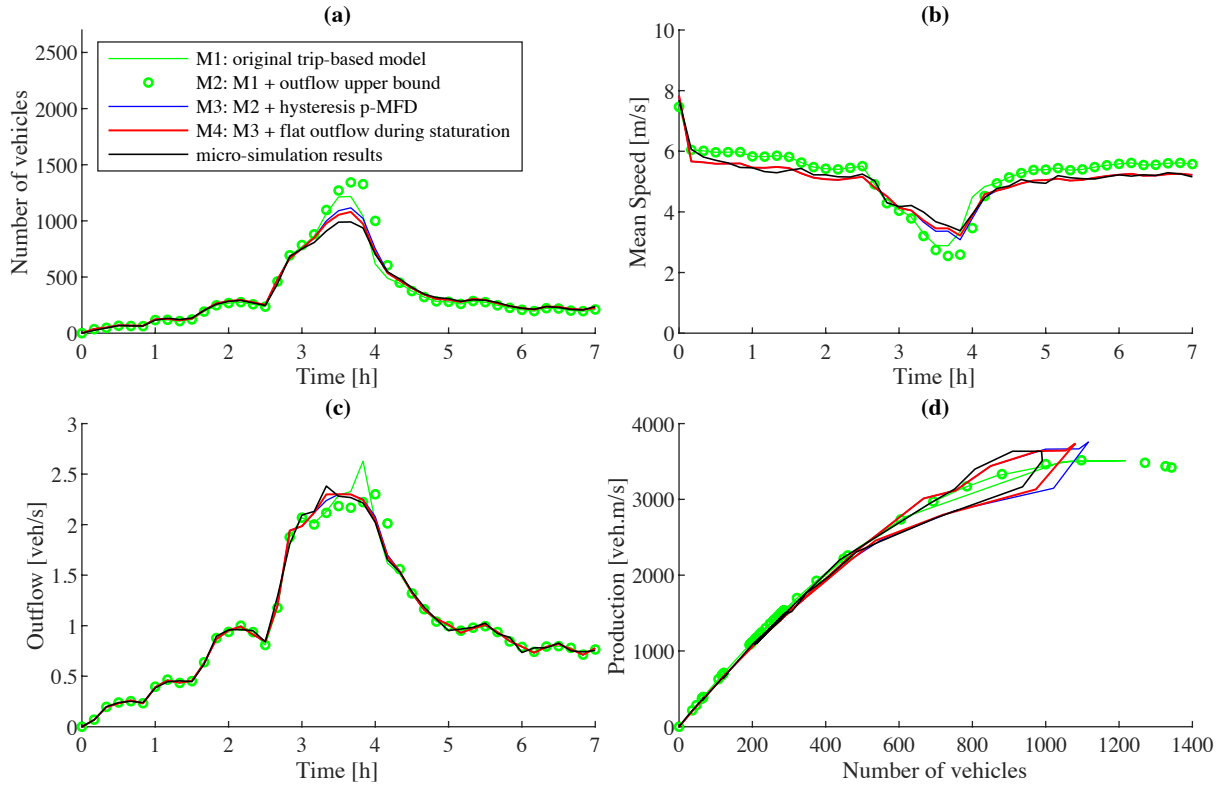


Figure 8: Comparison between trip-based MFD models and micro-simulation for an urban district (Lyon 6th, France) – (a) Time-evolution of the number of vehicles inside the area (b) Time-evolution of the mean speed (c) Time-evolution of the outflow (d) p-MFD.

We are now repeating the benchmark of MFD models for this most complex network setting. They are fed by exactly the same demand as the micro-simulator considering four kind of trips (external→internal, internal→external, internal→internal, external→external). The trip lengths coming from the micro-simulation have been averaged into a single value for the MFD models, which is equal to 1506 m, independent of the trip origins and the destinations. The theoretical (unimodal) p-MFD is defined as the quadratic regression of all micro-simulation observations in Figure 7c while the hysteresis p-MFD is composed of two quadratic regressions of the same observations but separated into two categories: loading and recovery.

Figure 8 shows the comparison with the micro-simulation for the different versions of the trip-based models. It appears that the results confirm all the conclusions we drew in the arterial case. The original trip-based model (M1) is very accurate in free-flow conditions, whatever the considered aggregated variables, here n , V and f_{out} . When the network reaches saturation, when $3.2 < t < 4.1$ h, some discrepancies start to appear. First, the number of vehicles is overestimated by more than 25% during this time period. In the meantime, the mean speed is underestimated, see Figure 8b, and the outflow is first underestimated during the end of the loading phase and then overestimated during the beginning of the unloading phase, see Figure 8c. The overestimation of the outflow is directly the consequence of the causality effect when the reduction of the inflow speeds up all vehicles as explained in the arterial case. To patch this problem, we proposed to introduce an upper-bound that corresponds to the maximal outflow (M2). This solution corrects the outflow values but here tends to increase the overestimation of the number of vehicles and the underestimation of the mean-speed. When looking at the green circles, it appears that the correction influences the outflow (and the other variables) slightly before the outflow peak that is observed for M1. This is because the bound is implemented as a minimal time period between two trip endings while the observations are here aggregated over 10 min to match the micro-simulation outputs. So, the bound is active from time $t=3.5$ h to time $t=4.1$ h. By reducing the outflow, we mechanically increase the accumulation in the system and worsen the results provided by M1. Introducing the p-MFD with hysteresis instead of the theoretical MFD (M3) significantly improves the accuracy of the trip-based results. This is because the mean speed is much better estimated, see Figure

8b. The outflow and number of vehicles are now very close to the micro-simulation results during saturation. Imposing the outflow to be maximal during the saturation phase (M4) is not really necessary here as the causality effect is not playing a major role at the end of the loading phase as long as we are using the hysteresis p-MFD. This also because the p-MFD is not exhibiting in that case a flat region at capacity, so the system is considered to be in free-flow up to the direct saturation, i.e. the outflow constraint is almost never active. Finally, Figure 8d shows the resulting p-MFD for the different trip-based formulations. It is not surprising that only versions M3 and M4 are able to reproduce the hysteresis pattern while M1 and M2 are following the unimodal theoretical MFD.

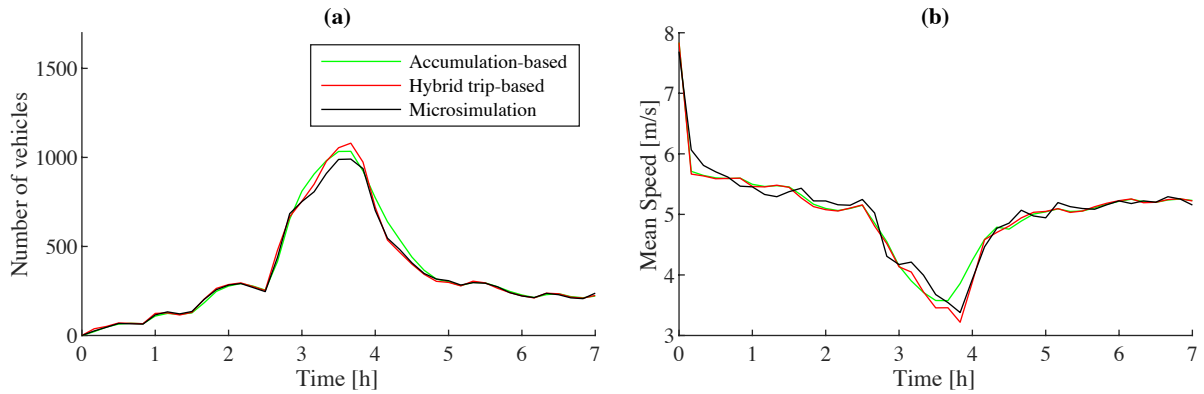


Figure 9: Comparison between MFD models and micro-simulation for an urban district (Lyon 6th, France) – (a) Time-evolution of the number of vehicles inside the area (b) Time-evolution of the mean speed.

To conclude this section, we are now comparing the accumulation-based results with the most accurate trip-based version (M4) and the micro-simulation. This last trip-based version has been named as hybrid during the arterial case analysis as the two patches (bounds on outflow) make the trip-based model very close to the accumulation one. Figure 9a presents the time-evolution of the number of vehicles while Figure 9b shows the time-evolution of the mean-speed. The results here are the confirmation that the hybrid model and the accumulation-based models are very close during the saturation phase. They are also very close during the loading of the network while the mean traffic states are in free-flow. Note that we are not observing the discrepancies of the accumulation-based model here when the demand is increasing step-by-step because the model outputs have been averaged over 10 min to fit the micro-simulation observations. Otherwise, there would have been slight discrepancies at each demand step related to the inaccurate estimation of the experienced travel times. During the network recovery the accumulation-based model does not provide a perfect match with the micro-simulation compared to the trip-based model. This is because the accumulation-based model is not able to reproduce the sharp decline of the outflow with the proper time-lag after the sharp decrease of the inflow corresponding to the beginning of the recovery phase, see Figure 7b. The accumulation-based is smoothing too much the outflow decreasing profile (it starts sooner but last longer), which explains the temporary overestimation of the number of vehicles and mean speed between times 4.1 and 4.6 h.

The comparisons of the accumulation-based, the trip-based and the micro-simulation outputs lead to exact same results for the Lyon 6th network and the arterial case. The global dynamics pattern looks very similar and the more accurate MFD model appears also here to be the hybrid one (the trip-based model with the two patches and the hysteresis p-MFD). What is impressive to our opinion is the accuracy of the MFD models compared to micro-simulation even when considering fast-varying demand profiles related to realistic peak hour patterns and a real network. It should be noted that in this paper we only investigated single reservoir settings of MFD models, which are certainly more relevant at the scale of an urban district than a full city.

4. Conclusion

This paper investigates the network aggregate dynamics under a demand profile that typically corresponds to peak hours. The first major conclusion is that the p-MFD exhibits a clockwise hysteresis pattern under such circumstances with a higher increasing branch during loading and a lower decreasing

branch during recovery. Such a hysteresis loop starts appearing when the network switches from free-flow to saturation and only disappears when all effects of congestion have disappeared. We demonstrated that this pattern can be explained by the local congestion mechanisms at internal bottlenecks. Previous studies in the literature have emphasized the role of unbalanced flows at intersections as a key trigger for the hysteresis profile. We showed here that another complementary explanation can be found in the traffic dynamics itself. Despite our conclusions have been obtained in the arterial case, we were able to confirm the observations for more complex network designs. Even if the arterial case may look over-simplistic we have the strong conviction that it is a good proxy for what happens when multiple vehicles are crossing a reservoir with multiple origins and destinations. More complex patterns may of course be observed for more general network but the dynamics of a peak hour is mainly that internal bottlenecks start being activated and cause congestion propagating backwards. Then, it is reasonable to assume while very difficult to formally prove that the general dynamics have similarities, i.e. higher flow upstream of bottlenecks during the network loading and lower flow at the same location when the network is unloading. This reason for the hysteresis pattern on MFD curves has not yet been properly documented in the literature to our best knowledge and this paper is filling this gap. Note that ring-road settings may be seen as another relevant option to mimic urban dynamics during peak hour in general networks but we believe the opposite. We are convinced this is not the right way because this kind of setting is very influenced by the fact that upstream and downstream boundaries are identical. This setting is certainly relevant for freeway road sections but in urban networks the congestion appears in our view because of the activation of local bottlenecks that are usually triggered by capacity sharing at intersections. Even if traffic has no clear direction in a general network, there is no reason for the demand upstream of bottlenecks to be defined by the flow that is able to cross them. This is why we decided to favor the arterial setting to better understand the connections between local and global traffic states. Again, the results we obtained in section 3 with more complex urban network tends to prove we were right even no final conclusions can be made due to the complexity of unravelling the physical mechanisms at such a large scale.

The second major conclusion of this paper is that the o-MFD exhibits a counter-clockwise hysteresis loop in the same situation. The main reason is that the outflow is sustained to its maximal value until all vehicles that experienced congestion in the reservoir have left. The fact that such hysteresis patterns are observed clearly highlights the limitation of the steady-state approximation that is used to analytically derive MFD based on the cuts method. In particular, the p-MFD and o-MFD patterns under dynamic loading are demand sensitive on the contrary to the theoretical MFD. More generally, the steady-state approximation should be related to slow varying demand profile to remain valid, which appears not to be the case during network loading and recovery. This means that the MFD we observe based on a dynamic network loading is different from what we would observe if we perform multiple loadings with a constant demand values and then consider the final stationary network state. The latter fits the steady-state approximation and will provide results in accordance with the cuts method with no hysteresis. The former is what is classically obtained from experimental observations over a day. Fortunately, some studies have concluded on the repeatability of real network loading and unloading patterns, e.g. (Lopez et al., 2017). This means that the calibrated hysteresis p-MFD from one particular day is certainly valid for multiple similar days where the global demand pattern is not very different.

This remark has important consequences in terms of MFD models design and calibration. In this paper, we showed that the accumulation-based model should better be set-up with a well-defined o-MFD. The problem is that the o-MFD is hardly observable in practice and that the relation that permits to easily derive it from the p-MFD is also only valid in steady state. In this paper, we showed that there is no point in such a case in refining the calibration of the p-MFD considering the hysteresis loop because it leads to an inconsistent o-MFD when applying the proportionality factor. The key element to obtain good simulation results is to accurately estimate the maximal outflow value and to ensure that this value is sustained during all the saturation period. This is exactly what is predicted by the theoretical MFD (derived from the cuts method). On the contrary, we showed that the trip-based model benefits from a refined p-MFD setting considering the hysteresis loop. This is because the trip-based core component is the speed-MFD and taking the hysteresis loop into account improves the speed estimations.

Another key conclusion about MFD models is that none is perfect. The accumulation-based is accurate during saturation but provides inaccurate travel time predictions in free-flow. The trip-based model outperforms the accumulation-based in free-flow but fails to accurately predict the network aggregate traffic states during saturation. This is because outflow values are widely inconsistent when the internal speed condition is applied due to the causality effect. We showed in this paper that two simple patches can resolve this issue. Interestingly, it appears that the trip-based model with the patches behaves like the accumulation-based model during saturation. This is because the patches control the system outflow in the same way as the accumulation-based model. So, the best MFD model to represent the aggregated dynamics of a network appears to be hybrid acting like the trip-based in free-flow and like the accumulation-based in saturation. Interestingly, (Mariotte and Leclercq, 2018) shows that same conclusions hold when reproducing congestion spillbacks in a multi-reservoir context. Finally, this paper provides some contributions to the actual debate the MFD community about how the entry-flow function should be defined. Here, we have a clear definition of the entry-flow function for an arterial with the distinction between the maximal entry flow that applies before the network full saturation and the maximal network flow μ that applies during saturation.

This paper aims to pave the long way towards a complete understanding on the validity of MFD models and their proper setting. While the literature on MFD is now very rich, there is surprisingly not so much comparisons of MFD simulation results with more detailed observations about network dynamics based on real data or more refined traffic models. The next-step is certainly to consider multi-reservoir systems and see how merging, diverging and path descriptions make the MFD simulation still valid compared to the integration of the local dynamics.

Acknowledgements

This study has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No 646592 – MAGnUM project).

References

- Arnott, R., 2013. A bathtub model of downtown traffic congestion. *Journal of Urban Economics*, 76(1):110-121.
- Buisson, C., Ladier, C., 2009. Exploring the impact of homogeneity of traffic measurements on the existence of Macroscopic Fundamental Diagrams. *Transportation Research Record*, 2124:127-136.
- Daganzo, C.F., 2011. On the macroscopic stability of freeway traffic. *Transportation Research Part B*, 45:782-788.
- Daganzo, C.F., 2007. Urban gridlock: macroscopic modelling and mitigation approaches. *Transportation Research Part B*, 41(1):49-62.
- Daganzo, C.F., 2005a. A variational formulation of kinematic waves: basic theory and complex boundary conditions. *Transportation Research Part B*, 39(2):187-196.
- Daganzo, C.F., 2005b. A variational formulation of kinematic waves: Solution methods. *Transportation Research Part B*, 39(10):934-950.
- Daganzo, C.F., Geroliminis, N., 2008. An analytical approximation for the macroscopic fundamental diagram of urban traffic. *Transportation Research Part B*, 42(9):771-781.
- Daganzo, C.F., 2003. *A Theory of Supply Chains*. Springer: 125 p.
- Dial, R.B., 1971. A probabilistic multipath traffic assignment model which obviates path enumeration. *Transportation Research*, 5:83-111.
- Fosgerau, M., 2015. Congestion in the bathtub. *Economics of Transportation*, 4(4):241-255.
- Friesz, T.L., Luque, J., Tobin, R.L., Wie, B.W., 1989. Dynamic network traffic assignment considered as a continuous time optimal control problem. *Operations Research*, 37(6):893-901.
- Godfrey, J., 1969. The mechanism of a road network. *Traffic Engineering and Control*, 11(7):323-327.
- Geroliminis, N., Sun, J., 2011. Hysteresis phenomena of a Macroscopic Fundamental Diagram in freeway networks. *Transportation Research Part A*, 45:966-979.

- Geroliminis, N., Daganzo, C., 2007. Macroscopic Modeling of Traffic in Cities #07-0413. *Proceedings of the 88th Transportation Research Board Annual Meeting (TRB)*. Washington: Transportation Research Board, 16p.
- Hajiahmadi, M., Knoop, V., De Schutter, B., Hellendoorn, H., 2013. Optimal dynamic route guidance: A model predictive approach using the macroscopic fundamental diagram. In: *IEEE 16th International Conference on Intelligent Transportation Systems (ITSC)*, 1022– 1028.
- Knoop, V. L., Hoogendoorn, S.P., 2014. Network transmission model: a dynamic traffic model at network level, #14-1104. *Proceedings of the 93rd Transportation Research Board Annual Meeting (TRB)*, 15 p.
- Knoop, V.L., Hoogendoorn, S.P., 2013. Empirics of a Generalized Macroscopic Fundamental Diagram for Urban Freeways. *Transportation Research Record*, 2391, 133-141.
- Lamotte, R., Geroliminis, N., 2017. The morning commute in urban areas with heterogeneous trip lengths. *Transportation Research Part B*, in press: <https://doi.org/10.016/j.trb.2017.08.023>.
- Laval, J.A., Leclercq, L., Chiabaut, N., 2018. Minimal parameter formulations of the dynamic user equilibrium using macroscopic urban models: Freeway vs City Streets revisited. *Transportation Research part B*, 117-part B, 676-686.
- Laval, J.A., Chilukuri, B.R., 2016. Symmetries in the kinematic wave model and a parameter-free representation of traffic flow. *Transportation Research part B*, 89:168-177.
- Laval, J.A., Castrillon, F., 2015. Stochastic approximations for the macroscopic fundamental diagram of urban networks. *Transportation Research part B*, 81(3):904-916.
- Laval, J.A., Leclercq, L., 2013. The Hamilton-Jacobi partial differential equation and the three representations of traffic flow. *Transportation Research part B*, 52:17-30.
- Leclercq, L., Parzani, C., Knoop, V.L., Amourette, J., Hoogendoorn, S.P., 2015. Macroscopic traffic dynamics with heterogeneous route patterns. *Transportation Research part C*, 55:292–307.
- Leclercq, L., Geroliminis, N., 2013. Estimating MFDs in Simple Networks with Route Choice. *Transportation Research part B*, 57:468-484.
- Leclercq, L., Laval, J.A., Chevallier, E., 2007. The Lagrangian coordinates and what it means for first order traffic flow models. In: Allsop, R.E., Bell, M.G.H., Heydecker, B.G. (Eds), *17th ISTTT*, Elsevier, London, 735-753.
- Lentzakis, A. F., Ware, S. I., Su, R., 2016. Region-based dynamic forecast routing for autonomous vehicles. In: *IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, 1464–1469.
- Lighthill, M.J., Whitham, J.B., 1955. On kinematic waves II: A theory of traffic flow in long crowded roads. *Proceedings of the Royal Society*, A229:317-345.
- Lopez, C., Leclercq, L., Krishnakumari, P., Chiabaut, N., Van Lint, H., 2017. Revealing the day-to-day regularity of urban congestion patterns with 3D speed maps. *Scientific Reports*, 7(1):14029.
- Mahmassani, H.S., Saberi, M., Zockaie, A., 2013. Urban network gridlock: Theory, characteristics, and dynamics. *Transportation Research Part C*, 36:480-497.
- Mahmassani, H.S., Williams, J.C., Herman, R., 1987. Performances of Urban Traffic Networks. In: (Gartner, N.H., Wilson, N.H.M., Eds) *Proceedings of the 10th International Symposium on Transportation and Traffic Theory*, 1-20.
- Mahmassani, H.S., Williams, J.C., Herman, R., 1984. Investigation of Network-Level Traffic Flow Relationships: Some Simulation Results. *Transportation Research Record*, 971:121-130.
- Mariotte, G., Leclercq, L., 2018. Flow exchanges in multi-reservoir systems with spillbacks. *Submitted to publication*.
- Mariotte, G., Leclercq, L., Laval, J.A., 2017. Macroscopic urban dynamics: analytical and numerical comparisons of existing models, *Transportation Research part B*, 101:268-282.
- Mazlounian, A., Geroliminis, N., Helbing, D., 2010. The spatial variability of vehicle densities as determinant of urban network capacity. *Philosophical Transactions of Royal Society A*, 368(1928):4627-4648.
- Merchant, D.K., Nemhauser, G.L., 1978. A model and an algorithm for the dynamic traffic assignment problem. *Transportation Science*, 12(3):183–199.

- Newell, G.F., 2002. A simplified car-following theory: a low-order model. *Transportation Research B*, 36(3) :195-205.
- Newell, G.F., 1993. A simplified theory of kinematic waves in highway traffic, I general theory, II queuing at freeway bottlenecks, III multi-destination flows. *Transportation Research Part B*, 27(4):281–313.
- Richards, P.I., 1956. Shockwaves on the highway. *Operations Research*, 4:42-51.
- Saberi, M., Mahmassani, H.S., 2013. Hysteresis and Capacity Drop Phenomena in Freeway Networks: Empirical Characterization and Interpretation. *Transportation Research Record*, 2391:44-55.
- Saberi, M., Mahmassani, H.S., 2012. Exploring Properties of Networkwide Flow-Density Relations in a Freeway Network. *Transportation Research Record*, 2315:153-163.