Estimating MFDs in simple networks with route choice ☆

Ludovic Leclercq [a,*], Nikolas Geroliminis [b]

[a] LICIT, IFSTTAR, ENTPE, Université de Lyon, Rue Maurice Audin, 69518 Vaulx-en-Velin Cedex, France
[b] Urban Transport Systems Laboratory (LUTS), Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

ARTICLE INFO

ABSTRACT

The concept of the Macroscopic Fundamental Diagram (MFD) is elegant and attractive because it provides a global view of traffic behavior and performance at a network level. However, recent research shows that the MFD shape can be influenced by local traffic heterogeneities. Notably, route choices and heterogeneous local capacities may drive uneven (in space) or inconsistent (in time) distributions of congestion and then affect the shape and the scatter of the MFD. We are far from having a global understanding of the connections between local phenomena and the resulting MFD. This paper first aims to improve existing MFD estimation method for a succession of links with traffic signals. The new method overcomes previous limitations, notably regarding to the topology and signal settings regularities, by fully utilizing the receipts of the variational theory. Then, a single network with several parallel routes is investigated. MFDs on different routes are estimated with the variational method and then aggregated in a unified MFD for stationary and dynamic conditions and different sorts of equilibria (user and system optimum). It appears that the flow distribution among routes smoothly varies with respect to the total flow either in free-flow or congestion situations. Such a distribution is much more rough for system optimum, where it presents some discontinuities and is far from equity. This means that a control strategy able to lead such a network to the perfect system optimum would be hard to tune, especially in the congested regime. However, being able to determine the MFD corresponding to the system optimum provides a valuable reference to estimate the current efficiency of the considered network. Case studies for different simple networks and insights for generalization at the city level are proposed.

© 2013 Elsevier Ltd. All rights reserved.

1. Introduction

Traffic data from loop detectors at one location often appear very scattered, notably in urban areas. Indeed, local data aggregation in time or by lane often mixes inconsistent traffic states. Dedicated filtering methods (e.g. Cassidy, 1998; Chiabaut and Leclercq, 2011), can then be applied to calibrate the local Fundamental Diagram (FD). Recent observations from empirical data in downtown Yokohama (Geroliminis and Daganzo, 2008) show that by spatially aggregating the highly scattered plots of flow vs. density from individual loop detectors in homogeneous networks, the scatter almost diminishes and a well-defined Macroscopic Fundamental Diagram (MFD) exists between space-mean flow and density. The idea of an MFD with an optimum accumulation belongs to Godfrey (1969) and similar approaches were introduced later by Herman and

---

Prigogine (1979) in the context of the two-fluid model, Mahmassani et al. (1984) and Daganzo (2007). The verification of its existence with dynamic features is recent (Geroliminis and Daganzo, 2008). This work shows, using a field experiment in downtown Yokohama, (i) that urban neighborhoods can exhibit an MFD relating the number of vehicles to space-mean speed (or flow), (ii) there is a robust linear relation between the neighborhood's average flow and its total outflow (rate vehicles reach their destinations) and (iii) the MFD is a property of the network infrastructure and control and not of the demand, i.e. space-mean flow is maximum for the same value of vehicle density and not sensitive to time-dependent origin–destination (O–D) tables.

Property (i) is important for modeling purposes as details in individual links are not needed to describe the congestion level of cities and its dynamics. It can also be utilized to introduce simple control strategies to improve mobility in homogeneous city centers building on the concept of an MFD (Daganzo, 2007; Haddad and Geroliminis, 2012; Keyvan-Ekbatani et al., 2012; Geroliminis et al., 2013; Haddad et al., 2013). The main logic of the strategies is that they try to decrease the inflow in regions with points in the decreasing part of an MFD and maximize the outflow for regions with high attraction for destinations. Property (ii) is important for monitoring purposes as flow can be easily observed with different sensors, see for example (Van Lint and Hoogendoorn, 2010), while outflow cannot. Property (iii) is important for control purposes as hierarchical active traffic management schemes can be developed without a detailed knowledge of O–D tables. Despite these recent findings for the existence of MFDs with low scatter, these curves should not be a universal law and network topology and O–D tables might affect the shape and scatter under different cases. In particular, networks with an uneven (in space) and inconsistent (in time) distribution of congestion may exhibit traffic states that are below the upper bound of an MFD and too scattered. Property (i) then no longer holds. Recent findings from empirical and simulate data (Geroliminis and Sun, 2011; Mazloumian et al., 2010; Knoop et al., 2012) have identified the spatial distribution of vehicle density in the network as one of the key components that affect the scatter of an MFD and its shape. They observe well-defined relations between flow and density when link density variance is constant. They also observe low scatter relationships between network flow and variance of link density for a given network. In other words, the average network flow is consistently higher when link density variance is low for the same network density, but higher densities can create points below an MFD when they are heterogeneously distributed. Daganzo et al. (2011) showed for simple networks with 2 interconnected rings, that networks with densities in the congested regime can produce strong instabilities and bifurcations and lead the system to gridlock. In a similar direction (Laval, 2010) drives some concerns on property (iii). Indeed, he shows with micro-simulation on idealized and rectangular networks, situations when variations in O–D tables and/or route choices induce scattering on the MFD, exactly because they modify the spatial distribution of density. Property (iii) should then be further investigated with a particular attention to the crucial question of the influence of route choices and network equilibrium on the MFD shape. This is one of the research objectives of the current work. Other recent studies related to MFD can be found in Ji et al. (2010), Knoop et al. (2012), and Saberi and Mahmassani (2012).

Efficient estimation methods are required to study MFD on networks. The most common method averages flow and occupancy data from loop detectors in the study region (e.g. Geroliminis and Daganzo, 2008). Courbon and Leclercq (2011) have shown that such method may introduce significant bias in the MFD depending on the loop detector distribution over the networks. Probe vehicles can also be used to estimate MFDs. However, a sufficient number of observations is required and some kind of probe vehicles may have specific behavior that can also introduce additional bias, e.g. taxis. Analytical methods that derive the MFD from the network infrastructure and control parameters are then really appealing. Daganzo and Geroliminis (2008) have set the foundation of such methods by introducing moving observers and applying some key features of the variational theory (Daganzo, 2005a,b). Analytical formulae are provided in this paper but only for networks with strong regularities in topology and signal settings. Otherwise, the provided bounds for the MFD are not necessarily tight or simulations are needed to enumerate all possible paths. Geroliminis and Boyaci (2012) propose an algorithm that extends the variational method to more general cases but some regularity conditions still remain. These references do not provide any analytical tools to identify the effect of route choice in the shape of the MFD.

The first aim of the present paper is to propose an accurate and elegant method to estimate MFDs with the variational method. It consists in defining a suitable graph in the plane (t,x) to encompass all relevant paths for moving observers. This graph is proved to be sufficient. It directly provides all the necessary cuts that tightly define the upper envelop of the MFD. Regularity conditions are no longer required and this estimation is also valid for irregular topologies, that are more realistic. The second aim of the paper is to investigate the relationship between route choice and the MFD shape in order to clarify property (iii). A simple network with several parallel routes will support the exposition and both user equilibrium and system optimum conditions will be considered. MFDs on different routes are estimated with the variational method and they are aggregated in a unified network MFD for stationary and dynamic conditions and different sorts of equilibria. The paper also discusses the feasibility and equity in the implementation of such equilibria. This research sheds some light in the integration of the route choice in an MFD framework for networks, which should be considered in the development of active traffic management schemes.

This paper is organized as follow: Section 2 summarizes the essential facts of variational theory applied to MFD estimation. Section 3 presents the advanced MFD estimation method and discusses its properties. Section 4 deals with aggregated MFD corresponding to a parallel network with route choices. Stationary and non-stationary traffic states are investigated for user equilibrium and system optimum. Finally, Section 5 presents a discussion.

2. Background on variational theory

This section recaps the main insights on variational theory (VT) applied to MFD estimation, see Daganzo and Geroliminis (2008) for details. A complete description on the variational theory can be found in Daganzo (2005a,b), and Daganzo and Menendez (2005).

A hyperlink is defined as a series of m successive links with traffic signals, the same number of lanes and homogeneous traffic conditions, i.e. well balanced in- and out-flows at intersections. Link i is defined by its length l_i and its signal settings (green G_i, cycle C_i and offset from a common reference δ_i). The total length of the hyperlink is $L = \Sigma l_i$.

Traffic on each link is described by the kinematic wave model (Lighthill and Whitham, 1955; Richards, 1956) and the same triangular FD with free-flow speed u, wave speed w, and jam density κ. The capacity during green time is equal to $s = wu\kappa/(w + u)$. The centerpiece of the VT is to express the kinematic wave model through a least-cost path problem. Valid paths start from the initial boundary condition and move in the (t,x) plane with a local speed v that should always fall in the range $[-w,u]$. The cost function $r(v)$, is linear and decreases from $r(-w) = w\kappa$ to $r(u) = 0$ when the FD is triangular. Note that $r(0) = s$. Physically, this cost function defines the maximum local passing rate at which traffic can overtake an observer moving along a valid path. Red signals correspond to point bottlenecks with fixed locations and limited time durations. In VT, they can be assimilated to shortcuts with a null cost (Daganzo and Menendez, 2005). Thus, when a portion of a path matches with a red signal, r is no longer equal to s but falls to 0. Finally, let consider all the valid paths that connect two points in the (t,x) plane. All these paths have the same mean speed V. Let $\Delta(P)$ be the mean cost along path P. $\Delta(P)$ is calculated from $r(v)$, including the effect of shortcuts. We can then define the quantity $R(V)$ that corresponds to the upper bound at which traffic can overtake any observer that travels at mean speed V:

$$R(V) = \inf_P(\Delta(P); V_P = V) \tag{1}$$

Daganzo and Geroliminis (2008) associate to any hyperlink and infinite hyperlink with periodic characteristics in space, i.e. the original hyperlink is copied upstream and downstream with a period L. They have then proved that the MFD for such an infinite hyperlink corresponds to the upper envelop of a set of lines defined by a single parameter, i.e. the mean speed V of the observer paths. These lines are called "cuts" because they individually impose constraints of the form: $Q \leqslant KV + R(V)$, where Q is the flow and K, the density. Note that this envelope is tight and that the MFD is then defined by:

$$Q = \inf_V(KV + R(V)) \tag{2}$$

where V should vary among all the valid speed range $[-w,u]$, see Fig. 1a. The above property formally requires that the time duration of paths, where $R(V)$ is calculated, to be infinite. An alternative solution to considering an infinite periodic hyperlink is to switch the observer path to the downstream border of the hyperlink, when it reaches the upstream border when $V > 0$ (or reciprocally when $V < 0$), see Fig. 1b. Note that the signal setting on both borders is identical because it corresponds to the same traffic signal in the periodic infinite pattern, see Fig. 1b. Thus, we only have to estimate $R_j(V)$ on several successive time windows j whose duration is $T = L/V$. $R(V)$ is equal to the mean value of $R_j(V)$ when j tends to infinity. In practice, the law of large numbers tell us that a sufficient number of well distributed time windows is sufficient to calculate $R(V)$. In practice, this number depends on the irregularity of the signal settings but is usually lower than ten. The key question is to develop an efficient method to estimate $R_j(V)$.
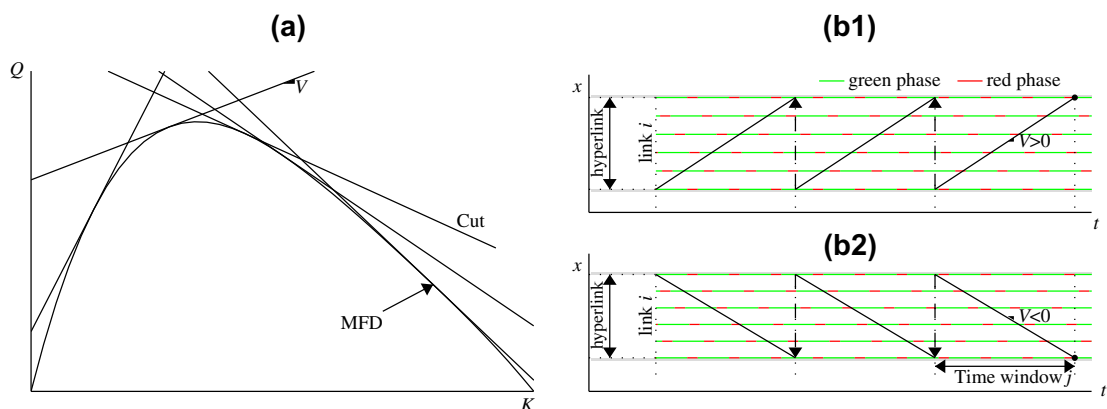

Fig. 1. (a) MFD defined by a family of cuts, and (b) decomposition of an observer path with speed V over the hyperlink.

3. Advanced variational methods for MFD estimation on a single hyperlink

3.1. A sufficient variational network without regularity conditions

Estimating the value of $R_j(V)$ for all V can be tedious. We then propose to discretize cut speeds values by only considering paths between the ends of red times on upstream and downstream borders, see Fig. 2a. Note that forward moving paths ($V > 0$) start at the upstream border while backward moving paths ($V < 0$) start at the downstream border ($V < 0$). Backward and forward moving paths have the same properties except a reverse travelling direction. Then, we will restrict our exposition to $V > 0$. We will later integrate both forward and backward paths to construct a sufficient variational network. Cut speeds can take the following discrete values $V_k = L/kC_m$; k is an integer such that $V_k \leqslant u$, i.e. $k \geqslant L/uC_m$. Different k values define time windows associated to different paths and consequently to cuts. Different starting points with the same value of k define a set of different time windows j associated to the same global path and cut k. Time windows will be referred as $T_{j,k}$ where j is the index of the starting point and k the index of the path speed. Remember that cut k is fully defined by V_k and that $R(V_k) = \text{mean}(R_j(V_k))$.

Let consider the two external points A and B associated to $T_{j,k}$, see Fig. 2b. Eq. (1) requires that all valid paths with mean speed V_k be investigated to estimate $R_j(V_k)$. Hopefully, the number of paths can be dramatically reduced with some VT properties. First, Daganzo (2005b) has demonstrated that when the cost function r is linear, i.e. the FD is triangular, any valid forward moving paths between A and B, i.e. within the lozenge ACBC′, has the same mean cost with the two border paths $A \rightarrow C \rightarrow B$ and $A \rightarrow C' \rightarrow B$ that connect A and B with the speed u and then 0 ($A \rightarrow C \rightarrow B$) or 0 and then u ($A \rightarrow C' \rightarrow B$), see Fig. 2b. Then, without bottlenecks only one path has to be considered. The bottlenecks question is addressed in (Daganzo
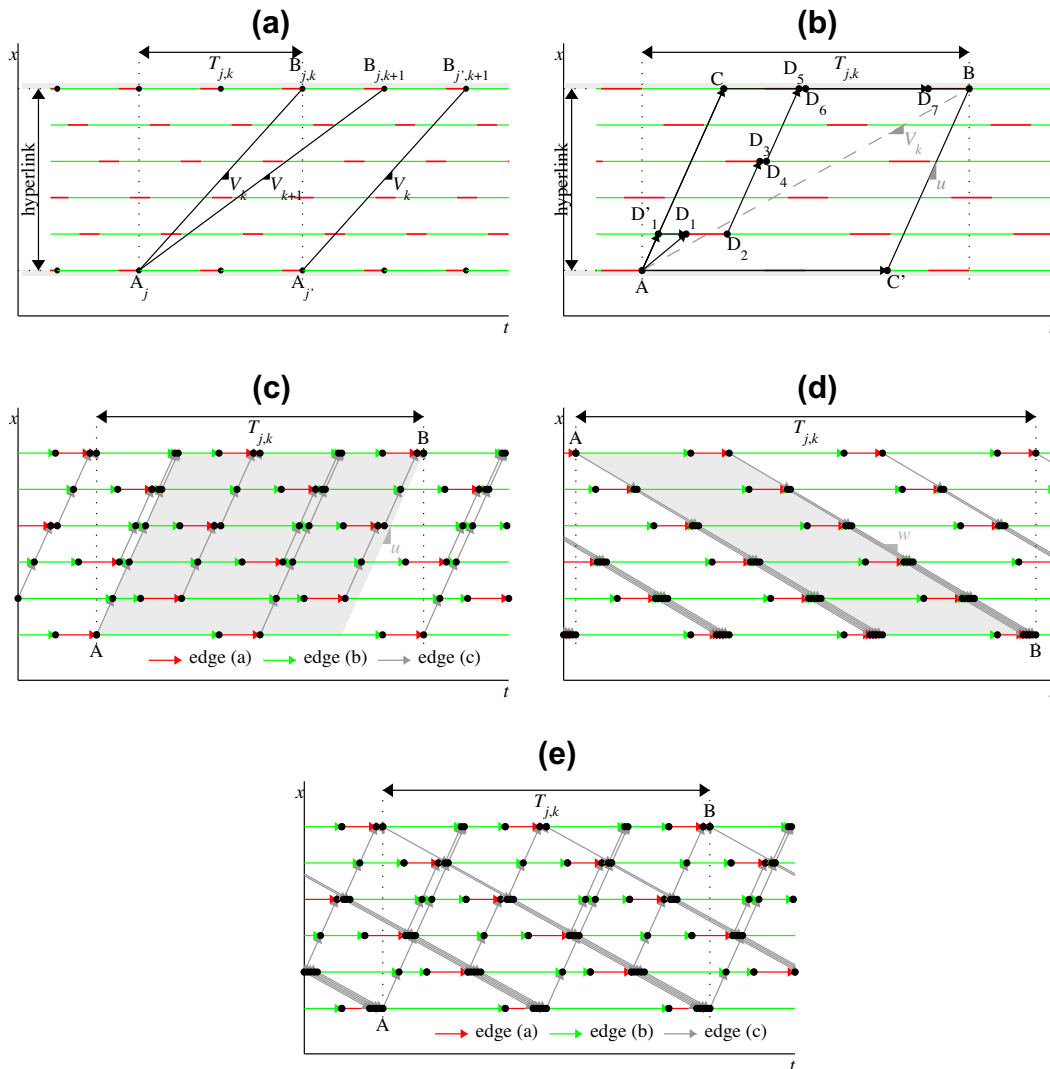

Fig. 2. (a) Discrete paths and associated time windows, (b) possible optimum paths for a given time window with shortcuts, (c) variational graph for backward moving cuts estimation, (d) variational graph for congested cuts estimation, and (e) merged global variational graph.

and Menendez, 2005) through the shortcut and the optimal path theorems and their corollaries. Our problem is homogeneous because the FD is the same in all links. Thus, optimal paths between A and B with shortcuts, i.e. least-cost paths, are necessary composed with one or more pieces of the following three types: (i) continuous sections of bottlenecks, (ii) least duration access and egress sub-paths and (iii) inter-bottleneck sub-paths with speed u. Type (i) means that an optimal path that reaches a bottleneck will stay on the associated shortcut until its ends unless the path has to reach the final point B. Type (ii) means that optimal paths maximize the time spent in shortcuts. Type (iii) means that the end of any shortcuts should be connected to all other reachable shortcuts with a speed u.

For example, path A $\rightarrow$ D$_i$ $\rightarrow$ B in Fig. 2b connects points A and B and respects all of the above conditions. We now have to construct all the possible paths that respect the above conditions to be sure to capture the optimal path with bottlenecks. Note that sub-paths between two bottlenecks with speed lower than u can be substituted by sub-paths with speed u and then 0 for the same costs. For example, sub-path A $\rightarrow$ D$_1$ can be replaced by sub-path A $\rightarrow$ D$_1'$ $\rightarrow$ D$_1$ in Fig. 2b. Thus, possible paths only experiment speeds equal to u or 0 (and 0 or $-w$ for backward observers) and should maximize the time spent in shortcuts. A convenient way to enumerate all the possible paths is to draw a graph with the following three types of edges, see Fig. 2c:

a. the red phases of all the traffic signals;
b. the green phases of all the traffic signals;
c. the paths with speed u that start from the ends of all red times of each signal and propagate until reaching another edge (a). Vertices should be added anytime such a path crosses edges (b) and on edge (a).

Note that a reverse graph can also be defined for determining backward moving cuts. The only difference is that edges (c) should have a speed equal to $-w$ and a cost equal to $w\kappa$, see Fig. 2d. We then combine all types of edges for forward and backward moving observers and we create a full graph, a merged global variational graph, see Fig. 2e (Daganzo, 2013). The above graph in Fig. 2e is sufficient because all paths between any points A and B that respect conditions (i) to (iii) are included by construction. Note also that we do not need to consider new vertices, which are the intersection of forward and backward observers in the middle of the links. The above graph is sufficient because all paths between A and B that respect conditions (i) to (iii) are included by construction. Indeed, point A is connected with all the shortcuts, shortcuts are connected with speed u when possible and time on shortcuts is maximized because edges (c) only start at end of red phases. Least mean cost between A and B can then easily be derived with a classical shortest path algorithm, e.g. the Dijsktra's algorithm. Note that cost r is s on edges (a) and 0 on edges (b) and (c). Note also that we can directly draw the sufficient graph for the whole hyperlink and then identify points A_j and $B_{j,k}$ that define any time window $T_{j,k}$, see Fig. 2c. Shortest path algorithms usually directly provide the shortest path between an origin and all the vertices. Calculations for all $T_{j,k}$ can then easily be optimized. Note that a reverse graph can also defined for determining backward moving cuts. The only difference is that edges (c) should have a speed equal to $-w$ and a cost equal to $w\kappa$, see Fig. 2d. Finally, a remarkable property should be mentioned in the regular case, i.e. $l_i = l$, $G_i = G$, $C_i = C$ and $\delta_i = i\delta$. In that case, the network presents a periodic pattern and all the time window $T_{j,k}$ are identical for a given k. Thus, only one starting point has to be considered because $R(V_k) = R_j(V_k)$ for all j. This greatly simplifies the cuts calculation. Note that by considering independently forward and backward moving observers in graphs of Fig. 2c and Fig. 2d is not sufficient, as a shortest path might consist of both forward and backward movements in the graph of Fig. 2e.

It is interesting to notice that the proposed graph encompasses the practical cuts defined in (Daganzo and Geroliminis, 2008; Geroliminis and Boyaci, 2012). These cuts correspond to observers that only move forward at speed 0 or u, that stop at red times but sometimes also at green times (to experiment several mean speed). In the regular case, these cuts can conveniently be parameterized with γ, the number of link an observer travels before stopping, see (Daganzo and Geroliminis, 2008). The point is that these practical cuts can correspond to possible optimum paths but not necessary to the optimal one, as the observers of these cuts do not travel backward and forward, but only in one direction. That is why Daganzo and Geroliminis (2008) mention that these cuts provide an upper bound that not always tightly defines the MFD, especially in cases of short links. The variational graphs proposed in this paper overcome this limitation because the shortest path algorithm investigates all the possible optimum paths, by moving forward and backward in the sufficient graphs defined in Fig. 2e. The resulting cuts then provide a tight envelop for the MFD without any regularity conditions. The only restriction is that the estimated MFD is discretized because we only investigate discrete values for V. Note finally that it is possible to demonstrate that k values strictly higher than m can be disregarded because the associated cuts are never tight.

Fig. 3a presents the resulting forward and backward moving cuts with forward and backward average speed calculated for a particular hyperlink (its characteristics are provided in the figure legend). The corresponding variational graphs are provided in Fig. 2c–e. Note that none of the two variational graphs provide the stationary cut corresponding to the minimum capacity $S = \min(sG_i/C_i)$ observed at the most constraining traffic signals. This cut can easily be added. Note that this cut is not a tight one for the specific case. Finally, Fig. 3b shows the only relevant cuts that fully define the MFD.

3.2. Cuts properties for regular hyperlinks

Only the variational graphs and the practical cuts (Daganzo and Geroliminis, 2008) methods provide the MFD without requiring simulated or measured traffic data. They can both be considered as semi-analytical methods. Practical cuts are very intuitive and can be directly connected with real observer paths, notably in the regular case (same characteristics for all links and signals) through parameter γ. Practical cuts are also useful because in case of irregular urban networks with multiple
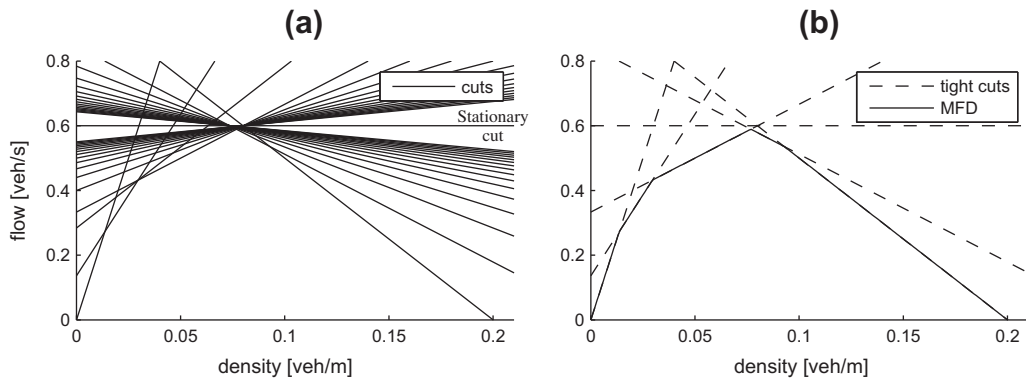
Fig. 3. (a) Whole sets of cuts for a given regular hyperlink ($m = 5$; $l_i = 200$ m; $C_i = 100$ s; $G_i = 75$ s; $\delta_i = i * 62$ s; $u = 20$ m/s; $w = 5$ m/s; $\kappa = 0.2$ veh/m), and (b) relevant cuts and resulting MFD.
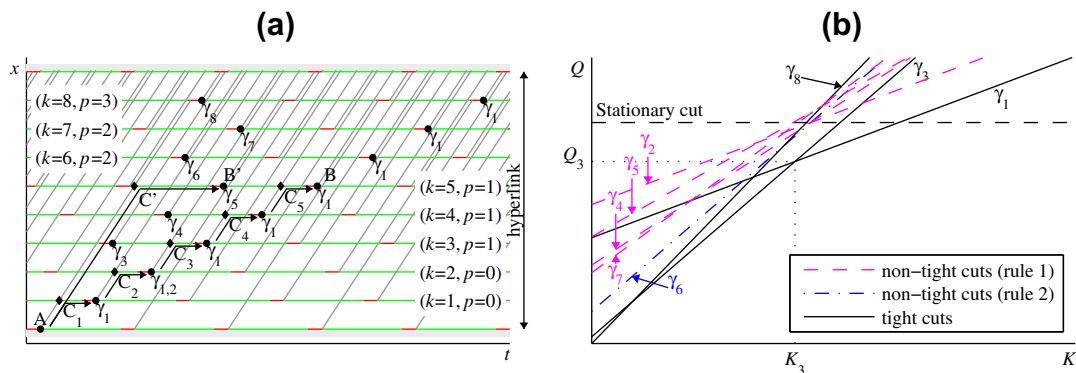

Fig. 4. (a) Signal-timing diagram with paths of all practical cuts γ_k ($m = 9$; $l_i = 200$ m; $C_i = 100$ s; $G_i = 85$ s; $\delta_i = i * 59$ s; $u = 10$ m/s; $w = 5$ m/s; $\kappa = 0.175$ veh/m), and (b) tight cuts selection.

route choices, it is not possible to construct a unique sequence of links between origins and destinations to develop a sufficient graph of Fig. 2e. Thus, one might need to construct and solve separately the forward and backward VT graphs of Fig. 2c and Fig. 2d, by assuming some regular properties of links (e.g average link length and signal settings). Note that only the complete graph of Fig. 2e is sufficient, while practical cuts provide only an upper bound. We will show in this section that the variational graphs shed light on some key properties of practical cuts when the hyperlink is regular. First, when the signal timing is regular, all alternative paths within the variational graphs to a practical cut path have a similar or a higher cost. Indeed, all shortcuts have the same duration and offsets between two successive links are identical, see for example Fig. 2c and d. Thus, practical cuts provide a tight MFD estimation for regular hyperlinks. Second, it appears that a pre-scanning step can be introduced to directly eliminate the γ values associated to non-tight cuts. This step only requires a simple visual inspection of the signal timings.

Let consider a regular hyperlink and the associated signal-timing diagram, see Fig. 4a. Geroliminis and Boyaci (2013) prove that the practical cut γ_1, i.e. $\gamma = 1$, plays a crucial role, because its intersection with the stationary cut always defines the lower bound of the range of density where the capacity is maximum (among the practical cuts). But, this cut is not sufficient to estimate the capacity in case the "stationary observer" cut is non-tight. The path P_1 associated to this cut starts from point A and reaches the end of each intermediary link i at point C_i, see Fig. 4a. It then stops until the traffic signal turns green again after a red period, e.g. at point B for link $k = 5$, see Fig. 4a. Path P_k associated to cut γ_k only stops at the end of link k (point C′) and then restarts at point B′, see Fig. 4a. We firstly look to the intersection point (K_k, Q_k) between both cuts γ_1, $Q = R(V_1) + V_1 K$, and cut γ_k, $Q = R(V_k) + V_k K$. Let T_1 (respectively T_k) be the time distance between A and B (respectively A and B′). By multiplying both cut equations by respectively T_1 and T_k and by noticing that $V_1 T_1 = V_k T_k = kl$, it comes that:

$$Q_k = \frac{R(V_1)T_1 - R(V_k)T_k}{T_1 - T_k} = \frac{N_1 - N_k}{T_1 - T_k} \tag{3}$$

with $N_1 = R(V_1)T_1$ (respectively $N_k = R(V_k)T_k$) that corresponds to the increase in N along path P_1, i.e. A → C_i → B (respectively P_k, i.e. A → C′ → B′). Let first disregard all the shortcuts, i.e. red signals are considered to be green. In that case, $N_1 - N_k$ could simply be estimated by the difference in N along the straight path B′ → B. These two points are always separated by an inte-
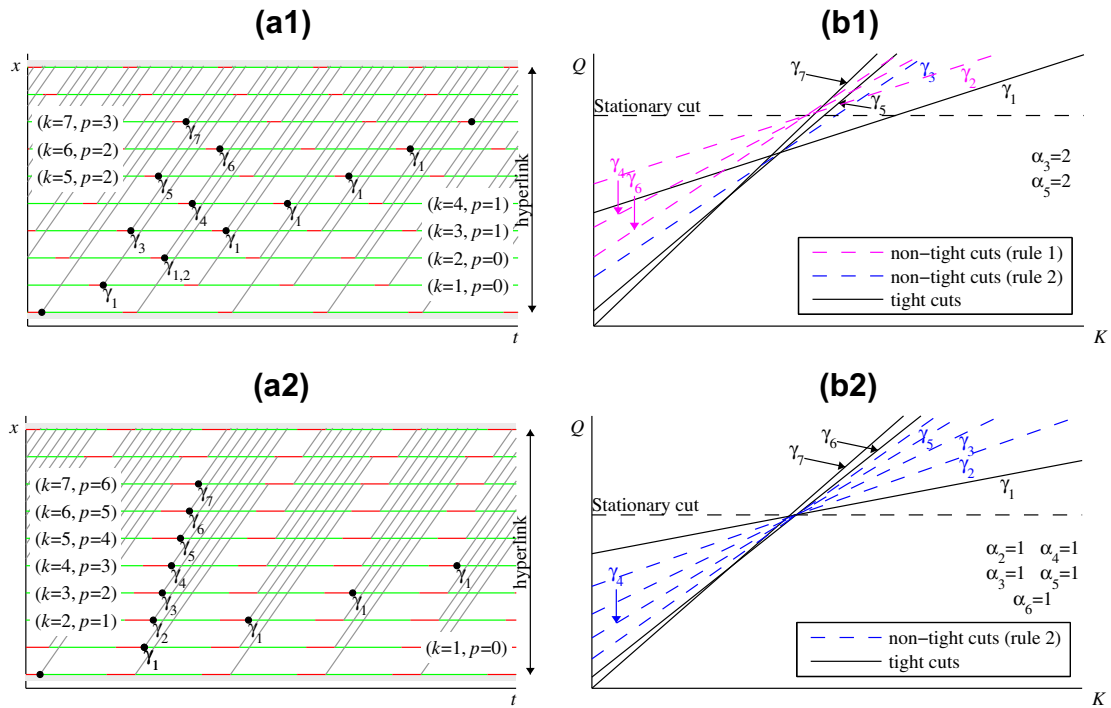
Fig. 5. (a) Signal-timing diagram with paths of all practical cuts γ_k, and (b) tight cuts selection (1) $m = 9$; $l_i = 200$ m; $C_i = 100$ s; $G_i = 85$ s; $\delta_i = i * 64.5$ s; $u = 10$ m/s; $w = 5$ m/s; $\kappa = 0.175$ veh/m (2) $m = 9$; $l_i = 200$ m; $C_i = 100$ s; $G_i = 70$ s; $\delta_i = i * 9.5$ s; $u = 10$ m/s; $w = 5$ m/s; $\kappa = 0.175$ veh/m.

ger number of signal cycles p that can be easily determined by a visual inspection of the signal-timing diagram, see Fig. 4a. An analytical expression for p can be derived, but it is tedious and beyond the scope of this paper. The difference in N along $B' \to B$ is then equal to pCs. Now, let reintroduce the shortcuts, i.e. the red phase. We should then consider that P_k experiments one shortcut between A and B' while P_1 experiments k shortcuts between A and B. Each shortcut reduces the increase in N along a path of $(C - G)s$. Thus, $N_1 - N_k = pCs - (k - 1)(C - G)s$. Finally, note that $T_1 - T_k$ is simply equal to pC. Thus, the value of Q_k is:

$$Q_k = \left(1 - \alpha_K \left(1 - \frac{G}{C}\right)\right) s \text{ with } \alpha_k = \frac{k - 1}{p} \tag{4}$$

Let A_p be the set of cuts γ_k that corresponds to the same value of p. For example, $A_1 = \{\gamma_3, \gamma_4, \gamma_5\}$ in Fig. 4a. Within set A_p, when k increases, α_k increases and thus Q_k decreases (4). Furthermore, V_k decreases. Let denote $k(p)$ the minimum value of k for a given set A_p. It appears that all cuts γ_k with $k > k(p)$ are above the lower envelope defined by cuts γ_1 and $\gamma_{k(p)}$, e.g. γ_4 or γ_5 fall above the couple (γ_1, γ_3) in Fig. 4b. They can then be disregarded as non-tight – rule 1. Note this rule also holds for A_0. In that latter case, all cuts are parallel because they have the same speed. Cut γ_1 is then always the tightest cut (among the practical cuts) because the associated path experiences the highest number of shortcuts. This first selection rule reduces the number of practical cuts to consider by half in Fig. 4b. The resulting set is thus $\{\gamma_1, \gamma_3, \gamma_6, \gamma_8\}$. This set can further be reduced by noticing that $\forall p \ V_{k(p)} \leqslant V_{k(p+1)}$. Thus, when $\alpha_{k(p+1)} \leqslant \alpha_{k(p)}$, cut $\gamma_{k(p)}$ can be disregarded – rule 2a. When $\alpha_{k(p+1)} > \alpha_{k(p)}$, cut $\gamma_{k(p+1)}$ can be disregarded only if $V_{k(p)} = V_{k(p+1)}$) – rule 2b. This second rule (rules 2a and 2b) disqualifies cut γ_6, in Fig. 4b. Note that this second rule does not hold for $\gamma_{\max}$ that should always be kept. Indeed, the associated path experiments the highest speed and a null cost. Cut $\gamma_{\max}$ then always defines a tight bound near $K = 0$. Applying both rules 1 and 2 does not always provide the minimal set of practical cuts that tightly defines the MFD but almost eliminate all the non-significant cuts.

Fig. 5 presents two more contrasted examples. For the first one, the initial set $\{\gamma_1 \ldots, \gamma_7\}$ is reduced to $\{\gamma_1, \gamma_3, \gamma_5, \gamma_7\}$ by rule 1 and further reduced to $\{\gamma_1, \gamma_5, \gamma_7\}$ by rule 2 because $\alpha_5 = \alpha_3$. Fig. 5b1 shows that this final set is minimum. The second example typically represents cases when no cuts have the same speed. In such a case, rule 1 does not eliminate any cuts but rule 2 only keeps $\gamma_{\max - 1}$ because all α_k are equal to 1. The final set is then $\{\gamma_1, \gamma_6, \gamma_7\}$, see Fig. 5b2. This case is the only one where the set can be further reduced by disregarding γ_1. Indeed, when all α_k are equal to 1, Q_k is always equal to Gs/C. Thus, all cuts intersect with the stationary cut and the relevant part of γ_1 is above this last cut, see Fig. 5b2.

We have performed extensive tests and we never found more than 3 forward moving cuts within the final set. We should also mention that the above selection method could be applied to backward moving cuts without any modification. Of course, the signal-timing diagram should be analyzed in the reverse direction with a travelling speed equal to $-w$ but Eq. (4) and rules 1 and 2 are identical. Thus, the proposed selection method directly focuses on the most relevant forward or
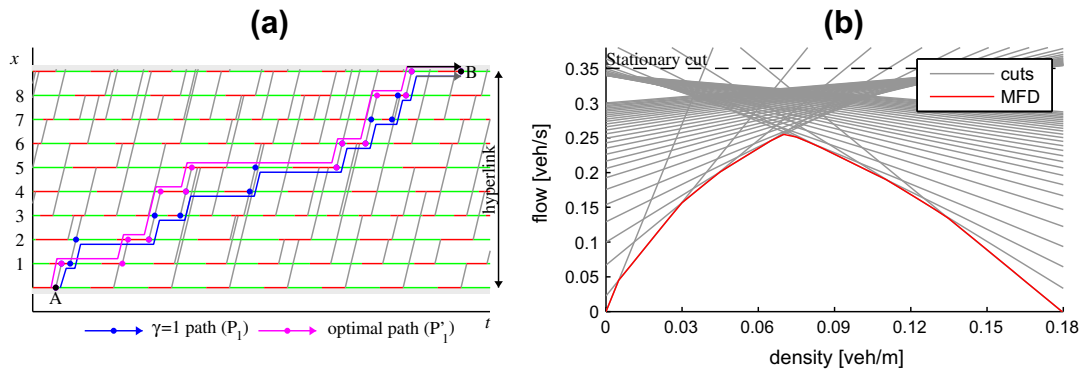
Fig. 6. (a) Forward variational graph ($m = 9$; $l_i = 100$ m; $C_1 = 90$ s; $C_2 = 140$ s; $C_3 = 55$ s; $C_4 = 110$ s; $C_5 = 190$ s; $C_6 = 80$ s; $C_7 = 45$ s; $C_8 = 130$ s; $C_9 = 100$ s; $\varepsilon = 0.6$; $G_i = \varepsilon C_i$; $\delta_1 = \delta_8 = 5$ s; $\delta_2 = \delta_7 = 21$ s; $\delta_3 = \delta_6 = 35$ s; $\delta_4 = \delta_5 = 45$ s; $\delta_9 = 0$ s; $u = 10$ m/s; $w = 5$ m/s; $\kappa = 0.175$ veh/m), and (b) relevant cuts and resulting MFD.

backward moving cuts. It may pave the way to make more efficient experimental estimation methods because it significantly reduces the required number of real observers. Note that even for regular networks, practical cuts cannot guarantee a tight MFD, because shortest paths can be defined in the merged full VT graph of Fig. 2e.

3.3. Cuts properties for irregular hyperlinks

In the general case, i.e. for irregular hyperlinks, the practical cuts do not always lead to a tight MFD estimation and might not be even well-defined. This is exemplified with the hyperlink presented in Fig. 6a. Let consider the path P_1 associated to the cut $\gamma = 1$ and that travels the hyperlink between A and B. The mean cost value along this path is 0.244 veh/s. The variational graph for the forward moving observer exhibits an alternative path P'_1 between the same points with a lower cost: 0.184 veh/s. P'_1 corresponds to the optimal (least-cost) path for the forward observer (not necessarily optimal for the merged full VT graph). The error driven by the practical cuts method for irregular hyperlinks may be really significant: here $R_{AB}(V_1)$ has been overestimated by 33%.

Fig. 6b shows the associated cuts estimated by separate forward and backward graphs. In this specific example, these cuts provide a tight MFD, same with the one obtained by the merged global VT graph. Note that we can no longer consider a single starting point for the paths because the signal-setting diagram is not regular. Here, we use ten time windows j for each V_k value to estimate $R(V_k) = \text{mean}(R_j(V_k))$. This seems sufficient because the ratio between the standard deviation of $R_j(V_k)$ and its means is always lower than 18% (respectively 6%) for all the positive (respectively negative) values of V_k. The first ten successive red endings at the hyperlink upstream border define the ten time windows.

Fig. 6b shows that the capacity of the MFD is significantly lower than the minimum capacity defined by the stationary cut (the drop is equal to 27%). Thus, the irregular signal timings generate a lot of negative interactions between signals and greatly decrease the hyperlink global efficiency. Fig. 6b also shows that the required number of cuts to define the MFD is significantly higher than for regular cases. This example is of course unrealistic because signal timings are rarely so contrasted in the real field. However, it clearly highlights that variational graphs are a powerful and efficient method to estimate MFD in all cases.

4. Route choice, user and system equilibria on parallel hyperlinks

In the sequel of the article, we will focus on a simple theoretical network composed of M parallel hyperlinks. The variational method of the previous section is used to determine the MFD on each hyperlink. We look for methods to determine the equivalent MFD (eMFD) for the whole network with different route choice assumptions. Notably, user and system optima will be investigated under steady states and dynamic conditions. Note that vehicles choose their route when entering the network and have then no other opportunity to change their decision due to the network structure. Different network structures are discussed in future work.

4.1. From MFDs to eMFD

Each hyperlink i is defined at a macroscopic level by its total length L_i, its input and output flows f_i and q_i, its accumulation n_i, its mean speed v_i and its density k_i. The MFD Q_i of hyperlink i provides the relation between the output flow and the accumulation, i.e. $q_i = Q_i(n_i)$. The travel production Z_i (Edie, 1963) is equal to q_iL_i. The total input and output flows, f and q, and the total accumulation n for the whole system are given by: $f = \Sigma f_i$, $q = \Sigma q_i$, $n = \Sigma n_i$. The total network can then be substituted for an equivalent hyperlink (e-hyperlink) whose traffic states are defined by these variables. Let denote Q the eMFD associated to this hyperlink, i.e. $q = Q(n)$. The question to tackle is to determine $Q(n)$ for different route choice hypothesis. Because

Table 1

Definition of both case studies for parallel hyperlinks ($l_{i,j}$ corresponds to the length of link j in hyperlink i).

	Case 1 ($M = 2$)	Case 2 ($M = 3$)
Hyperlink 1	$G = 35$ s; $l_{1,1} = 400$ m ($L_1 = 800$ m)	
	$\delta = 7$ s	$\delta = 23$ s
Hyperlink 2	$G = 30$ s; $l_{2,1} = 300$ m ($L_2 = 700$ m)	
	$\delta = 13$ s	$\delta = 23$ s
Hyperlink 3		$G = 27$ s; $\delta = 29$ s$l_{3,1} = 200$ m ($L_2 = 600$ m)
Common parameters	Number of links in each hyperlink: $m = 5$	
	Traffic signals: $C = 50$ s FD: $u = 13$ m/s; $w = 5$ m/s; $\kappa = 0.2$ veh/m	
	Links: $l_{i,2} = 100$ m; $l_{i,3} = 50$ m; $l_{i,4} = 200$ m; $l_{i,5} = 50$ m	
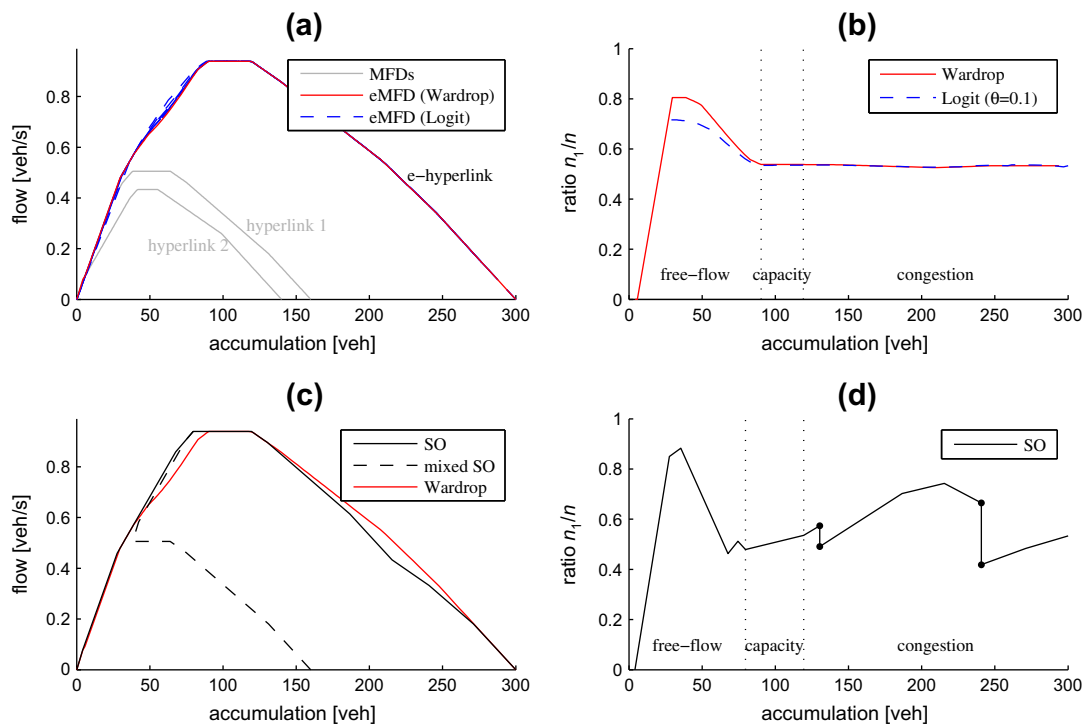
Fig. 7. Static case 1 (a) eMFD under user optima, (b) distribution of accumulation across hyperlinks for all different equilibrium states under user optima, (c) eMFD under SO, and (d) distribution of accumulation across hyperlinks for all different equilibrium states under SO.
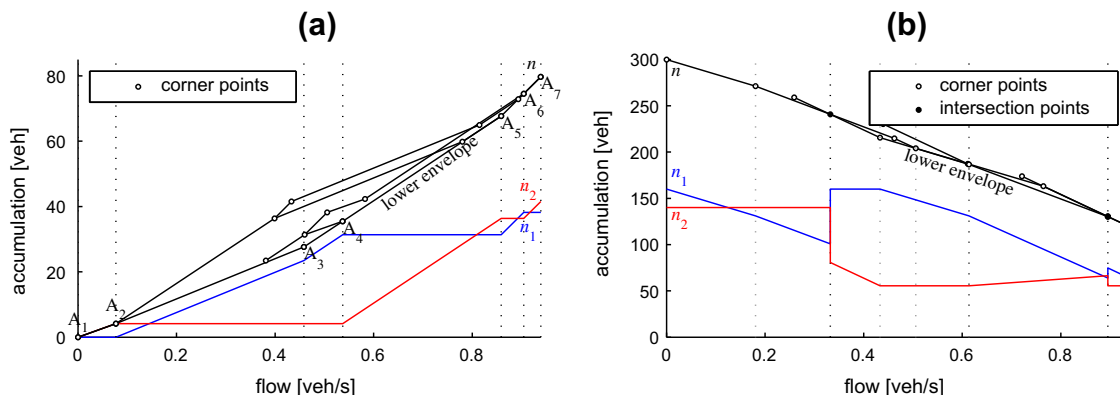
Fig. 8. Solution method for determining the eMFD under static SO (a) free-flow, and (b) congestion.
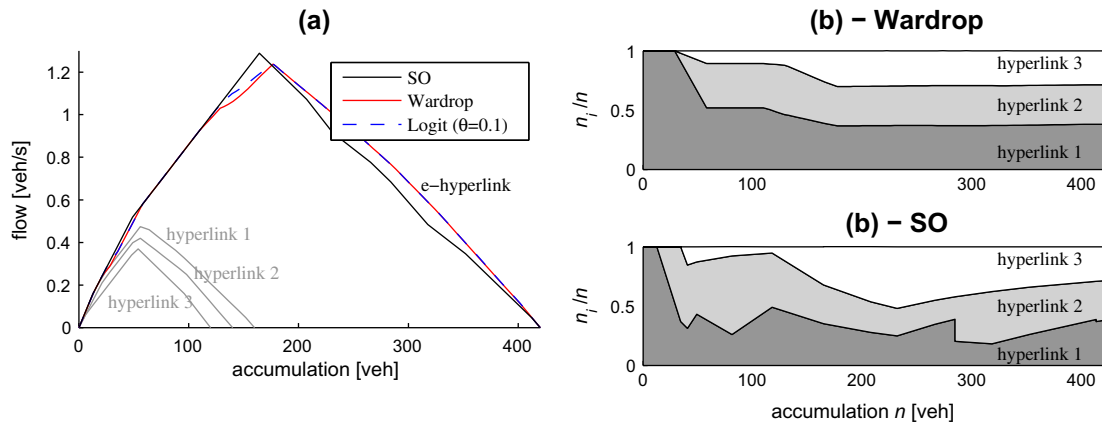
Fig. 9. Static case 2 (a) eMFD under user and system optima, and (b) distribution of accumulation across hyperlinks for all different equilibrium states.

$Q(n) = \Sigma Q_i(n_i)$, it suffices to determine the distribution of n_i (or q_i) across hyperlinks associated to any values of q both in free-flow and congested situations.

Defining the length L, the mean speed v and density k for the e-hyperlink is really appealing because the full network can then be approximated by a single hyperlink and its MFD. However, this is not trivial because all hyperlinks have different lengths. To cope this issue, we propose to fix the length of the e-hyperlink to the mean value of the hyperlink lengths, i.e. $L = (\Sigma L_i)/M$. Then, the density k can be defined as n/L. Finally, the expression of v is:

$$V = \frac{q}{k} = \frac{\sum n_i \varphi_i v_i}{\sum n_i} \text{ with } \varphi_i = L/L_i \tag{5}$$

Note that this definition of v does not correspond to the classical definition from Edie (1963) for a network due to the term φ_i. This is because we do not consider here the mean speed over the original network but over an equivalent network where all hyperlinks would have a length L and a speed $v_i' = \varphi_i v_i$ in order travel time be the same. This is necessary to fix the length of the e-hyperlink because the original Edie's definition is based on average trip length that depends on the accumulation in hyperlinks. Note that the travel production Z is not equal to qL but should be calculated as $\Sigma q_i L_i$. However, qL provides a reasonable approximation of the total travel production when the number of hyperlinks is high enough and flow q_i is not correlated with length L_i of individual hyperlinks.

Finally, the exposition in this section will be supported by two case studies. Case 1 corresponds to a network with two parallel hyperlinks while case 2 has three hyperlinks. Table 1 presents the detailed definition of both cases with notably link lengths and signal settings in each hyperlink. Note that the MFDs for each hyperlink are drawn in Figs. 7a (case 1) and 9a (case 2).

4.2. Equilibria under steady state conditions

4.2.1. User equilibria

In this section we focus on two classical models for the user equilibria: Wardrop's principle (1952) and the Logit model (Dial, 1971). The traffic states are assumed to be steady in all hyperlinks. They are then fully described by the MFDs Q_i. Such MFDs are piecewise linear because they are estimated with the variational method presented in Section 3. All Q_i are then expressed as:

$$Q_i(n_i) = \min_j (r_{i,j} + c_{i,j} n_i) \tag{6}$$

where $r_{i,j}$ and $c_{i,j}$ are respectively the y-intercept and the slope of cuts j that define the MFD Q_i. Note that the MFD is now expressed with respect to the accumulation n_i. Cut slopes provided by the variational method should then be divided by L_i to obtain $c_{i,j}$ values. The cut with the highest slope is referred with $j = 0$. Note that $v_i^* = L_i c_{i,0}$ corresponds to the highest mean speed of traffic in hyperlink i due to the signal timing. $r_{i,0}$ is always equal to 0. The reader can verify that Q_i can also be expressed with respect to v_i:

$$Q_i(v_i) = \min_j \left(q_j = \frac{r_{i,j} v_i}{v_i - L_i c_{i,j}} \text{ with } q_j \geq 0 \text{ and } r_{i,j} > 0 \right) \tag{7}$$

Let consider v_1 be the mean speed in hyperlink 1 (hyperlink label sequencing has no influence). We first derive the values of v_i in all other hyperlinks ($i > 1$) with respect to v_1 under both user equilibria. When the network satisfies the Wardrop's principle (1952), all hyperlinks with non-zero flows have the same travel time $\tau_i = L_i/v_i$. Thus, v_i is simply equal to:

$$v_i = \min\left(\frac{L_i}{L_1} v_1, v_i^*\right) \tag{8}$$

The Logit model assumes that the ratio of flows between routes i and 1 only depends on the difference in travel times between both routes:

$$q_i/q_1 = e^{-\theta(\tau_i - \tau_1)} \tag{9}$$

where θ is the parameter of the Logit model. By noticing that $q_i = Q_i(v_i)$ and $\tau_i = L_i/v_i$, (9) defines an implicit expression of v_i with respect to v_1 (the explicit solution can easily be estimated with any classical solver):

$$\frac{Q_i(v_i)}{e^{-\theta \frac{L_i}{v_i}}} = \frac{Q_1(v_1)}{e^{-\theta \frac{L_1}{v_1}}} \tag{10}$$

Eqs. (8) and (10) respectively provide the relations between any v_i and v_1 for both considered user equilibria, i.e. $v_i = F_i(v_1)$. The parametric expression of the global states q and n for the whole network can then be derived with respect to v_1:

$$\begin{cases} q = \sum_i q_i = \sum_i Q_i(v_i) = \sum_i Q_i(F_i(v_1)) \\ n = \sum_i n_i = \sum_i \frac{q_i L_i}{v_i} = \sum_i \frac{Q_i(F_i(v_1))L_i}{F_i(v_1)} \end{cases} \tag{11}$$

This defines a simple method to determine the eMFD for user optima. We only have to test all the possible values for v_1 between 0 and v_1^* and apply (11) and (8) (respectively (11) and (10)) to determine the Wardrop (respectively the Logit) equilibrium. Fig. 7a presents the resulting eMFD in case 1 and both equilibria. Several θ values from 0.05 to 0.4 with an increment of 0.05 are tested for the Logit model. Fig. 7a highlights that both equilibria only differ at the network level in the free-flow region for accumulation values between 40 and 90 veh. This is not surprising because in congestion speed values are low and then comparatively closer in all hyperlinks. The flow allocation is then nearly constant. This is confirmed by Fig. 7b that shows the evolution of the ratio n_1/n with respect to the total accumulation. This ratio stabilizes to 0.54 when traffic states reach the capacity region and remains constant in congestion. Note that this ratio smoothly varies for both equilibria. This means that the system is more likely to adapt to equilibrium when upstream demand or downstream supply change.

4.2.2. System optimum

The system optimum (SO) for M parallel hyperlinks and a steady total outflow q corresponds to the flow allocation q_i (or n_i) that is solution of:

$$\begin{cases} \min\left(\sum_i q_i \tau_i\right) = \min\left(\sum_i n_i\right) \\ \sum_i q_i = q \text{ with } q_i \geq 0 \text{ and } q_i = Q_i(n_i) \end{cases} \tag{12}$$

The system (12) cannot be solved through classical optimization technics because Q_i is non-continuously differentiable. Let consider $\{n_i^*\}_{i \in [1,M]}$ be the optimal solution of (12) for a given total flow value q^*. A small increment Δn_i of the accumulation n_i^* in hyperlink i will change the flow value:

$$Q_i(n_i^* + \Delta n_i) = Q_i(n_i) + Q_i'(n_i^*)\Delta n_i \tag{13}$$

This first order Taylor expansion is locally exact because all Q_i are piecewise linear. Let denote $\Delta n = \Sigma \Delta n_i$ the total variation of the accumulation inside of the network and Δq the associated variation of the total flow. From (13), it comes that:

$$\Delta q = \Delta n \sum_i Q'(n_i^*)\beta_i \text{ with } \beta_i = \Delta n_i / \Delta n \tag{14}$$

The optimal solution of (12) for a total flow value $q^* + \Delta q$ is given by the set of β_i values that minimizes Δn, or equivalently that maximizes $\Sigma Q'(n_i^*)\beta_i$. This quantity admits an upper bound equal to $\max(Q'(n_i^*))$ that is achieved when $\beta_{i^*} = 1$ with $i^* = \text{argmax}(Q'(n_i^*))$ because the problem is linear. Thus, it appears that when the total flow varies from Δq the optimal solution of (12) is only modified in hyperlink i^*, where the slope of the MFD is maximum. Furthermore, the variation of the accumulation in this hyperlink is equal to $\Delta n_{i^*} = \Delta q / \max(Q'(n_i^*))$. This provides an efficient solution method to determine an optimal solution from an existing one by incrementing the value of q. However, in some cases (12) has several solutions for a given flow value q. In such a case, the incremental method will only target one of these solutions. Further increments may then lead to suboptimal solutions because all the possible initial states have to be tested with the incremental method to determine the global optimum.

This solution method can be further improved to circumvent this drawback. A point $(n_i, Q_i(n_i))$ is said to be a break point if n_i corresponds to the intersection between two linear branches of Q_i, i.e. $Q_i(n_i)$ is non differentiable. A set of $\{n_i\}$ is said to be a corner point if all n_i are break points. By applying the incremental method from a corner point, it comes that the accumulation is only modified in hyperlink i^* until n_{i^*} reaches the next break point on Q_{i^*}. Indeed, all $Q_i'(n_i)$ are constant until next break point is reached. The incremental method makes then possible to connect any corner points with an optimal path that

is a segment. If a corner point is an optimal solution of (12), the connected corner point is at least a local optimum. It corresponds to a global optimum if the former point is the unique solution of (12) for the associated value of q. Thus, if we apply the incremental method from all the possible corner points we then only have to select the set of segments that define the lower envelope of n with respect to q. In practice, corner points are divided in two groups with regards to the associated traffic conditions: fully free-flow or fully congested. Note that this decomposition highly reduces the number of points to consider but only permits to estimate the SO when all hyperlinks experiment the same traffic conditions, either free-flow or congested. Mixed conditions will be addressed at the end of this subsection. Note also that each group contains a corner point that corresponds to a trivial optimal solution when $q = 0 : \{n_i^* = 0\}_{i \in [1,M]}$ and $\{n_i^* = n_{i,\max}\}_{i \in [1,M]}$ where $n_{i,\max}$ is the maximal accumulation in hyperlink i.

Let exemplify the above solution method in case 1, see Fig. 8. Fig. 8a (respectively Fig. 8b) presents all the corner points and the optimal segments for free-flow (respectively congested) conditions. These figures also show the evolutions of n_1 and n_2 with respect to q under SO. In free-flow, it appears that the lower-envelope that defines the optimal solution is always composed with complete segments and corner points. In congestion, segments may intersect when defining the lower envelope. This introduces intermediate intersection points, see Fig. 8b. Note that such intersection points are related to total flow values q when (12) has multiple solutions. For example in Fig. 8b, when q = 0.33 veh/s, (n_1 = 100.8, n_2 = 140) and (n_1 = 160, n_2 = 80.8) correspond to the same total minimal accumulation n = 240.8. The optimal evolutions of n_1 and n_2 with respect to q can then be discontinuous. Such intersection points are never observed in free-flow because the constraint domain in (12) is convex.

The solution method can then be simplified in free-flow. Starting from corner point $\{n_i = 0\}$ (see point A_1 in Fig. 8a), the incremental method can be applied to determine which hyperlink should be firstly loaded. The flow should be assigned to this hyperlink until the next break point be reached on the corresponding MFD. This defines a new corner point (see point A_2 in Fig. 8a) from where the incremental method can be again applied. By iterating this process (see points A_i in Fig. 8a), the optimal solution can be determined for any q values. Furthermore, it appears that the optimal evolutions of n_1 and n_2 with respect to q are continuous in free-flow.

Fig. 7c summarizes the eMFD in case 1 under SO. Fig. 7d shows the evolution of the ratio n_1/n with respect to the total accumulation. It appears that the eMFD is no longer concave in congestion. This is due to the existence of intersection points that drive sudden switches in the optimal allocation among hyperlinks. Such switches correspond to discontinuities in the evolution of the ratio n_1/n with respect to the total accumulation; see solid dots in Fig. 7d. Note that this evolution presents a rough pattern compared to the one observed for user equilibria. This means that a control strategy able to lead the system to the perfect SO would be hard to tune. SO should then be seen as a target that is not necessarily reachable notably in dynamic situations. Furthermore, it appears that the SO may be far from equity. For example in case 1, Fig. 8b shows that for total flow values between 0 and 0.43 veh/s, the optimal solution consists in sacrificing one hyperlink to let all the flow exit the other ($n_1 = n_{1,\max}$ or $n_2 = n_{2,\max}$).

Finally, let now deal with mixed situations when some hyperlinks of the system are congested while others are in free-flow. Let denote by I (respectively J) the set of hyperlinks that experiment free-flow (respectively congested) conditions. The eMFD for each subset can easily be determined with the previous solution method. Each subset can then be represented by the associated e-hyperlink. The problem comes down to determining the eMFD for a system of two initial hyperlinks 1 and 2 in reverse states, respectively free-flow and congestion. Let consider two situations (n_1,n_2) and (n_1^*, n_2^*) that correspond to the same total flow, i.e. $Q_1(n_1) + Q_2(n_2) = Q_1(n_1^*) + Q_2(n_2^*) = q$. Suppose now that $Q_1(n_1^*) < Q_1(n_1)$, it comes that $n_1^* < n_1$ because hyperlink 1 is in free-flow. Furthermore, $Q_2(n_2^*) > Q_2(n_2)$ and $n_2^* < n_2$ because hyperlink 2 is congested. Thus, $n_1^* + n_2^* < n_1 + n_2$. It appears that the optimal solution of (12) for q should minimize the flow in the free-flow hyperlink. When the value of q is lower that the capacity of MFD 2, all the flow should be assigned to hyperlink 2 and hyperlink 1 is empty. This optimal solution is like the Braess paradox because closing a hyperlink improves the SO. However, one should mention that vehicles that are no longer delayed inside the parallel network would surely be delayed upstream. That is why considering mixed cases for parallel network and system optimum are not really relevant in practice and why we have mainly focused on homogeneous situations. Fig. 7c illustrates the eMFD corresponding to case 1 with hyperlink 1 in free-flow and hyperlink 2 in congestion, see dashed line.

4.2.3. Comparison of steady system and user equilibria

Fig. 7c compares SO with Wardrop equilibrium in case 1. This graph shows that for low flow values (either in free-flow and in congestion) all equilibria are equivalent. The main different are observed for accumulation values between 42 and 90 veh and between 140 and 270 veh. Note that in that case, the global capacity of the system is the same for SO and Wardrop equilibria. Thus, the main difference between both equilibria is that the mean speed is always higher for SO than for Wardrop for a given value a flow.

Fig. 9a presents the eMFDs resulting from the different equilibria (SO, Wardrop and Logit with θ = 0.1) in case 2. First, it confirms that the eMFD for SO is no longer concave in congestion. Second, it also confirms that Wardrop and Logit equilibria only diverge in free-flow near the capacity ($130 \leqslant n \leqslant 174$). SO and user equilibria diverge in the same free-flow region and in congestion. Third, it appears that SO increases the maximum capacity by 4.2% compared to Wardrop and Logit (1.29 veh/s v.s. 1.23 veh/s). This can be explained by noticing that the range of accumulation when the capacity is achieved on each individual MFD is reduced to a single point. User equilibria then mix free-flow and congested traffic states in each hyperlink when reaching the capacity. SO corresponds to the situation when the system has been ideally controlled in order to maintain free-flow conditions in each hyperlink until the maximal capacity is reached.

Fig. 9a also shows that Logit equilibrium is closer to SO than Wardrop equilibrium in free-flow near the capacity ($130 \leqslant n \leqslant 174$). Such a result can also be observed in Fig. 7a and c in case 1. This result is not universal. Indeed, it can be easily demonstrated that, for the same total flow q, Logit equilibrium $\{n_{L,i}\}$ always increases the speed (and then reduces the flow) in the hyperlink i' where the highest flow is observed at Wardrop equilibrium $\{n_{W,i}\}$, i.e. $n_{L,i'} < n_{W,i'}$. Logit equilibrium then enhances the global efficiency of the system only if the slope of $Q_{i'}(n_{W,i'})$ is lower than the slopes of $Q_i(n_{W,i})$ with $i \neq i'$. Thus, for all values of θ value, Logit equilibrium is either worse or better than Wardrop equilibrium for a given value of q.

Fig. 9b presents the evolution of the ratios n_1/n and n_2/n with respect to the total accumulation for Wardrop and SO in case 2. The same conclusions as in case 1 can be drawn. The evolution of the flow distribution among hyperlinks is smooth for Wardrop and stabilizes when reaching the capacity. The corresponding evolution for SO is much more rough and presents some discontinuities in congestion. This confirms that even if SO improves the efficiency of the system in congestion (the accumulation is roughly reduced by 5–8%), it would be hard to get through a dedicated control strategy because the targeted flow distribution is not stable. Note that such a strategy should not only control input but also output flows in all hyperlinks to be effective in congestion.

4.3. Equilibria under dynamic conditions

4.3.1. User equilibrium in the dynamic case

Consider now the same parallel network but a time dependent total inflow $f(t)$ that enters the network and is distributed (chooses) to each hyperlink i, i.e. $\Sigma f_i(t) = f(t)$. Traffic conditions in each hyperlink are still expressed by the MFD, $Q_i(n_i)$. We restrict here the user equilibrium to the Wardrop's principle. In that case, travel times τ_i should be equal for all routes with non-zero flows.

$$\tau_i = \frac{n_i}{Q_i(n_i)} = \frac{L_i n_i}{Z_i(n_i)} = \tau \ \forall_i \mid f_i > 0 \tag{15}$$

Note that (15) also provides the expression of τ_i with respect to the travel production Z_i (vh-km travelled per unit time) because this quantity is often used when dealing with MFD. All the above equations can also be expressed by replacing Q_i by Z_i/L_i.

In the dynamic case where conditions change with time, the necessary conditions for $\tau_i(t)$ for an equilibrium can be expressed in an equivalent way of (15) as:

$$\tau_i(0) = \tau(0) \ \text{and} \ \frac{d\tau_i(t)}{dt} = \dot{\tau}(t) \forall t > 0 \tag{16}$$

The dynamic equation for each hyperlink, expressed as a flow conservation equation is:

$$\frac{dn_i}{dt} = f_i - Q_i(n_i) \tag{17}$$

By combining (15)–(17) and using the chain rule we get (time t is omitted from symbol n_i for simplicity):

$$\frac{d\tau_i}{dt} = \frac{d\tau_i}{dn_i} \frac{dn_i}{dt} = \left(\frac{Q_i(n_i) - n_i Q_i'(n_i)}{(Q(n_i))^2} \right) (f_i(t) - Q_i(n_i)) = \dot{\tau}(t) \forall i \tag{18}$$

Given initial conditions for accumulations $n_i(t = 0)$ and total inflow $f(t)$, one can solve at every time t a $(n+1) \times (n+1)$ linear system of equations $Ax = b$ and identify $f(t)$ and $\tau(t)$ with

$$a_i = \left(\frac{Q_i(n_i) - n_i Q_i'(n_i)}{(Q(n_i))^2} \right) \ \text{and} \ b_i = 1 - \frac{n_i Q_i'(n_i)}{Q_i(n_i)} \tag{19}$$

$$A = \begin{bmatrix} \begin{bmatrix} a_1 & & 0 \\ & & \\ 0 & & a_n \end{bmatrix} & \begin{matrix} -1 \\ \\ -1 \end{matrix} \\ \begin{matrix} 1 & \ \ \ 1 \end{matrix} & 0 \end{bmatrix} \quad x = \begin{bmatrix} f_1 \\ \\ f_n \\ \tau \end{bmatrix} \ \text{and} \ b = \begin{bmatrix} b_1 \\ \\ b_n \\ f \end{bmatrix} \tag{20}$$

This solution is attractive and computationally elegant because it proves that the dynamic case problem can be decomposed under UE conditions in different time steps and be solved in a sequential way. Given also that conditions at static and dynamic case are expressed by Wardrop's principle as shown in (15), it is expected that the two developed methodologies will provide the same eMFDs, even if the static case does not consider dynamic equations (17). This means that the dynamics of the network of M parallel hyperlinks can be described by a much simpler equation, which is the summation of all dynamic equations (17), i.e. $dn/dt = f(t) - Q(n)$, where $Q(n)$ is the eMFD for the network. Thus, one can solve the dynamic case by only considering a static solution. Nevertheless, if the dynamics of each hyperlink have to be estimated, then Eqs. (17) and (20) have to be applied. This solution will give additional insights on the evolution of inflows and split ratios $f_i(t)/f(t)$.
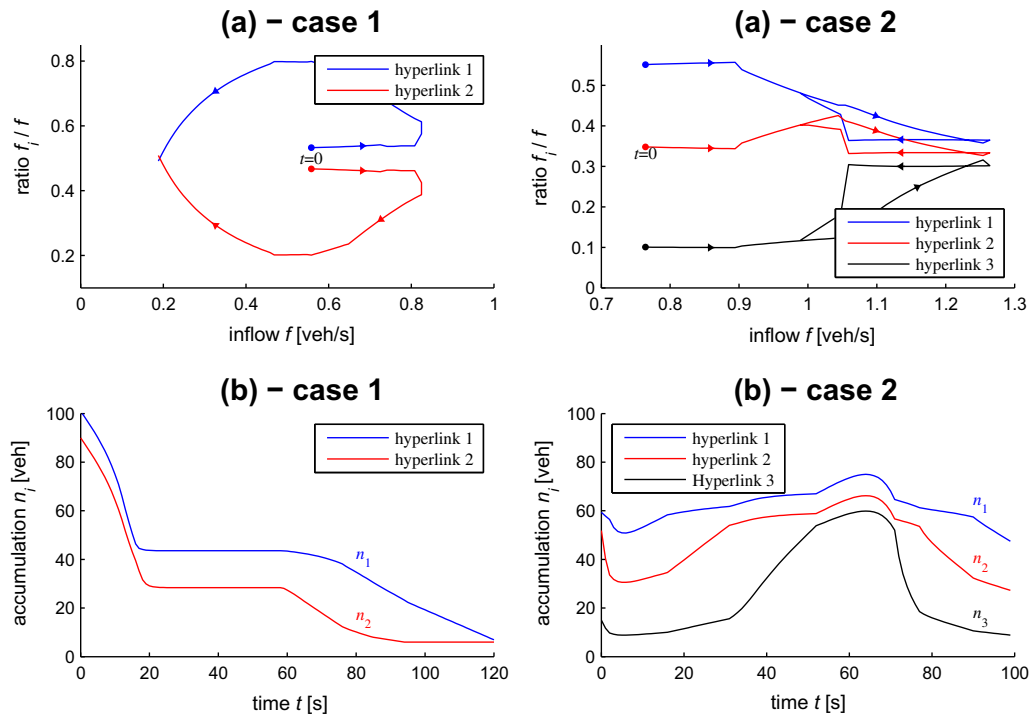
Fig. 10. Dynamic user equilibrium (a) evolution of the split ratio with respect to the total inflow, and (b) evolution of the hyperlink accumulation with respect to time.

To investigate this, let's now consider an application of the above methodological framework for the same network structures as described in Table 1 and for a trapezoidal time-dependent demand. For case 1 with $M = 2$ hyperlinks $f(t) = 0.56 + \min(0.02t, 0.27, 1.4 - 0.01t)$, while for case 2 with $M = 3$ hyperlinks $f(t) = 0.74 + \min(0.01t, 0.53, 1.13 - 0.01t)$. Case 1 starts from congested initial conditions, while case 2 starts from uncongested initial conditions. The demand profile creates states in all regions of the MFDs and eMFD for both cases. Notably, the shape of the eMFD is identical with the one presented in Figs. 7a and 9a for UE in the static case. The demand split ratios, as expressed by f_i/f are different compared to the onset and the offset of congestion, as shown in Fig. 10a1 and a2. Note for case 1 that the system starts and ends with almost equal split ratios, while for the rest of the simulation much higher split follows hyperlink 1, even for the same total inflow f, because it has a higher MFD than hyperlink 2 for the same accumulation. For case 2, conclusions are different. Note that in the offset of congestion, when the system returns from congested to uncongested states, the split ratio for hyperlink 3 is much higher for the same total inflow f, because travel times are much higher during the offset of congestion as expressed by hyperlink accumulations in Fig. 10b2. This shows that traffic networks are not memoryless and initial conditions significantly influence how congestion evolves. Note that even when accumulation is constant for a period of time ($t = 20$ to $t = 60$ in Fig. 10b1) for both links (case 1), the split ratios are completely different and a large hysteresis is observed. Conclusions for SO under static and dynamic cases are completely different. This is discussed in the next subsection.

4.3.2. System optimum in the dynamic case

While UE provides identical aggregated solutions in the static and dynamic case, SO has very different characteristics both for congested and uncongested situations. In the system optimum case (SO), total inflow $f(t)$ is distributed in the hyperlinks in such a way to maximize the total outflow or minimize the total vehicle hours travelled (VHT), similar to Eq. (12) but in the dynamic case. By expressing the dynamics in the discrete version of Eq. (17), then the optimization problem can be formulated as follows for total VHT.

$$
\begin{cases}
\max_{f_i^t} J = \sum_{t=1}^{T} \sum_{i=1}^{M} n_i^t \\
n_i^{t+1} = n_i^t + f_i^{t+1} - Q_i(n_i^t) \\
\sum_{i=1}^{M} f_i^{t+1} = f^{t+1}, f_i^{t+1} \geq 0, n_i^{t+1} \geq 0 \; \forall i, t
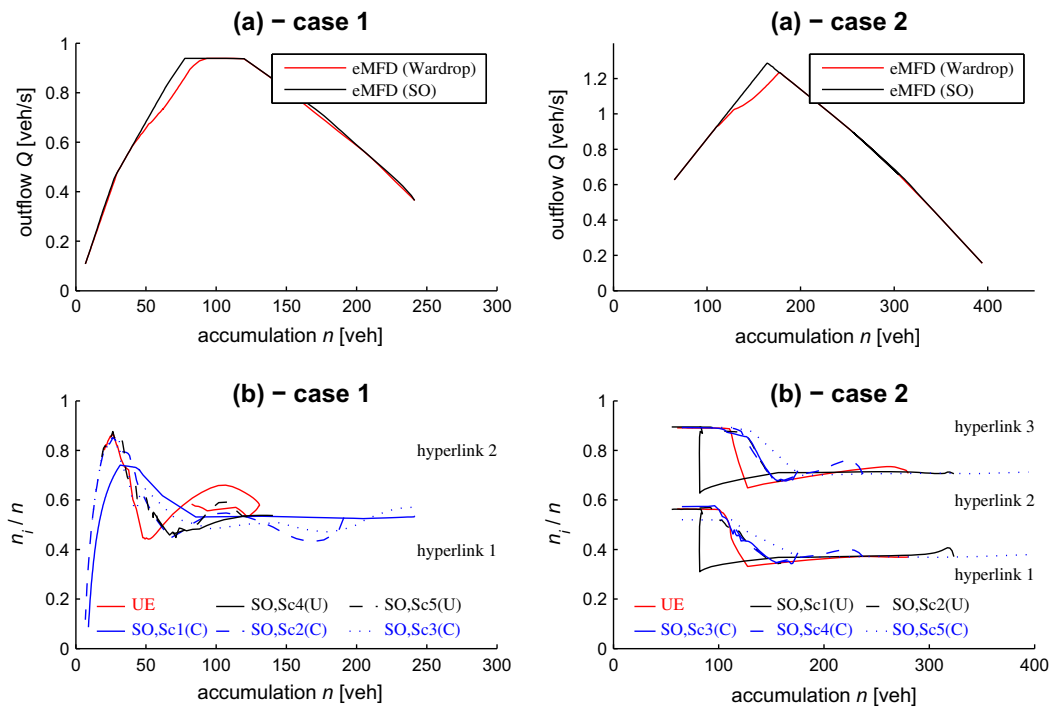\end{cases}
\tag{21}
$$

Fig. 11. Dynamic system optimum (a) eMFD (UE and SO), and (b) distribution of the accumulation across hyperlinks with respect to the total accumulation.

Problem (21) is hard to solve to optimality, as the number of local optima might be large. Nevertheless, by utilizing a multi-start search combined with a classical optimization tool (e.g. steepest decent or generalized reduced gradient (GRG) method) a close to optimal solution can be obtained, as local optima can be avoided due to multi-start search. In a more sophisticated way, one can utilize a model predictive control (MPC) approach to solve these types of problems, see for example Geroliminis et al. (2012). Alternatively, one can optimally solve problem (21) in a sequential way at each time step with a method similar to the one presented in Section 4.2 for the static SO case (maximum slope in the MFD). While, this method cannot guarantee a globally optimal solution, it is much faster to be implemented and can give good solutions if the problem is smooth. To guarantee globally optimal solution methods is beyond the scope of this work as this is a tough problem. Nevertheless, by solving a multi-start GRG method and the simpler sequential method for many different instances, results were less than 0.5% far. In the sequential method the optimal solution at time interval t, identifies initial and boundary conditions for the problem at time $t + 1$.

In the static case, accumulation n_i does not have any initial conditions as it is not governed by a dynamic equation (inflow is equal to outflow). Thus, for a given inflow, the SO solution tries to split the inflows in such a way to create the smallest possible total accumulation, and this is the reason that eMFD for SO is always on the left of the one for UE (see Figs. 7c and 9a). In the dynamic case, inflow is different than outflow and the split ratios cannot be estimated from the static case. Also, the resulting eMFD is not identical in the two cases, and it creates states of higher outflow for a given total accumulation, n. Fig. 11a presents the eMFD for the UE and SO in the dynamic case for many different scenarios for $M = 2$ and $M = 3$ hyperlinks (cases 1 and 2). While the eMFD is almost identical in the uncongested regime, it has a different behavior in the congested regime. Fig. 11b also shows (in an analogy to the static case) how the distribution of n_i/n across hyperlinks changes with accumulation n for different scenarios, i.e. different initial values and demand profiles. Note that eMFDs for UE and SO are very close in the congested regime. This means that even if SO is applied for a congested system, savings will be small. For the scenarios that the system moves from congested to uncongested regime (C → U) the ratio n_i/n is very sensitive to initial conditions and demand profile. Thus, such a strategy will be difficult to implement in reality and without significant decrease in delays. With respect to uncongested states, SO can provide substantial improvements. It is clear in Fig. 11b(1&2) that for scenarios with initial conditions uncongested moving towards congestion (U → C), a SO solution can completely avoid congested regimes and provide large savings with respect to system delays. Note that all scenarios in Fig. 11b(1&2) create significant congestion under UE. For both the cases of $M = 2$ and $M = 3$, VHT savings for U → C scenarios (Sc1, Sc2 for $M = 3$ and Sc4, Sc5 for $M = 2$) range 20–25%, while for C → U scenarios (Sc3, Sc4, Sc5 for $M = 3$ and Sc1, Sc2, Sc3 for $M = 2$) range only 5–11%. Note that in this case the SO under static conditions can provide larger improvements, but indeed it might not be an effective way to reach it in practice, except with very coercive measures where you place all the vehicles in one link. The observation of these almost unreachable states emphasizes that SO conditions are reachable and can provide travel delay improvements mainly during the network loading with uncongested initial conditions.

5. Discussion

In this paper, we further study the influence of route choice in the MFD for a network of parallel routes under different traffic conditions, static and dynamic case for user equilibrium and system optimum. In order to analytically obtain an MFD for different route topologies (link lengths, signal settings and number of lanes), we have tuned an MFD estimation based on variational theory. A sufficient graph is defined in the time–space plane (t,x) to encompass all relevant paths for moving observers. The new method provides all the necessary cuts to estimate an upper envelop of the MFD. The advantage of the method is that regularity conditions are no longer needed and this method can be directly applied for more realistic heterogeneous topologies. Furthermore, this method properly works even if the number of links inside a hyperlink is small. Note also that we can relax the conditions of the same number of lanes for all links of a hyperlink when estimating the MFD. Note finally that the condition of homothetic FD within the links can be relaxed. Indeed, one has only to adapt the slope of edges (c) in the variational graphs accordingly with the actual values of u and w in each link. The extension of the proof is based on the results provided in Section 4.2 of Daganzo's paper (2005b) about piecewise homogeneous problems (piecewise FD in space).

In the second part of the paper, we utilize the developed methodology to estimate MFDs for a number of parallel routes and combine them with different route choice rules (Wardrop's principle, Logit model and system optimum condition) to investigate how the global network performance (as expressed by an eMFD) is affected. The results are consistent with previous findings that identify the variance of link density (Mazloumian et al., 2010; Daganzo et al., 2011) as a key variable that can affect the shape of the eMFD. An interesting observation is that under static (inflow is equal to outflow for each route) and dynamic (system is governed by dynamic equations (17)) conditions the resulting eMFDs for UE under Wardrop's principle are identical. Nevertheless, the dynamic analysis provides further insights about the split ratios. It is clear that traffic networks, especially under capacity and congested conditions are not memoryless and a dynamic analysis is necessary to identify the traffic evolution, which can be very different in the onset and the offset of congestion even for the same demand values. With respect to the system optimum static and dynamic approach provide different results especially in the congested regime. Our findings indicate that under a dynamic approach, eMFDs for UE and SO are very close during congested conditions in the decreasing part of the curve. This means that even if SO is applied for a congested system, savings will be small. Nevertheless, the SO under static conditions can provide larger improvements even in the congested regime, but it might not be an effective way to reach it in practice, except with very aggressive and inequitable strategies, where most of the vehicles are forced to travel in one link. The observation of these unreachable states, emphasizes that SO conditions can provide effective and realistic travel delay improvements mainly during the network loading with uncongested initial conditions.

Ongoing research considers further extensions of this work for more complicated network structures with a system of serial and parallel hyperlinks. This task is challenging for various reasons. First, a larger number of alternative paths have to be evaluated to identify equilibrium conditions. Computational issues should be investigated, but it is expected that this problem will still have much smaller number of alternatives compared with the classical link-based approach. Second, methodologies to partition a network in hyperlinks with homogeneous level of congestion (to create low scatter in the MFDs) should be developed. The reason is that variational theory considers a series of links with similar density and errors might occur when applying this methodology for an inhomogeneous distribution of density. Nevertheless, congestion is not randomly distributed due to spatial correlations, i.e. a series of successive intersections load the main route with net positive inflows from cross streets. Thus, it is expected that heterogeneously congested transportation networks can be partitioned in a number of homogeneous clusters. Towards this direction, static clustering algorithms have been developed for heterogeneous networks (Ji and Geroliminis, 2012), but further research is needed for cases where the shape of the clusters changes over time. Third, spatiotemporal propagation of traffic congestion at the network level is not well-studied, while there is a vast literature (shockwave analysis, higher order traffic models, etc.) for freeway propagation. Towards this direction, the concept of shockwave analysis can be revisited, where (after partitioning), traffic state dynamics for serial hyperlinks can be described with generalized MFD shockwaves. This research is underway.

References

Cassidy, M.J., 1998. Bivariate relations in nearly stationary highway traffic. Transportation Research Part B 32 (1), 49–59.
Chiabaut, N., Leclercq, L., 2011. Wave velocity estimation through CVC-curve automatic analysis. Transportation Research Record 2249, 1–6.
Courbon, T., Leclercq, L., 2011. Cross-comparison of macroscopic fundamental diagram estimation methods. Procedia – Social and Behavioral Sciences 20, 417–426.
Daganzo, C.F., 2013. Private communication.
Daganzo, C.F., Gayah, V.V., Gonzales, E.J., 2011. Macroscopic relations of urban traffic variables: bifurcations, multivaluedness and instability. Transportation Research Part B 45 (1), 278–288.

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., 2007. Urban gridlock: macroscopic modeling 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., Menendez, M., 2005. A variational formulation of kinematic waves: bottlenecks properties and examples. In: Mahmassani, H.S. (Ed.), 16th ISTTT. Elsevier, London, pp. 345–364.

Daganzo, C.F., 1995. The cell transmission model, Part II: network traffic. Transportation Research Part B 29 (2), 79–93.

Dial, R.B., 1971. A probabilistic multipath traffic assignment model which obviates path enumeration. Transportation Research 5 (2), 83–111.

Edie, L.C., 1963. Discussion of traffic stream measurements and definitions. In: Almond, J. (Ed.), Proceedings of the 2nd International Symposium on the Theory of Traffic Flow. OECD, Paris, France, pp. 139–154.

Keyvan-Ekbatani, M., Kouvelas, A., Papamichail, I., Papageorgiou, M., 2012. Exploiting the fundamental diagram of urban networks for feedback-based gating. Transportation Research Part B 46 (10), 1393–1403.

Geroliminis, N., Haddad, J., Ramezani, M., 2013. Optimal perimeter control for two urban regions with macroscopic fundamental diagrams: a model predictive approach. IEEE Transactions on Intelligent Transportation Systems 14 (1), 348–359.

Geroliminis, N., Boyaci, B., 2013. The effect of variability of urban systems characteristics in the network capacity. Transportation Research Part B 46 (10), 1576–1590.

Geroliminis, N., Daganzo, C.F., 2008. Existence of urban-scale macroscopic fundamental diagrams: some experimental findings. Transportation Research Part B 42 (9), 759–770.

Geroliminis, N., Sun, J., 2011. Properties of a well-defined macroscopic fundamental diagram for urban traffic. Transportation Research Part B 45 (3), 605–617.

Godfrey, J., 1969. The mechanism of a road network. Traffic Engineering and Control 11 (7), 323–327.

Haddad, J., Ramezani, M., Geroliminis, N., 2013. Cooperative traffic control of a mixed network with two urban regions and a freeway. Transportation Research Part B 54, 17–36.

Haddad, J., Geroliminis, N., 2012. On the stability of traffic control in two-region urban cities. Transportation Research Part B 46 (9), 1159–1176.

Herman, R., Prigogine, I., 1979. A two-fluid approach to town traffic. Science 204, 148–151.

Ji, Y., Geroliminis, N., 2012. On the spatial partitioning of urban transportation networks. Transportation Research Part B 46 (10), 1639–1656.

Ji, Y., Daamen, W., Hoogendoorn, S., Hoogendoorn-Lanser, S., Qian, X., 2010. Macroscopic fundamental diagram: investigating its shape using simulation data. Transportation Research Record 2161, 42–48.

Knoop, V.L., Hoogendoorn, S.P., Van Lint, J.W.C., 2012. Routing strategies based on the macroscopic fundamental diagram. Transportation Research Record 2315, 1–10.

Laval, J.A., 2010. The effect of signal timing and network irregularities in the macroscopic fundamental diagram. In: Proceedings of the Summer Meeting of the TFTC TRB Committee, July 7–9, Annecy, France, 5 p.

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.

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.

Mazloumian, 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.

Richards, P.I., 1956. Shockwaves on the highway. Operations Research 4, 42–51.

Saberi, M., Mahmassani, H.S., 2012. Exploring the Properties of Network-wide Flow-Density Relations in a Freeway Network. Transportation Research Record 2315, 153–163.

Van Lint, J.W.C., Hoogendoorn, S., 2010. A robust and efficient method for fusing heterogeneous data from traffic sensors on freeways. Computer-Aided Civil and Infrastructure Engineering 25 (8), 596–612.

Wardrop, J.G., 1952. Some theoretical aspects of road traffic research. Institution of Civil Engineers 2, 325–378.