22nd International Symposium on Transportation and Traffic Theory

The morning commute in urban areas with heterogeneous trip lengths

Raphaël Lamotte^a, Nikolas Geroliminis^{a,*}

*^aSchool of Architecture, Civil and Environmental Engineering,
Urban Transport Systems Laboratory
Ecole Polytechnique Fédérale de Lausanne (EPFL)
GC C2 389, Station 18, 1015 Lausanne, Switzerland*

Abstract

This paper investigates the equilibrium properties of the morning commute problem at the network level with heterogeneous trip lengths. Congestion is modeled with a Macroscopic Fundamental Diagram relating the space-mean speed of a network to the vehicular accumulation. It is shown for a large class of scheduling preferences that if users have continuously distributed characteristics, the network accumulation at equilibrium is a continuous function of time. With $\alpha - \beta - \gamma$ preferences and under certain conditions, a partial First-In, First-Out (FIFO) pattern emerges at equilibrium among early and late users. This FIFO pattern is strict only within families of users having heterogeneous trip lengths and identical preferences, or vice versa. Simulation results confirmed that an attracting steady-state exists for a wide range of demands and that the predicted patterns are indeed observed.

© 2017 The Authors. Elsevier B.V. All rights reserved.

Peer review under responsibility of the scientific committee of the 22nd International Symposium on Transportation and Traffic Theory.

Keywords: traffic congestion, departure time, trip length, scheduling preferences, equilibrium, pricing

1. Introduction

The morning commute has historically been studied with point bottlenecks, following the seminal paper of Vickrey (1969). Urban networks however, cannot be modeled as collections of independent point bottlenecks. In fact, congestion propagates from one bottleneck to its neighbors, creating connected components of congestion that grow and may extend to the whole network (Ji et al., 2014). In the face of this complexity, an attractive solution consists in changing the modeling scale, from a point bottleneck to the network level (Small and Chu, 2003; Geroliminis and Levinson, 2009). This approach relies on empirically supported relationships (see e.g. Geroliminis and Daganzo (2008) and Buisson and Ladier (2009)), referred to as Macroscopic Fundamental Diagrams (MFD), which describe the dynamics

* Corresponding author. Tel.: +41-216-932-481.

E-mail address: nikolas.geroliminis@epfl.ch

of congestion under some homogeneity conditions with just a few variables such as accumulation, space-mean flow, trip completion rate and speed.

The consequences of this change of scale are, however, not fully understood yet. The first works combining an MFD with departure time choice called attention to the cost of hypercongestion, i.e. the phenomenon by which vehicle flow decreases with accumulation when accumulation exceeds a critical level (Small and Chu, 2003). Geroliminis and Levinson (2009) and Fosgerau and Small (2013) argued that by maintaining the system at the flow-maximizing accumulation, the benefits of congestion pricing with homogeneous users could be even greater than with Vickrey's bottleneck, as the duration of the peak hour could be shortened. Arnott (2013) showed that further gains can be obtained by maintaining the accumulation always *below* its flow-maximizing value, at a level that increases with the peak duration. While they reach similar conclusions, the aforementioned papers utilized different assumptions to deal with complex dynamics involving endogenous delays. Small and Chu (2003) and Geroliminis and Levinson (2009) assumed that travel time is entirely determined by the conditions at a single instant (e.g. at the arrival), Fosgerau and Small (2013) considered a piece-wise constant decreasing branch for the MFD of flow versus accumulation and Arnott (2013) assumed that at any time, all users have the same probability of exiting the network, independently of the time at which they started their trip. Yet, it is only very recently that Fosgerau (2015) and Daganzo and Lehe (2015) recognized how trip length heterogeneity challenges the fundamental FIFO assumption and started investigating its impacts on the morning commute.

Surprisingly however, Fosgerau (2015) and Daganzo and Lehe (2015) reached very different conclusions on the role of trip length. Using mathematically convenient but unconventional exponential-type scheduling preferences, Fosgerau (2015) showed that under some assumptions, the user equilibrium exhibits the so-called regular sorting property, i.e. two users differing only by their trip length sort according to a Last-In, First-Out (LIFO) pattern where the user with the longest trip starts earlier and finishes later. On the other hand, Daganzo and Lehe (2015) proved for the more conventional $\alpha - \beta - \gamma$ scheduling preferences but with a less realistic congestion mechanism (actually very similar to a point bottleneck) that the social optimum exhibits the FIFO property. Based on these theoretical considerations, Daganzo and Lehe (2015) also proposed a usage-based toll maintaining the accumulation near its flow-maximizing value and demonstrated numerically its benefits on the same realistic congestion mechanism as in Fosgerau (2015). The question of the prevailing sorting pattern with heterogeneous trip lengths remains not fully solved, despite its crucial role in congestion management.

This paper investigates the morning commute problem with inelastic demand and an MFD relating speed to accumulation. Our focus is primarily on the impact of trip length heterogeneity, but we also study the impact of heterogeneity in the scheduling preferences. In line with most of the literature, we rely on the so-called “fluid approximation” (Newell, 1982), in which the stochastic dynamics of a large number of agents are modeled by a deterministic real-valued process. With Vickrey's constant capacity bottleneck, this assumption, together with some convexity assumptions on the schedule penalties, permitted to prove the existence and the uniqueness of an equilibrium distribution of arrival times (Smith, 1984; Daganzo, 1985). In this paper, we leave aside the existence and uniqueness questions as they would require tedious derivations beyond the scope of the paper (even for a bottleneck model with much simpler dynamics, the proofs in Smith (1984) and Daganzo (1985) contain a significant number of modeling assumptions, but also very long derivations). Instead, we analytically characterize an hypothetical equilibrium and confirm using simulations that (i) an attracting steady-state exists for a wide range of demands and that (ii) the patterns theoretically predicted are indeed observed.

The complexity of the problem is mainly due to an endogenous delay in the dynamical model (more information is provided in Section 2.1). This makes a fully analytical solution intractable, as also pointed out by Arnott (2013). Nevertheless, in this work we are able to mathematically prove important properties of the equilibrium solution and develop a numerical simulation of the detailed model that confirms some of the proofs and provides further insight. More specifically, this paper has three main contributions. First, we show that with a continuum of users having continuously distributed characteristics, accumulation and speed are continuous functions of time. This result is valid for a large class of scheduling preferences and contrasts with the discontinuities resulting from the assumption of homogeneous trip lengths (Arnott et al., 2016). Second, we assume $\alpha - \beta - \gamma$ preferences and, under certain conditions, demonstrate that they result in a partial FIFO pattern among early and late users at equilibrium. This FIFO pattern is strict only within families of users having identical $\alpha - \beta - \gamma$ preferences and heterogeneous trip lengths, or vice versa. Yet, simulation results indicate that it influences the overall properties of the equilibrium, and in particular

the cost function. It is also shown that with this more detailed dynamic model, the proportion of on-time users may be larger than previously found under simpler approximations and that, under rapid demand variations, the social optimum may exhibit hypercongestion. This finding goes against the previously proposed pricing strategies and calls for further research on congestion pricing under rapidly varying demand.

Section 2 introduces the models of congestion and departure time choice, Sections 3 and 4 provide analytical treatments of the continuity and sorting properties, and Section 5 presents the simulation results.

2. Background

A typical morning commute is a stochastic process involving many thousands of agents. When dealing with such large numbers, it is common to rely on the so-called “fluid approximation”. Within this approximation, the number of users is considered as a real-valued variable, and shall be therefore referred to as the “mass” of users. This approximation greatly simplifies the study of the dynamics as accumulation can then potentially evolve continuously, and not necessarily with very small steps of one user. When a strictly positive mass of users departs or arrives simultaneously, the cumulative number of departures or arrivals exhibits a discontinuity. We refer to such phenomena as “mass departures” and “mass arrivals”. All the congestion models described hereafter rely on the fluid approximation.

2.1. Congestion models

Different congestion mechanisms require different models. Vickrey’s constant capacity bottleneck might be considered appropriate for point bottlenecks such as toll plazas. Congestible facilities however often exhibit hypercongestion. The idea of an MFD with an optimum accumulation was initiated by Godfrey (1969) and similar approaches were introduced later by Herman and Prigogine (1979), Mahmassani et al. (1984) and Daganzo (2007). Geroliminis and Daganzo (2008) was among the first works to empirically observe such macroscopic relationships at the level of a downtown neighborhood with dynamic features.

These works have produced two fundamental results. First, the speed $v(t)$, or equivalently, the network flow (or “production”) $P(t) = n(t)v(t)$, follows a well defined function of the accumulation $n(t)$ (the number of vehicles currently driving in the network) independently of the time dimension. This first result is referred to as the “speed-MFD” (or equivalently, as the “production-MFD”). In agreement with real-world observations, the speed-MFD function $\mathcal{V}(n)$ is assumed in this work to be positive, continuously differentiable and strictly decreasing on an interval $[0, n_{\text{jam}}]$. The free-flow speed $\mathcal{V}(0)$ is denoted v_f . The associated production-MFD is a unimodal curve, reaching its maximum at the so-called “flow-maximizing” or “critical” accumulation.

Second, the ratio of the production divided by the trip completion rate (or “outflow”) remains approximately constant and is equal to the (static) arithmetic average trip length of the trip-generating process. Thus, there is also a well-defined relation between the outflow and the accumulation. This second result is referred to as the “outflow-MFD” or “Network Exit Function”. Theoretical considerations suggest that the outflow-MFD holds if (i) the production-MFD exists and (ii) conditions are slowly-evolving. While there are empirical and simulation-based verifications of such assumptions, this should not be considered a universal law. Strong fluctuations in the demand that create fast evolving transient states, spatial heterogeneity of congestion or route choice effects can influence the trip length distribution of vehicles in the network (and the space-mean trip length) and the ratio of production over outflow might exhibit some variations (Mahmassani et al., 2013; Yildirimoglu and Geroliminis, 2014). Yet, if the aforementioned two assumptions hold, then the conservation law takes a straight-forward form: $\dot{n}(t) = I(t) - O(n(t))$, where $n(t)$ is the accumulation of vehicles inside the zone at time t , $\dot{n}$ denotes the time derivative of accumulation, $O(n)$ is the function described by the outflow-MFD and $I(t)$ is the inflow rate.

If conditions are changing rapidly however, the outflow-MFD might not provide a good approximation as it is “memory-less”, i.e. it ignores the history of the system. If, for instance, there is a sharp increase in the inflow in the uncongested regime, the accumulation increases instantaneously, so the outflow predicted by the outflow-MFD increases as well. Intuition suggests that the outflow should only increase after the newly entered users have completed their trip. For a user with trip length l_0 entering at time t_0 , this would take a time τ_0 that is uniquely determined by

$$\int_{t_0}^{t_0+\tau_0} \mathcal{V}(n(u)) du = l_0. \quad (1)$$

This might actually form the basis for an alternative zone-based dynamical modeling based on the speed-MFD. Consider the flow entering the zone $I(t)$ as given for all times t and assume that $I(t) = 0$ for all times $t < 0$. If the times of exits follow Eq. (1), then the accumulation at time t is given by

$$n(t) = \int_0^t I(s) \left(1 - F_{l,s} \left(\int_s^t \mathcal{V}(n(u)) du \right) \right) ds, \quad (2)$$

where $F_{l,s}(\cdot)$ is the cumulative distribution function of the trip length for users entering the network at time s (the associated probability density function is denoted $f_{l,s}(\cdot)$). Differentiating this equation leads to¹:

$$\dot{n}(t) = I(t) - \mathcal{V}(n(t)) \int_0^t I(s) f_{l,s} \left(\int_s^t \mathcal{V}(n(u)) du \right) ds. \quad (3)$$

Note that the outflow rate depends not only on the current value of accumulation but also on all the previous values.

As highlighted by Eq. (3), the phenomena occurring with a speed-MFD may be difficult to analyze in the time domain. Instead, it is often convenient to change the frame of reference and to study the dynamics of congestion over some distance measure. More specifically, we can introduce the bijection $f : \mathbb{R} \rightarrow \mathbb{R}$ given by $f(t) = \int_0^t v(s) ds$, which associates to a time t the distance traveled by a virtual user from the origin of time to t .² In the same fashion as we use t to denote a particular time, let $x = f(t)$ denote a particular distance traveled and let V be the speed function in the x space: $V(x) = v(f^{-1}(x))$.

We can now define the function $\tau(t_a, l)$ that associates to an arrival time $t_a \in \mathbb{R}$ and a trip length $l \in \mathbb{R}^+$ the corresponding travel time. In the time domain, $\tau(t_a, l)$ is only implicitly specified by the constraint $\int_{t_a - \tau(t_a, l)}^{t_a} v(s) ds = l$. In the distance domain however, τ admits an explicit expression:

$$\tau(t_a, l) = \int_{f(t_a) - l}^{f(t_a)} \frac{1}{V(s)} ds. \quad (4)$$

Note that f is continuous and since V is positive almost everywhere at equilibrium, τ is continuous as well. In addition, if $v(t)$ is continuous, then $V(x)$ is also continuous and τ is continuously differentiable. In particular,

$$\frac{\partial \tau}{\partial t_a}(t_a, l) = \frac{v(t_d) - v(t_a)}{v(t_d)}, \quad (5)$$

where the departure time t_d is actually a function of t_a and l , given by $t_d(t_a, l) = t_a - \tau(t_a, l)$.

Small and Chu (2003), Geroliminis and Levinson (2009), Fosgerau and Small (2013) and Arnott (2013) all adopted dynamics based on outflow-MFDs. Although they did not explicitly state the conservation law (Eq. (3)) or the outflow expression, Fosgerau (2015) and Daganzo and Lehe (2015) also used dynamics based on the speed-MFD, respectively for theoretical derivations and simulations.

2.2. Scheduling preferences

The simplest way to think about scheduling preferences is probably in terms of marginal utility of time. Every minute spent at home (resp. at work) has some value $h(t)$ (resp. $w(t)$) that depends on the time t of the day. Typically, the marginal utility of time at work increases sharply around the official work starting time. The marginal utility of time spent driving is often assumed to be zero. With such an assumption, the utility associated to the arrival time t_a for a user with trip length l is given by³:

$$U_l(t_a) = \int_0^{t_d(t_a, l)} h(s) ds + \int_{t_a}^0 w(s) ds. \quad (6)$$

¹ Note that this equation slightly differs from the one in Lamotte and Geroliminis (2016), where the term $\mathcal{V}(n(t))$ had been mistakenly omitted. The numerical results from the event-based simulation were not impacted as they relied on a formulation that is equivalent to Eq. (2).

² Note that f might not be bijective in some out-of-equilibrium situations reaching (or converging towards) gridlock, but we discard these pathological cases to focus on the equilibrium properties.

³ Utility is only defined up to a constant. It might be more intuitive for instance to use the beginning of the day as the lower limit of the integral of h and e.g. midday as the upper limit of the integral of w .

Since $t_d(t_a, l)$ is continuous, $U_l(t_a)$ is also continuous. The same preferences can also be expressed in terms of travel time (TT) cost and schedule delay (SD) cost. The user cost is then

$$C_l(t_a) = -U_l(t_a) = \text{TT}(t_d(t_a, l), t_a) + \text{SD}(t_a), \quad (7)$$

where $\text{TT}(t_d, t_a) = \int_{t_d}^{t_a} h(s) ds$ and $\text{SD}(t_a) = \int_0^{t_a} w(s) - h(s) ds$.

If the marginal utility of time spent at home is constant ($h(t) = \alpha$) and the marginal utility of time spent at work is piece-wise constant ($w(t) = \alpha - \beta$ for all times $t < t^*$ and $w(t) = \alpha + \gamma$ for all times $t \geq t^*$), the formulation with marginal utilities of time boils down to the widely used $\alpha - \beta - \gamma$ preferences (Arnott et al., 1990):

$$\text{TT}_{\alpha-\beta-\gamma}(t_d, t_a) = \alpha(t_a - t_d) \quad (8a)$$

$$\text{SD}_{\alpha-\beta-\gamma}(t_a, t^*, \beta, \gamma) = \max(\gamma(t_a - t^*), \beta(t^* - t_a)). \quad (8b)$$

The set of $\alpha - \beta - \gamma$ preferences is simply a sub-set of the preferences that can be described by the marginal utility functions $h(t)$ and $w(t)$.

The use of constant and piecewise constant marginal utilities of time has often been questioned. In fact, the empirical evaluations of marginal utilities systematically relied on more complex models than the simple $\alpha - \beta - \gamma$ preferences. Small (1982) allowed for a discrete jump in utility for late arrivals and for a flexible range of arrival times with lower penalties. Hendrickson and Plank (1984) estimated two quadratic penalty functions for early and late arrivals, which are both equivalent to affine marginal utility functions. Tseng and Verhoef (2008) estimated marginal utilities at different times of the day in a non-parametric way, and Hjorth et al. (2015) used exponential marginal utility functions. However, when comparing different models with stated preference data, Hjorth et al. (2015) found that piecewise constant and affine specifications outperformed the exponential ones. While all agree that constant marginal utilities are not ideal, there is no consensus on the parametric form that should be used. For instance, Hendrickson and Plank (1984) found a marginal utility of time at work that decreases with the time of day while the results of the other works mentioned above suggest an increase.

The continuity results in Section 3 are obtained with only minor requirements on the scheduling preferences. Then, the subsequent results assume $\alpha - \beta - \gamma$ preferences for simplicity but numerical applications in Section 5 suggest that the results obtained are also valid for a larger class of similar but smooth utility functions.

3. Continuity of accumulation over time

Besides being extremely numerous, commuters are also very diverse. This heterogeneity is often neglected for the sake of tractability and intuition. Yet, this section shows that trip length heterogeneity radically changes the equilibrium patterns when considering a system governed by a speed-MFD. Indeed, Arnott et al. (2016) showed that with homogeneous trip length, there exists an equilibrium which exhibits piece-wise constant accumulation and non-overlapping trip intervals. We demonstrate that when users have continuously distributed characteristics and especially trip lengths, such an equilibrium cannot occur as accumulation should be continuous over time. More specifically, the continuity result demonstrated in this section requires the following assumption.

Assumption 1 (Distributed characteristics). *Assume that:*

- All users have marginal utilities of time at home and at work $h(t, \Theta)$ and $w(t, \Theta, T^*)$, where Θ is a vector of individual-specific parameters and T^* is a continuously distributed scalar parameter.
- For all Θ , $t \rightarrow h(t, \Theta)$ is positive and continuous everywhere and for all Θ and T^* , $t \rightarrow w(t, \Theta, T^*)$ is continuous everywhere, except possibly at $t = T^*$.
- Trip length and trip length conditioned on T^* are continuously distributed variables.

Although its statement is technical, Assumption 1 is actually quite reasonable. It simply requires continuous marginal utility functions (but still allows for a discontinuity in the marginal utility at work at a single specific time T^*) and that some user characteristics (trip length, T^* and trip length conditioned on T^*) are continuously distributed. Note that continuous distributions are overwhelmingly common in nature and there are strong reasons to believe that

trip length and T^* are no exception. Indeed, the trip length distribution arises from a very large number of pairs of origins and destinations (primarily home and work locations). Similarly, the distribution of desired exit times from the road network (when the marginal utility is likely to be discontinuous) depends on both the distribution of official work starting times (which is admittedly not continuously distributed) and on the (continuous) distribution of time spans required to reach the workstations from the time the vehicle exits the road network modeled by the MFD (e.g. when entering a parking lot). Note finally that the continuity result obtained is also corroborated by many empirical studies (see for example Geroliminis and Daganzo (2008); Buisson and Ladier (2009)).

To prove continuity of accumulation, we prove that there cannot be mass departures or mass arrivals at equilibrium, or that they have to exactly cancel out each other. This result only makes sense in the context of the fluid approximation as it requires that a single agent does not have a mass, but only a density.⁴

We first need to describe the framework considered and define some basic notions. A *realization* of the morning commute is an allocation of departure times (or equivalently, arrival times) to all users. An *equilibrium* is a realization such that no user can reduce her cost by unilaterally changing her decision. A *socially optimal realization* minimizes the sum of the costs of all users, excluding potential tolls (which are considered pure financial transfers).

Consider a population of mass $N > 0$. In general, an allocation of departure times or arrival times might be such that there are masses of users that depart or arrive simultaneously. If mass departures and mass arrivals do not exactly cancel out, they create discontinuities in the accumulation. For any realization, we can define the set of times T_{dis} for which such discontinuities occur and m_i the (relative) change in accumulation at time $i \in T_{\text{dis}}$. If, for instance, there is a mass departure at time i but no mass arrival, then m_i is the mass of departures at time i . If there is a mass arrival but no mass departure, then the mass of arrivals is given by $-m_i$. If we further define $d_c(t)$ and $a_c(t)$ as the continuous components of the departures and arrivals, the variation in accumulation during some period $[t_1, t_2]$ is given by $\int_{t_1}^{t_2} d_c(t) - a_c(t) dt + \sum_{i \in T_{\text{dis}} \cap [t_1, t_2]} m_i$. We now introduce two Lemmas that are useful to prove the continuity of accumulation in Proposition 1.

Lemma 1. *If trip length is continuously distributed, then for any realization of the morning commute, the set of users that have discontinuities both at their departure and arrival has Lebesgue measure zero.*

Proof. Clearly $\sum_{i \in T_{\text{dis}}} |m_i| \leq 2N$, so the set $\{m_i, i \in T_{\text{dis}}\}$ is summable, and consequently countable. As there is a bijection between each element of $\{m_i, i \in T_{\text{dis}}\}$ and T_{dis} , T_{dis} is countable as well, so its Lebesgue measure (in the time domain) is zero.

Now, one can associate to any pair of times (t_1, t_2) in $(T_{\text{dis}})^2$ the distance separating them: $\int_{t_1}^{t_2} v(u) du$. Since T_{dis} is countable, the set of distances separating two events of T_{dis} is also countable. Since trip length is a continuously distributed variable, the set of users having exactly these trip lengths is of measure zero. $\square$

Lemma 2. *Left and right-hand limits of the the accumulation exist everywhere for any realization of the morning commute.*

Proof. The variation in accumulation during some period $[t_1, t_2]$ is given by $\int_{t_1}^{t_2} d_c(t) - a_c(t) dt + \sum_{i \in T_{\text{dis}} \cap [t_1, t_2]} m_i$. The first term converges to zero when t_1 tends towards t_2 , regardless of whether the endpoints t_1 and t_2 are included or not. For the second term however, the distinction matters as there might be a point mass at t_2 . If t_2 is excluded, the function $t_1 \mapsto \sum_{i \in T_{\text{dis}} \cap [t_1, t_2]} |m_i|$ is weakly decreasing and remains strictly positive for all $t_1 \in (-\infty, t_2)$. Hence, it converges towards its infimum as $t_1 \rightarrow t_2^-$. This infimum is equal to zero as all point masses occurring at times strictly smaller than t_2 are excluded progressively as $t_1 \rightarrow t_2^-$. As $|\sum_{i \in T_{\text{dis}} \cap [t_1, t_2]} m_i| \leq \sum_{i \in T_{\text{dis}} \cap [t_1, t_2]} |m_i|$, $|\sum_{i \in T_{\text{dis}} \cap [t_1, t_2]} m_i|$ also converges towards zero as $t_1 \rightarrow t_2^-$. If we define a sequence of times $(t_i)_{i \in \mathbb{N}}$ that converges from below towards t_2 but remains strictly smaller, $(n(t_i))_{i \in \mathbb{N}}$ is a Cauchy sequence in $\mathbb{R}$, hence it converges. The reasoning is the same to show that $\lim_{u \rightarrow t_2^+} n(u)$ exists. $\square$

Proposition 1 (Continuity). *Consider a system governed by a speed-MFD and a population satisfying Assumption 1. If a realization of the morning commute is an equilibrium, then for this realization the accumulation is a continuous function of time.*

⁴ A discrete version of this result would be along the lines of “if users all have different trip lengths and desired arrival times, then no user departs or leaves exactly at the same time as another one” but it would be more cumbersome to prove, if it can be proven at all.

Proof. In order to prove the continuity of the accumulation function, we assume the existence of a discontinuity at some time $t_{\text{dis}} \in T_{\text{dis}}$ and show that it implies that some user is not at equilibrium. We first assume that the discontinuity is an accumulation decrease, which implies the existence of a mass arrival. Since T^* is continuously distributed, only a set of measure 0 can have $T^* = t_{\text{dis}}$ (we will say such users are “on time” by analogy with the $\alpha - \beta - \gamma$ preferences, even though this does not necessarily have the same meaning with other types of schedule penalty functions). Thus, most of the users arriving at t_{dis} are not on time. Using Lemma 1, we can select a user from the mass that is not on time such that there was no discontinuity at the time of her departure t_d . Let t^* and θ denote the values of T^* and Θ of this specific user.

Lemma 2 guarantees that left and right-hand limits of accumulation exist everywhere. Consequently, the utility function admits left and right derivatives (with respect to the arrival time) everywhere. A local necessary condition for the selected user to be at equilibrium is that $\dot{U}_-(t_a) \geq 0 \geq \dot{U}_+(t_a)$ where $\dot{U}_-$ and $\dot{U}_+$ denote the left and right hand derivatives. Using Eq. (6) and Eq. (5), this is equivalent to

$$\frac{v(t_{\text{dis}}^-)}{v(t_d)} h(t_d, \theta) - w(t_{\text{dis}}, \theta, t^*) \geq 0 \geq \frac{v(t_{\text{dis}}^+)}{v(t_d)} h(t_d, \theta) - w(t_{\text{dis}}, \theta, t^*), \quad (9)$$

where $v(t_{\text{dis}}^-) = \lim_{u \rightarrow t^-} v(u)$ and $v(t_{\text{dis}}^+) = \lim_{u \rightarrow t^+} v(u)$ denote the left-handed and right-handed limits of the speed v at the discontinuity. Since $h(t_d, \theta) > 0$, Eq. (9) is equivalent to $v(t_{\text{dis}}^-) \geq v(t_{\text{dis}}^+)$, which is inconsistent with a brutal decrease of accumulation. Thus, a discontinuous decrease of accumulation cannot occur at equilibrium under the assumptions made.

The case of a discontinuous accumulation increase is treated in a similar manner by selecting a user that departs at the discontinuity. Among the users departing with this mass, only a zero-measure set can be on-time (because we assumed that both desired arrival time and trip length conditioned on desired arrival time are continuously distributed and only one trip length corresponds to a pair of departure and arrival time) and only a zero-measure set can arrive at another discontinuity (Lemma 1). Thus, we can select a user who departs at t_{dis} , is not on time and arrives at a time t_a with no discontinuity. The local equilibrium condition (9) becomes

$$\frac{v(t_a)}{v(t_{\text{dis}}^-)} h(t_{\text{dis}}, \theta) - w(t_a, \theta, t^*) \geq 0 \geq \frac{v(t_a)}{v(t_{\text{dis}}^+)} h(t_{\text{dis}}, \theta) - w(t_a, \theta, t^*),$$

i.e. $v(t_{\text{dis}}^-) \leq v(t_{\text{dis}}^+)$, which is inconsistent with a brutal increase of accumulation. $\square$

In the case where only the trip length is distributed and its distribution does not have any point mass, then one can show with a similar proof that the accumulation $n(t)$ is continuous everywhere except at the common desired arrival time. Note also that while Proposition 1 precludes the existence of an equilibrium consisting only of departure and arrival masses, departure and arrival masses might still coexist if they occur simultaneously and exactly cancel out.

4. Properties of the equilibrium with $\alpha - \beta - \gamma$ preferences

Depending on the congestion mechanism and on the scheduling preferences used, the first and second-order equilibrium conditions may lead to different sorting properties. For instance, Fosgerau (2015) showed for users with homogenous exponential-type scheduling preferences and a speed-MFD that, provided some endogenous assumptions on the equilibrium, these conditions translate into a Last-In-First-Out pattern where all trips overlap and shorter trips are “included” within longer trips, i.e. start later and finish earlier. We demonstrate in this section that with $\alpha - \beta - \gamma$ preferences and a speed-MFD, these conditions constrain the evolution of speed (see Proposition 2) in such a way that, provided the accumulation over time follows a unimodal curve, a partial FIFO sorting emerges (see Proposition 3). If parameters of the schedule penalty functions also vary between individuals, then they might also act locally as a sorting criterion, just as trip length (see Proposition 4). An important consequence of the use of $\alpha - \beta - \gamma$ preferences is that sorting applies even among users having different desired arrival times.

4.1. First and second order equilibrium conditions with $\alpha - \beta - \gamma$ preferences

Assume some user (t^*, l) with $\alpha - \beta - \gamma$ preferences arrives at equilibrium at time t_a . If her schedule penalty function SD is locally continuously differentiable (i.e. $t_a \neq t^*$), the first order equilibrium condition ($\frac{\partial U}{\partial t_a}(t_a, t^*, l) = 0$) translates

into

$$\frac{\partial \tau}{\partial t_a}(t_a, l) = -\frac{1}{\alpha} \frac{\partial \text{SD}}{\partial t_a}(t_a, t^*). \quad (10)$$

Similarly, if $\tau(t_a, l)$ and $\text{SD}(t_a, t^*)$ are twice differentiable, the second order equilibrium condition becomes

$$\frac{\partial^2 \tau}{\partial t_a^2}(t_a, l) \geq -\frac{1}{\alpha} \frac{\partial^2 \text{SD}}{\partial t_a^2}(t_a, t^*). \quad (11)$$

With $\alpha - \beta - \gamma$ preferences, these conditions can be stated as follows:

Proposition 2. Consider a system governed by a speed-MFD and a population of users satisfying Assumption 1. For a user with $\alpha - \beta - \gamma$ preferences, trip length l and desired arrival time t^* , the equilibrium departure and arrival times t_d and t_a satisfy:

- if $t_a < t^*$, then $v(t_a) = \frac{\alpha - \beta}{\alpha} v(t_d)$;
- if $t_a > t^*$, then $v(t_a) = \frac{\alpha + \gamma}{\alpha} v(t_d)$;
- if $t_a = t^*$, then $\frac{\alpha - \beta}{\alpha} v(t_d) \leq v(t_a) \leq \frac{\alpha + \gamma}{\alpha} v(t_d)$.

In addition, if $t_a \neq t^*$ and $v(t)$ is differentiable at t_a and t_d :

$$\frac{\dot{v}(t_d)}{v(t_d)^2} \geq \frac{\dot{v}(t_a)}{v(t_a)^2}. \quad (12)$$

Proof. For early and late users, the proportionality result between the speeds at departure and at arrival is obtained by combining the first order equilibrium condition (Eq. (10)) with Eqs. (8b) and (5). Similarly, Eq. 12 is obtained from the second order condition (Eq. (11)) by noticing that $\frac{\partial^2 \text{SD}}{\partial t_a^2}(t_a, t^*) = 0$ for all early and late users, by differentiating Eq. (5) and reinserting it into Eq. (11). Note that if $v(t)$ is differentiable at t_a and t_d , then the second order differentiability of the travel time at (t_a, l) is a direct consequence of its definition in Eq. (4). Regarding on-time users, although their utility function is not differentiable at their arrival time, it admits left and right derivatives. At equilibrium $\frac{\partial_- u}{\partial t_a}(t_a, t^*, l) \geq 0$ and $\frac{\partial_+ u}{\partial t_a}(t_a, t^*, l) \leq 0$, i.e.

$$\begin{cases} \frac{\partial \tau}{\partial t_a}(t_a, l) \leq -\frac{1}{\alpha} \frac{\partial_- \text{SD}}{\partial t_a}(t_a, t^*), \\ \frac{\partial \tau}{\partial t_a}(t_a, l) \geq -\frac{1}{\alpha} \frac{\partial_+ \text{SD}}{\partial t_a}(t_a, t^*). \end{cases} \quad (13)$$

Combine these equations with Eqs. (8b) and (5) to obtain the result for on-time users. □

Let us provide further intuition for the proposition above. The necessary conditions for some user to be at equilibrium stated in Proposition 2 already impose severe constraints on the shape of the equilibrium. Indeed, if we define the function $Q_l(t) = \frac{v(t)}{v(t_d(t, l))}$, the set of candidate arrival times at which a user may be at equilibrium is $Q_l^{-1}\left(\left\{\frac{\alpha - \beta}{\alpha}\right\}\right) \cup Q_l^{-1}\left(\left\{\frac{\alpha + \gamma}{\alpha}\right\}\right) \cup \{t^*\}$, where $Q_l^{-1}(I)$ denotes the (possibly empty) preimage of $I \subset \mathbb{R}$. Thus, provided desired arrival times are sufficiently similar, the exact value of t^* does not influence the equilibrium arrival times of early and late users, and users sort themselves based on their trip length and $\alpha - \beta - \gamma$ coefficients.

This suggests that the $\alpha - \beta - \gamma$ coefficients and trip length play a similar role and essentially sort users. The desired arrival time however seems to influence the equilibrium arrival time in a very distinct manner. Given some $\alpha - \beta - \gamma$ coefficients and trip length, the function that associates to a desired arrival time t^* the corresponding equilibrium arrival time would be the identity function for some range of times such that arriving on time is optimal, and would be piece-wise constant for the complement of this range. In the remainder of this section, we will focus on the effect of the $\alpha - \beta - \gamma$ coefficients and trip length. The reason for this choice is that the effect of t^* is either trivial (when users arrive exactly on time), or requires comparing the costs of candidate equilibrium arrival times, which means that the variations of accumulation must be known over long time intervals. Thus, understanding the effect of the $\alpha - \beta - \gamma$ coefficients and trip length is a preliminary step to understand the effect of t^* .

Note that this is very similar to the situation with a bottleneck model. When desired arrival time plays little role, the dynamics can be solved easily. For instance, with a S-shape cumulative distribution of desired arrival times and homogeneous $\alpha - \beta - \gamma$ coefficients, the dynamics only depend on two particular points of the distribution of t^* (Smith, 1984; Daganzo, 1985). Yet, when heterogeneity in desired arrival time is combined with heterogeneity in $\alpha - \beta - \gamma$ coefficients or when the desired arrival time does not have a S-shape cumulative distribution, numerical methods are necessary.

4.2. From a demand peak to a congestion peak

Various types of assumptions may be used to limit the influence of heterogeneity in desired arrival times. The S-shape assumption of Smith (1984) and Daganzo (1985) is an excellent example, being a reasonable idealization and still allowing for strong results in the case of a constant capacity bottleneck. So far, we have been unable to adapt this result to dynamics imposed by a speed-MFD, although early simulation results suggest a similar law might exist. Yet, as with a constant capacity bottleneck, our analytical investigations have shown that the exact distribution of desired arrival times is of little significance. What primarily matters is whether the distribution leads to a unimodal evolution of accumulation or to several peaks.

To obtain the sorting results of Section 4.3, we chose to consider the existence of a single peak as granted. More specifically, we made the following endogenous assumption.

Assumption 2 (Single peak). *The distribution of users' characteristics is such that at equilibrium: (i) all commuters travel within an interval $[t_1, t_2]$ and (ii) accumulation exhibits a single peak, i.e. there exists $t_p \in (t_1, t_2)$, such that for all $t \in [t_1, t_p]$, accumulation is strictly increasing and for all $t \in [t_p, t_2]$, accumulation is strictly decreasing.*

This assumption has the great advantage that it can be tested empirically. In fact, several empirical observations suggest that unimodal evolutions of accumulation are quite common (Parthasarathi et al., 2011; Yildirimoglu et al., 2015).⁵ It also greatly simplifies the analysis, as evidenced by the following corollary of Proposition 2.

Corollary 1. *Consider a system governed by a speed-MFD and assume that users' characteristics are such that the accumulation $n(t)$ is continuous and exhibits a single peak, reached at time t_p , as specified in the Assumptions 1 and 2. Consider a group of users with $\alpha - \beta - \gamma$ preferences. At equilibrium, early users depart and arrive before t_p while late users depart and arrive after t_p .*

Alternatively, it is shown in Appendix that sorting results similar to those presented in Section 4.3 can be obtained with only *exogenous* assumptions on the distribution of user characteristics (with, in particular, the assumption that desired arrival times are restricted to a compact subset of the time space). It requires however some additional precautions and, in the end, the sorting results apply only out of the range where desired arrival times are distributed. In that sense, it is similar to the numerous works assuming homogenous desired arrival times for Vickrey's bottleneck model, but with a more tedious derivation due to the state-dependent speed. We did not include this in the body of the paper as we favor realism, simplicity and far-reaching results over strict exogeneity of assumptions.

4.3. Sorting properties

This section includes two main sorting results. The first one (Proposition 3) builds on the Propositions 1, 2 and on the single peak assumption (Assumption 2) to provide a FIFO sorting result based on trip length that is valid for any pair of users having similar $\alpha - \beta - \gamma$ preferences and arriving both early (or both late). The second result (Proposition 4) is an extension of the first providing some complementary insight about the sorting phenomena emerging among users having *different* scheduling preferences, at the cost of several additional endogenous assumptions.

Proposition 3. *Consider a system governed by a speed-MFD and a population of users satisfying Assumptions 1 and 2. Consider two users having $\alpha - \beta - \gamma$ preferences and arriving early (resp. late) at equilibrium at the times $t_{a,1}$*

⁵ Smaller peaks with faster variations (e.g. on a time scale of 15 min) are often observed on top of the main peak but they might be considered as less predictable and therefore less important in the choice of departure time.

and $t_{a,2}$. Denote $(l_1, t_1^*, \alpha_1, \beta_1, \gamma_1)$ and $(l_2, t_2^*, \alpha_2, \beta_2, \gamma_2)$ their respective trip lengths, desired arrival times, α , β and γ coefficients. Assume without loss of generality that for instance $t_1^* \geq t_2^*$. If $\frac{\beta_1}{\alpha_1} = \frac{\beta_2}{\alpha_2}$ (resp. $\frac{\gamma_1}{\alpha_1} = \frac{\gamma_2}{\alpha_2}$) and $t_{a,1} \leq t_2^*$ (resp. $t_{a,2} \geq t_1^*$), then the user that arrives the earliest (resp. latest) has the longest trip length.

Proof. In the case where both users are early (resp. late), assuming $t_{a,1} \leq t_2^*$ (resp. $t_{a,2} \geq t_1^*$) is equivalent to assuming that both users would remain early (resp. late) after switching their arrival times.

Independently of the order of desired arrival times, let us use the index $i \in \{1, 2\}$ for the user that has the shortest trip length and j for the other. In addition to the notations previously defined, let us also define $t_{d,i}$ and $t_{d,j}$ the departure times of users i and j and $t'_{d,i}$, a departure time such that user i arrives exactly at $t'_{d,i} = t_{a,j}$. Since the original situation was an equilibrium, user i cannot change her decision unilaterally and reduce her cost. Therefore, the cost C'_i of departing at $t'_{d,i}$ must be at least as great as the cost C_i of departing at $t_{d,i}$: $C'_i \geq C_i$.

Let us compare the cost for user j of arriving at $t_{a,i}$ (denoted C'_j) and the cost of arriving at $t_{a,j}$ (denoted C_j). In the case both users are early:

$$\begin{aligned} C'_j &= \alpha_j \tau(t_{a,i}, l_j) + \beta_j (t_j^* - t_{a,i}), \\ &= \alpha_j (\tau(t_{a,i}, l_i) + \tau(t_{d,i}, l_j - l_i)) + \beta_j (t_i^* - t_{a,i}) + \beta_j (t_j^* - t_i^*), \\ &= \frac{\alpha_j}{\alpha_i} C_i + \alpha_j \tau(t_{d,i}, l_j - l_i) + \beta_j (t_j^* - t_i^*). \end{aligned}$$

Similarly, $C_j = \frac{\alpha_j}{\alpha_i} C'_i + \alpha_j \tau(t'_{d,i}, l_j - l_i) + \beta_j (t_j^* - t_i^*)$.

Equilibrium requires that $C'_j \geq C_j$, i.e. $\alpha_j (\tau(t_{d,i}, l_j - l_i) - \tau(t'_{d,i}, l_j - l_i)) \geq \frac{\alpha_j}{\alpha_i} (C'_i - C_i)$, and that $C'_i \geq C_i$. Hence,

$$\tau(t_{d,i}, l_j - l_i) \geq \tau(t'_{d,i}, l_j - l_i). \quad (14)$$

By applying Corollary 1, both users departed and arrived during the onset of congestion, so speed strictly decreases for the entire duration considered (Assumption 2). Thus, Eq. (14) imposes that $t_{d,i} \geq t'_{d,i}$, and therefore $t_{a,i} > t_{a,j}$.

If both users are late at equilibrium, two cases should be distinguished, depending on whether users i and j can exchange their arrival times and still depart both during the offset of congestion. If they cannot (user j would have to depart during the onset), then $t_{a,i} < t_{a,j}$. If they can, then β_i and β_j should simply be replaced by $-\gamma_i$ and $-\gamma_j$ and the equations above remain valid, leading to Eq. (14). Since the period considered is after the peak, speed strictly increases during the entire period so Eq. (14) imposes that $t_{d,i} < t'_{d,i}$ (i.e. $t_{a,i} < t_{a,j}$). $\square$

The ordering of arrivals demonstrated trivially implies the ordering of departures for early users (if $l_i < l_j$ and $t_{a,i} > t_{a,j}$, then $t_{d,i} > t_{d,j}$). For late arrivals, this is not true in general but it is true with the assumptions made as (i) speed is monotonously increasing during the offset of congestion and (ii) Proposition 2 imposes that if speed is greater at $t_{a,j}$ than at $t_{a,i}$, it is also greater at $t_{d,j}$ than at $t_{d,i}$. This leads to the following corollary.

Corollary 2. Consider a system governed by a speed-MFD and a population of users satisfying Assumptions 1 and 2. At equilibrium, departures and arrivals among a family of early (resp. late) users having the same ratio β/α (resp. γ/α) follow a First-In-First-Out order.

This is in sharp contrast with the LIFO pattern found by Fosgerau (2015). The difference stems from the types of scheduling preferences considered. In Fosgerau (2015), the marginal utilities at home and at work decrease and increase exponentially with time, such that users just try to be at the good place at the good time, i.e. travel around the time when both utility curves intersect. Note that this is valid independently of the level of congestion, so the system would unavoidably reach gridlock if there is a number of users having identical scheduling preferences that is greater than the jam accumulation. Here, the marginal utilities of time are respectively constant and piece-wise constant so earliness and lateness is more acceptable. Since travel time is relatively more important for users with long trip lengths, they naturally avoid the peak.

Proposition 3 is however limited to early users having identical ratios $\frac{\beta}{\alpha}$ or late users having identical ratios $\frac{\gamma}{\alpha}$. Since the coefficients α , β and γ are actually continuously distributed in the real world, the scope of this proposition might appear limited. However, one can actually extend Proposition 3 to cover such cases by applying the implicit function theorem (see Proposition 4).

For any early or late user, let r denote the user's active ratio of scheduling preferences, i.e. $r = \frac{\beta}{\alpha}$ for early users and $r = \frac{\gamma}{\alpha}$ for late users. Let also σ denote the sign of $\frac{\partial \text{SD}}{\partial t_a}(t_a, t^*)$ (+1 for late users and -1 for early users).

Proposition 4. Consider a user with trip length $\tilde{l}$, desired arrival time $\tilde{t}^*$ and $\alpha - \beta - \gamma$ preferences that arrives early or late at equilibrium, at time $\tilde{t}_a$. Denote $\tilde{r}$ her active ratio of scheduling preferences.

If (i) speed is continuously differentiable at $t_d(\tilde{t}_a, \tilde{l})$ and $\tilde{t}_a$, (ii) $\tilde{t}_a$ is the only arrival time that minimizes the user's cost (i.e. the global optimum is unique) and (iii) $\frac{\partial^2 \tau}{\partial t_a^2}(\tilde{t}_a, \tilde{l}) > 0$; then there exists an open set U around $(\tilde{l}, \tilde{t}^*, \tilde{r})$ such that there exists a unique function $t_a(l, t^*, r)$ on U such that a user with trip length l , desired arrival time t^* and active ratio of scheduling preferences r in U would be at equilibrium by arriving at $t_a(l, t^*, r)$.

The function $t_a(l, t^*, r)$ is continuously differentiable on U and its derivatives are given by:

$$\frac{\partial t_a}{\partial l}(l, t^*, r) = \frac{v(t_a)\dot{v}(t_d)}{v(t_a)^2\dot{v}(t_d) - v(t_d)^2\dot{v}(t_a)}; \frac{\partial t_a}{\partial r}(l, t^*, r) = -\sigma \frac{v(t_d)^3}{v(t_a)^2\dot{v}(t_d) - v(t_d)^2\dot{v}(t_a)}; \frac{\partial t_a}{\partial t^*}(l, t^*, r) = 0; \quad (15)$$

where t_d and t_a are functions of l , t^* and r .

Proof. Since the user considered does not arrive on time at equilibrium and speed is continuously differentiable at $t_d(\tilde{t}_a, \tilde{l})$ and $\tilde{t}_a$, the first order equilibrium condition is $\frac{\partial \tau}{\partial t_a}(t_a, l) + \sigma r = 0$. Using Eq. (5), this is equivalent to $g(t_a, t^*, l, r) = 0$, where $g(t_a, t^*, l, r) = \frac{v(t_d(t_a, l)) - v(t_a)}{v(t_d(t_a, l))} + \sigma r$.

Since v is continuously differentiable at $t_d(\tilde{t}_a, \tilde{l})$ and $\tilde{t}_a$, $t_d(t_a, l) = t_a - \tau(t_a, l)$ is also continuously differentiable (see Eq. (5)). Thus g is continuously differentiable. In addition, $\frac{\partial g}{\partial t_a}(t_a, l) = \frac{\partial^2 \tau}{\partial t_a^2}(t_a, l) > 0$. Hence, we can apply the implicit function theorem to obtain the existence of a unique function $t_a(l, t^*, r)$ on a open neighborhood around $(\tilde{l}, \tilde{t}^*, \tilde{r})$ such that a user characterized by (l, t^*, r) would be at equilibrium by arriving at $t_a(l, t^*, r)$. Furthermore, $\frac{\partial g}{\partial l} = \frac{v(t_a)\dot{v}(t_d)\frac{\partial t_d}{\partial l}}{v(t_d)^2} = -\frac{v(t_a)\dot{v}(t_d)}{v(t_d)^3}$; $\frac{\partial g}{\partial t_a} = \frac{v(t_a)^2\dot{v}(t_d) - v(t_d)^2\dot{v}(t_a)}{v(t_d)^3}$; $\frac{\partial g}{\partial t^*} = 0$; and $\frac{\partial g}{\partial r} = \sigma$. The implicit function theorem then provides the expressions of the partial derivatives of $t_a(l, t^*, r)$ given in the proposition. $\square$

The expressions of the three partial derivatives in Eq. (15) provide three local sorting results. If there is a single peak, $\dot{v}(t_d)$ is negative (resp. positive) for early (late) users. In addition, the condition (iii) is equivalent to $v(t_a)^2\dot{v}(t_d) - v(t_d)^2\dot{v}(t_a) > 0$. Hence the equilibrium arrival time locally decreases (increases) with the trip length, in agreement with Proposition 3. Similarly, $\frac{\partial t_a}{\partial r}(l, t^*, r) > 0$ so the equilibrium arrival time increases (resp. decreases) with the ratio $\frac{\beta}{\alpha}$ (resp. $\frac{\gamma}{\alpha}$), i.e. when schedule penalties are relatively more important. Finally, the fact that $\frac{\partial t_a}{\partial t^*}(l, t^*, r) = 0$ confirms that among early users, the desired arrival time plays no sorting role. Note that Newell (1987) obtained the same last two sorting properties (without using the gradient of t_a) for Vickrey's bottleneck. In fact, these qualitative properties result directly from the first order equilibrium condition, and only the quantitative properties of the gradient depend on the congestion mechanism.

Finally, let us discuss the assumptions used in Proposition 4. The differentiability assumption in Condition (i) is stronger than the continuity result shown in Proposition 1. Yet, we believe that in the same way we proved the continuity of the accumulation requiring only that the distributions of some characteristics of the population do not include any point-mass, the differentiability of the accumulation could be obtained by requiring only the continuity of the probability density functions of the trip length, of the desired arrival time and of the schedule penalty parameters. Although we do not prove this result, the intuition behind it is the following. Continuously distributed desired arrival times ensure the continuity of on-time departure and arrival rates. Continuously distributed trip lengths ensure the continuity of early and late departure and arrival rates. Then, continuously distributed schedule penalty parameters ensure that the transition between e.g. early and on-time departure does not occur simultaneously for many types of users. Condition (ii) simply discards some borderline cases and Condition (iii) is the second order sufficient optimality condition. Note that assuming Condition (i) and that the user is early or late at equilibrium already ensures that $\frac{\partial^2 \tau}{\partial t_a^2}(\tilde{t}_a, \tilde{l}) \geq 0$. However, a strict inequality is necessary to ensure that the conditions are also satisfied in an open neighborhood around the user considered.

5. Simulation-based analysis

Simulation-based and analytical approaches are complementary. While analytical considerations provide useful understanding about the key variables involved in congestion mechanisms, simulation can test more complicated configurations and provide clues as to whether systems with higher degrees of freedom and uncertainties can be described with little error by elegant analytical models. Indeed, most of the results proven in the previous sections require some assumptions that are simplifications of the real world (e.g. strict $\alpha-\beta-\gamma$ preferences or any homogeneity assumption). Simulation allows for the relaxation of these assumptions. Furthermore, the stability and the “power of attraction” of equilibria is just as important as their exact characterization. In fact, only a stable and attractive equilibrium is likely to be similar to the ever-changing real world traffic conditions. While these characteristics are usually extremely difficult to identify analytically, simulation inherently addresses these issues as it mimics the evolution of demand from day to day.

5.1. Description of the simulation

The simulation-based method proposed mimics the day-to-day adaptation of real drivers to changing conditions in order to identify a potential equilibrium. The basic idea consists in iteratively 1) updating the departure time decisions and 2) running an event-based and agent-based simulation based on Eq. (2) to determine the evolution of congestion over time and the cost of each user. In order to provide some stability, the commonly used (Peeta and Mahmassani, 1995; de Palma and Marchal, 2002) method of successive averages (MSA) is applied on the departure time decisions. At each iteration, the departure time decisions of a fraction of the population (e.g. 5% or $\frac{100}{n}\%$, where n is the iteration number) were updated based on the results of the previous simulation. Alternatively, MSA could have been applied on the travel times by updating the decisions of all the agents but using average travel times instead of those of the last iteration. The algorithm used is described in more details in Lamotte and Geroliminis (2016).

A quadratic speed-MFD such that $V(n) = v_f \left(1 - \frac{n}{n_{\text{jam}}}\right)^2$ if $n \in [0, n_{\text{jam}}]$ and $V(n) = 0$ otherwise is used for all numerical applications (v_f denotes the free-flow speed). Note that this MFD has a critical accumulation $n_{\text{cr}} = n_{\text{jam}}/3$, that $V(n_{\text{cr}}) = 4/9 v_f$ and that the maximum production is $(4/27)n_{\text{jam}}v_f = (4/9)n_{\text{cr}}v_f$. To decrease the degree of freedom, the free-flow speed was assumed to be 1 and the trip length to be uniformly distributed between 0 and 3 in all numerical applications.

Ten types of $\alpha - \beta - \gamma$ preferences were created with $\alpha = 1$, $\beta = 0.4 + \frac{0.2k}{9}$, and $\gamma = 1.5 + \frac{k}{9}$, where $k \in \{0, 1, \dots, 9\}$ indicates the family. For each family, 400 agents were created with desired arrival time $(i/400 - 1/2)h$, where $i \in \{1, 2, \dots, 400\}$ and h represents the length of the interval with desired arrival times. Hence, the total population is represented by 4000 agents. Various demands are simulated by varying the mass of each agent (or equivalently, the value of the jam accumulation) but keeping the number of agents (i.e. the resolution of the simulation) constant. A value of $h = 5$ was used for most results except when the impact of h is explicitly studied. For each scenario, 2000 iterations were carried out to approximate the equilibrium.⁶

5.2. Validation of the theoretical findings and relaxation of some assumptions

The analytical approach presented in the previous sections relied on a number of assumptions to characterize an hypothetical user equilibrium. This section provides some insight on the validity and scope of these results using the agent-based simulation described in Section 5.1.

Figures 1a and b are the main tools for this investigation, providing respectively aggregated and disaggregated description of the dynamics. Fig. 1a represents the natural logarithm of the speed versus $x = f(t)$. Recall that $f(t)$ represents the distance traveled at time t by a virtual user that enters the network with the first user and never exits (see Section 2.1). Since both f and the natural logarithm are bijections, this graph is similar to the evolution of speed over time, but in a different coordinate space. One can readily observe that this is a single peak case. Fig. 1b represents

⁶ Fewer iterations would have often been sufficient but cases with high congestion levels or little heterogeneity typically require a large number of iterations to converge.

Fig. 1: (a) Evolution of the natural logarithm of speed ($\ln(V)$) as a function of x , the distance traveled by a virtual user. (b) Representation of all users in the space (x_d, x_a) , illustrating the value of x at the arrival versus the value of x at the departure. All users had $\alpha - \beta - \gamma$ preferences as described in Section 5.1, the demand N/n_{cr} was set to 3.

each agent by a single point with coordinates the value of $x = f(t)$ at the agent's departure (x_d) and arrival (x_a). Note that since $x_a - x_d$ represents the trip length, all the points are above the line $x_a = x_d$.

Let us define $x_p = f(t_p)$ the value of x that minimizes $\ln(V(x))$. In other words, speed reaches its minimum at t_p . First, note that all the colored points corresponding to early users in Fig. 1b are located on the left and below the point (x_p, x_p) . This is actually the case for all early users (which form lines parallel to those colored), so all early users depart and arrive during the onset of congestion. Similarly, all colored points corresponding to late users are on the right and above the point (x_p, x_p) , i.e. late users depart and arrive during the offset of congestion, as per Corollary 1.

The sorting properties can also be observed in Fig. 1b. For each user, the trip length may be read as $x_a - x_d$, i.e. the vertical distance between the point representing the user and the line $x_a = x_d$. Fig. 1b shows clearly that within the sets of early users having identical coefficients $\alpha - \beta - \gamma$, those with long trips are on the left and below those with shorter trips, meaning that they depart and arrive earlier, as per Proposition 3. Similarly, late users with the same coefficient γ sort such that those with short trips depart and arrive earlier than those with longer trips. It is also clear that early users with $\beta = 0.49$ depart and arrive later than those having approximately the same trip length but with $\beta = 0.4$. In between, the points are located along three lines, corresponding respectively to the families $\beta = 0.4 + \frac{0.2k}{9}$ with $k = 1, 2$ and 3. The same applies to late users, as per Proposition 4.

Finally, the joint analysis of Figs. 1a and b illustrates the first and second order equilibrium conditions studied in Section 4.1. Indeed, in the coordinate system $(x, \ln(V))$, the first order condition (10) simply becomes $\ln(V(x_a)) = \ln(V(x_d)) - \ln(\frac{\alpha}{\alpha-\beta})$ for early users and $\ln(V(x_a)) = \ln(V(x_d)) + \ln(\frac{\alpha+\gamma}{\alpha})$ for late users. These conditions are illustrated for two early users and one late user by the dashed lines of different colors joining the Figs. 1a and b. For any user (x_d^0, x_a^0) materialized in Fig. 1b, $\ln(V(x_d^0))$ can be read at the intersection of the vertical line $x_d = x_d^0$ with the curve $\ln(V)$ (in Fig. 1a). Similarly, $\ln(V(x_a^0))$ may be read by starting from the point materializing the user, drawing an horizontal line $x_a = x_a^0$ until the line $x_a = x_d$ in Fig. 1b and then a vertical line $x_d = x_a^0$ until the curve $\ln(V(x))$ in Fig. 1a. One can then check that $\ln(V(x_a^0))$ is at the same vertical level as the curve $\ln(V(x)) - \ln(\frac{1}{1-\beta})$ at x_d^0 for early users, or at the same level as $\ln(V(x)) + \ln(1 + \gamma)$ for late users. Besides, the second order condition (11) for early or late users becomes $\frac{d}{dx} \ln(V(x)) \Big|_{x=x_d} \geq \frac{d}{dx} \ln(V(x)) \Big|_{x=x_a}$. One can easily verify that the slope of $\ln(V(x))$ systematically decreases between the departure and the arrival of early and late users.

Fig. 2: (a) Cumulative departure and arrival curves with strict $\alpha - \beta - \gamma$ preferences and the smooth approximation $h(t) = \alpha$ and $w(t) = \frac{2+\gamma-\beta}{2} + \arctan(4(t-t^*)) \frac{\gamma+\beta}{\pi}$. (b) and (c): Contour plots representing the amount of lateness for all 400 users having $\beta = 0.4$ and $\gamma = 1.5$ with strict $\alpha - \beta - \gamma$ scheduling preferences (b) and their smooth approximation (c). The results were obtained with a demand $N/n_{cr} = 3$.

Last, we investigate whether the phenomena obtained for the $\alpha - \beta - \gamma$ can be expected to hold with more realistic scheduling preferences that do not exhibit a discontinuity in t^* . The simulation above was replicated with a smooth approximation of the $\alpha - \beta - \gamma$ preferences, modeled by $h(t) = \alpha$ and $w(t) = \frac{2+\gamma-\beta}{2} + \arctan(4(t-t^*)) \frac{\gamma+\beta}{\pi}$. The resulting cumulative departure and arrival curves are displayed in Fig. 2a, together with those obtained with strict $\alpha - \beta - \gamma$ preferences. While departures and arrivals occur almost exactly over the same periods, the evolution of the cumulative arrival curve is smoother with the smooth preferences than with the $\alpha - \beta - \gamma$ preferences. The peak is also less pronounced and the maximum accumulation (i.e. the vertical distance between the two curves) smaller.

Despite these minor differences in the dynamics, the analysis of the individual decisions suggests the existence of similar patterns. Fig. 2b and c illustrate the amount of lateness (a negative amount corresponds to earliness) of all the 400 users characterized by $\beta = 0.4$ and $\gamma = 1.5$ with both types of preferences. With strict $\alpha - \beta - \gamma$ preferences, only users with long trips choose not to arrive on time. They reschedule earlier or later depending on their desired arrival time t^* . With smooth preferences, the same observation is valid for users with long trips. In this case however, short trips are also rescheduled in such a way that arrival times remain close to desired arrival times. This is caused by the slow variations in utility around t^* : with smooth scheduling preferences, there exists a range of arrival times which are almost equivalently desirable. Thus, in both cases, users with short trips are approximately on time, while those with long trips are early or late, depending on their desired arrival time. Overall, $\alpha - \beta - \gamma$ preferences produce results that are remarkably similar to those associated with a smooth approximation, thus justifying the use of the $\alpha - \beta - \gamma$ preferences for analytical derivations.

5.3. Impact of demand on the user equilibrium

Demand severity is described throughout this paper in terms of total demand (N/n_{cr}) and duration over which desired arrival times are distributed (h). Fig. 3a and 3b illustrate the time series of accumulation and speed for different values of N/n_{cr} . As expected, the peak duration and the maximum congestion level both increase with the total demand (keeping h constant). Note that for the four scenarios considered, the maximum congestion level is reached for $t = -\frac{h}{2}$, i.e. at the earliest desired arrival time. With Vickrey's bottleneck model and homogeneous $\alpha - \beta - \gamma$ preferences, the maximum would be reached later, for $t = \frac{\gamma-\beta}{\beta+\gamma} \frac{h}{2}$. Intuitively, as users have to travel some distance in the congested area, they are necessarily in the network before their desired arrival time, so the accumulation peak occurs earlier than with a constant capacity bottleneck.

As shown by Fig. 3c, the detailed dynamic model used allows for a very large share of users arriving during the interval $[-h/2, h/2]$, even though large accumulations may be observed earlier. To explain this phenomenon, we analyze the trajectory of the outflow in the accumulation-outflow space (see Fig. 3f). The common steady-state (and static average trip length L) approximation assumes that this trajectory lays along a well-defined curve $O(n) = \mathcal{P}(n)/L$ (represented by a black dashed line in Fig. 3f). Here, the observed outflow is very far from this idealized model and during the interval $[-h/2, h/2]$, it is about twice as large as the outflow-MFD capacity. Two reasons can explain

Fig. 3: User equilibrium comparison for different demands of the (a) accumulation over time ; (b) speed over time ; (c) cumulative departure and arrival curves ; (d) decomposition of the average congestion cost per user into its components associated to delays, earliness and lateness. The sub-figures (e) and (f) represent the harmonic average of trip length of all users in the network as a function of time (smoothed with a moving average) and the observed outflow (also smoothed with a moving average) in the accumulation-outflow space, for the scenario $N/n_{cr} = 3.6$. The observed outflow is compared in (f) with $P(t)/L$ (where $L=3/2$) and with $P(t)/L(t)$, where $L(t)$ is the harmonic average trip length represented in (e). Desired arrival times are distributed on an interval of length $h = 5$.

this discrepancy. First, the average trip length of users in the network is time-dependent, as shown in Fig. 3e.⁷ By replacing the constant trip length of the outflow model by the time-dependent average trip length $L(t)$, an anti-clockwise hysteresis loop is obtained, represented by the red curve in Fig. 3f.⁸ Yet, this hysteresis loop remains smaller than the one observed. The rest of the discrepancy might be explained by the complex endogenous delay in Eq. (3). Indeed, users that are about to finish their trip are under-represented in the network during the onset of congestion but over-represented during the offset. Thus, even if all the generated trips had the same length, an hysteresis loop would still be obtained. Note that the hysteresis phenomenon obtained here is extremely severe but that it would be reduced with slower time variations.

This high outflow during the most desired period allows for very little schedule penalties overall, as shown in Fig. 3d. In contrast, schedule penalties account for half the congestion cost with a bottleneck of constant capacity and homogeneous users. Note also that the individual congestion cost increases approximately linearly with the demand, at least for the range $N/n_{cr} \in [1.8, 3.6]$. This is similar to Vickrey's bottleneck problem (Arnott et al., 1990) and to the results of Arnott (2013) but in sharp contrast with those of Fosgerau (2015), in which a LIFO sorting means that a population larger than the jam accumulation of users having homogeneous scheduling preferences cannot be served at a finite cost. This result is related to the sorting pattern and is expected to hold whenever the system converges towards an equilibrium.

⁷ The harmonic average considered in Fig. 3e would be equal to the arithmetic average trip length of the trip-generating process if the system was in a steady state.

⁸ The noise in the upper part of the loop is caused by users having trip lengths almost equal to 0 and traveling during the period $[-h/2, h/2]$.

Fig. 4: (a) Comparison for $h = 2, 3, 4$ and 5 of the average individual congestion cost as a function of the demand N/n_{cr} ; (b) time series of accumulation for a demand of $N/n_{cr} = 1.8$ and $N/n_{cr} = 3$; (c) proportions of early and late users as functions of the demand N/n_{cr} for desired arrival times distributed on intervals of length $h = 2$ and $h = 5$.

5.4. Impact of staggered work hours

Staggered work hours may be used instead or in addition to congestion pricing to flatten the peak and reduce schedule penalties (Henderson, 1981). With homogeneous $\alpha - \beta - \gamma$ preferences and a S-shape distribution of desired arrival times, Vickrey's bottleneck predicts a schedule delay reduction but no impact on the dynamics as long as all desired arrival times remain within the congested period (Vickrey, 1969).

With a speed-MFD however, Fig. 4 shows that staggered work hours do impact the dynamics, in a way that depends on the total demand N/n_{cr} . Fig. 4a compares the average congestion cost for different demands and ranges h of desired arrival times. Fig. 4b and Fig. 4c provide similar comparisons for the time series of accumulation (only for $N/n_{cr} = 1.8$ and 3) and the proportions of early and late users (only for $h = 2$ and 5). $h = 2$ may correspond to an initial state while $h = 5$ may correspond to a state after staggering work hours. Fig. 4b suggests that the impact of h on the dynamics of accumulation is larger when the demand is lower, i.e. when the range of desired arrival times is of the same order of magnitude as the peak duration. When the peak period is significantly longer than the range of desired arrival times, staggering work hours only slightly reduces the maximum congestion level. Yet, it still allows for significant cost reductions, as evidenced by Fig. 4a. These cost reductions are explained by the reduced congestion level but also by a large reduction of the proportions of early and late users (see Fig. 4c). Note that we only focused here on the impact of staggering work hours on the congestion cost. In reality, staggering work hours also entails other costs. For instance, for many types of professions, the productivity of a worker at a given time depends on whether others also work at that time (Henderson, 1981).

6. Discussion

This paper has investigated the equilibrium properties and appropriate congestion pricing strategies for the morning commute problem with a regional bottleneck modeled by a speed-MFD. Our investigations have shed light on the fundamental role of trip length heterogeneity. We first showed that with the fluid approximation, a continuously distributed trip length ensures that the main variables of the morning commute (accumulation, speed) are continuous functions of time for a large class of scheduling preferences. This essential property served then as the basis for several analytical derivations. We showed in particular that if the users' characteristics are such that the morning commute consists of a single peak, a FIFO sorting pattern naturally emerges within families of early and late users having identical $\alpha - \beta - \gamma$ preferences and heterogeneous trip lengths, or vice versa. Simulations suggest that these sorting properties influence the global characteristics of the equilibrium and produce patterns that are also observed with similar but smooth utility functions.

Yet, many questions remain open. The comparison of the present work with Fosgerau (2015) underlines the importance of empirical measurements. Different estimations of scheduling preferences reaching different conclusions have highlighted the complexity of an apparently simple choice. While this would not fully characterize the scheduling

preferences, the analysis of a large number of observations of real world trips may answer the question of the prevailing pattern (FIFO or LIFO). On a more theoretical note, an important issue is the determination of necessary and sufficient conditions for the emergence of a single peak. Early results towards this direction are detailed in Appendix and deserve to be extended. This might also be completed by empirical works investigating whether commuters schedule their trip mostly based on a smoothed profile of the dynamics (that would exhibit a single peak in most cases) or whether they also account for small oscillations.

The consequences of rapidly varying demand on congestion pricing should also be investigated. Different tolls have been proposed that maintain the accumulation below its critical value, either at a constant (Daganzo and Lehe, 2015) or at a time-varying level (Arnott, 2013). These tolls were shown to be optimal under some simplifying assumptions and are expected to provide significant benefits as long as demand does not vary too rapidly. Under more rapid variations however, such pricing strategies might prove inefficient and a complete change of paradigm may be needed. Consider for instance a system governed by a triangular production-MFD, such that the production-maximizing accumulation is N . Assume that there are exactly N long-distance travelers that travel (at free-flow speed) from 7 am to 8 am. Now assume that there is an additional short-distance traveler, who would like to arrive at 7:30 am. If the short trip length is short enough, it is socially optimal that all users arrive on-time, even if it creates transient hypercongestion.⁹ Note that the fact that demand results from scheduling preferences is of little importance. Indeed, even if we considered departures times as given, it would not be optimal to prevent the $N + 1^{th}$ user from entering the network until 8 am. Such a measure would impose 30 min delay to this user while her impact on the others can be arbitrarily small (with arbitrarily small trip lengths). Thus, when the dynamics are governed by a speed-MFD, the total distance traveled in the next time unit by all vehicles is not the only performance criterion: its distribution among users matters as well. In some situations, finishing as many trips as possible in the near future might be a more appropriate myopic objective than keeping the system in a state that maximizes the number of vehicle kilometers traveled (i.e. the production). This is also supported by recent control applications in networks governed by multi-region MFD dynamics (Ramezani et al., 2015). Estimation of the trip completion rate is difficult with traditional measurement methods (e.g. loop detectors) but will become more practical with the increasing use of trajectory data.

7. Acknowledgements

We would like to thank Richard Arnott for the very fruitful discussion. The authors were financially supported by the ERC Starting Grant METAFERW: Modeling and controlling traffic congestion and propagation in large-scale urban multimodal networks.

Appendix A. An approach of sorting relying only on exogenous assumptions

This Appendix details an alternative approach to sorting that only relies on exogenous assumptions. The main assumption is that desired arrival times are distributed in a compact subset of the time axis. Then, we show that accumulation cannot be decreasing before this time interval and cannot be increasing after it (Proposition 6). We refer to this result as the ‘single predominant peak’. Accumulation can be constant during some time period only if nobody departs or arrives during that time period (Lemma 3). Altogether, these results form a basis to derive a sorting result that is valid among users traveling exclusively out of the interval where desired arrival times are distributed (Proposition 7).

⁹ To prove this, first note that the situation proposed is a local optimum. Indeed, small changes of departure times only add schedule penalties and do not reduce travel time. Second, adding one on-time user with a very short trip length is equivalent to keeping the number of users constant but continuously switching the trip length of the $N + 1^{th}$ user from zero to some non-zero value. By continuity of the social cost function with respect to the $N + 1$ pairs of arrival times and trip length (gridlock is not possible here), as there is only one global optimum when the trip length of the $N + 1^{th}$ user is zero, the global optimum changes continuously in a small neighborhood around this situation. Thus, if the short trip is short enough, the situation proposed is the global optimum.

A.1. Preliminary results

While the concept is quite simple, the proofs are made more tedious by the utilization of a continuum of users. A drawback of this approach is that even with the full knowledge of the cumulative numbers of departures and arrivals $D(t)$ and $A(t)$, the departure and arrival *rate* functions are only determined up to a set of measure zero, so that one cannot be sure that there are actually users departing or arriving at any specific time. Fortunately, Proposition 5 allows us to circumvent this issue. Before stating it, we need to introduce some notations and a definition.

Let $\mathcal{Z} = \{(\alpha, \beta, \gamma, l, t^*) \in (\mathbb{R}^{++})^4 \times \mathbb{R}, \beta < \alpha\}$ denote the set of admissible characteristics for users with $\alpha - \beta - \gamma$ preferences. The condition $\beta < \alpha$ ensures that users arriving early do not remain in their vehicle until their desired arrival time t^* . Let also $C(z, t_a)$ denote the function from $\mathcal{Z} \times \mathbb{R}$ to $\mathbb{R}$ that maps a user with characteristics z and arrival time t_a to her experienced cost. Note that since the travel time function $\tau(t_a, l)$ is continuous, the cost function $C(z, t_a)$ is continuous as well.

Definition 1. A function $h : \mathbb{R} \rightarrow \mathbb{R}$ is said to be locally constant at $\tilde{t} \in \mathbb{R}$ if there exists $\epsilon > 0$, such that $h(t)$ is constant on $[\tilde{t} - \epsilon, \tilde{t} + \epsilon]$.

Proposition 5. Consider a system governed by a speed-MFD and a population of users whose characteristics are continuously distributed on a compact subspace Z of $\mathcal{Z}$. Consider an equilibrium realization of the morning commute. If the cumulative number of departures (resp. arrivals) is not locally constant at some $t \in \mathbb{R}$, then t is an equilibrium departure (resp. arrival) time for some user with characteristics $z \in Z$.

Proof. If $D(t)$ (resp. $A(t)$) is not locally constant, one can find for all $n \in \mathbb{N}$ a user that arrives (at equilibrium) in the interval $t_n \in [t - 1/n, t + 1/n]$. Denote z_n and t_n her characteristics and arrival time. Since all terms of $(z_n)_{n \in \mathbb{N}}$ are in the compact space Z , the Bolzano-Weierstrass theorem ensures that there exists a subsequence of $(z_n)_{n \in \mathbb{N}}$ that converges. Let z denote the limit of this subsequence. Because Z is closed, z belongs to Z . The continuity of the function $C(z, t)$ ensures that t is an equilibrium departure (resp. arrival) time for a user characterized by z . $\square$

Note that there might not be any user with characteristics z in the population but it does not matter. Note also that the statement “ D is not locally constant at t ” is equivalent to “there exists $m \in (0, N]$, such that $t = \inf\{u \in \mathbb{R}, D(u) \geq m\}$ ”. The time t is often abusively referred to as the departure time of the m^{th} user to depart, although as we have discussed, there might not be any user in the population who departs at t .

A.2. The single predominant peak result

We are now ready to define our assumptions and the single predominant peak result.

Assumption 3 (Distributed $\alpha - \beta - \gamma$ preferences). 1. All users have $\alpha - \beta - \gamma$ preferences, with coefficients belonging to a compact subspace Z of $\mathcal{Z}$. In particular, t^* belongs to an interval $[\underline{t}^*, \bar{t}^*]$.
 2. Desired arrival times (t^*) are continuously distributed.
 3. $\frac{\beta}{\alpha}$ and $\frac{\gamma}{\alpha}$ are bounded away from 0 for all users.
 4. Trip length and trip length conditioned on the desired arrival time are continuously distributed variables.

Note that Assumption 3 implies Assumption 1. The main additional requirements of Assumption 3 are that all users have $\alpha - \beta - \gamma$ preferences and that the users’ characteristics are distributed on a compact space. Requirement 3 is purely technical and quite reasonable.

Proposition 6 (Single predominant peak). Consider a system governed by a speed-MFD and a population satisfying Assumption 3. At equilibrium, accumulation is weakly increasing for all $t < \underline{t}^*$ and weakly decreasing for all $t > \bar{t}^*$.

Proof of Proposition 6. Requirement 3 of Assumption 3 allows us to define $\epsilon > 0$ such that $\frac{\alpha}{\alpha - \beta} > 1 + \epsilon$ and $\frac{\gamma + \alpha}{\alpha} > 1 + \epsilon$ for all users. Proposition 2 implies that between the departure and the arrival of an early (resp. late) user, speed must be divided (resp. multiplied) by at least $1 + \epsilon$. Then, the proof proceeds by contradiction: we demonstrate that an accumulation decrease before $\underline{t}^*$ or an increase after $\bar{t}^*$ implies the existence of a sequence of users whose speed at departure or arrival tends to infinity, which is not possible given the speed-MFD considered.

We detail the proof that $n(t)$ is weakly increasing for $t < \bar{t}^*$ and then explain how it can be adapted to show that $n(t)$ is weakly decreasing for $t > \bar{t}^*$. Suppose there exist $t_1 < t_2 < \bar{t}^*$ such that $n(t_1) > n(t_2)$ and let $u \in (v(t_1), v(t_2)) \cap (v(t_2)/(1 + \epsilon), v(t_2))$. Since Assumption 3 implies Assumption 1, Proposition 1 guarantees the continuity of accumulation and speed. The intermediate value theorem ensures the existence of a time $t'_1 \in [t_1, t_2]$ such that $v(t'_1) = u$. Define then $A = \{t \in [t'_1, t_2], n(t) = n(t'_1)\}$. A is clearly bounded, not empty and since $n(t)$ is continuous, A is closed and admits a largest element, that we denote t'_1 . Similarly, let t'_2 be the smallest element of $\{t \in [t'_1, t_2], n(t) = n(t_2)\}$. The interval $[t'_1, t'_2]$ constructed is such that for all t inside it, $n(t)$ belongs to $[n(t'_2), n(t'_1)]$.

The definition of t'_1 also implies that for all $\xi > 0$, there exists $t \in [t'_1, t'_1 + \xi]$ such that $n(t) < n(t'_1)$, hence $A(t) > A(t'_1)$, i.e. A is not locally constant at t'_1 . By applying Proposition 5, there exists an element $z = (\alpha, \beta, \gamma, l, t^*) \in \mathcal{Z}$ that would be at equilibrium by arriving at t'_1 . Hence:

$$\begin{aligned} C(z, t'_1) \leq C(z, t'_2) &\Leftrightarrow \alpha \tau(t'_1, l) + \beta(t^* - t'_1) \leq \alpha \tau(t'_2, l) + \beta(t^* - t'_2) \\ &\Leftrightarrow \alpha(\tau(t'_2, l) - \tau(t'_1, l)) \geq \beta(t'_2 - t'_1) \\ &\Leftrightarrow \alpha \int_{t'_1}^{t'_2} \frac{\partial \tau}{\partial u}(u, l) du \geq \beta(t'_2 - t'_1) \\ &\Leftrightarrow \alpha \int_{t'_1}^{t'_2} 1 - \frac{v(u)}{v(t_d(u, l))} du \geq \beta(t'_2 - t'_1) \quad (\text{using Eq. (5)}) \\ &\Leftrightarrow \int_{t'_1}^{t'_2} f(u) du \geq 0, \end{aligned} \tag{A.1}$$

where $f(t) = 1 - \frac{\beta}{\alpha} - \frac{v(t)}{v(t_d(t, l))}$.

Proposition 2 implies that $v(t'_1)/v(t_d(t'_1, l)) = 1 - \beta/\alpha$, i.e. $f(t'_1) = 0$. Since $v(u) \geq v(t'_1)$ for all $u \in [t'_1, t'_2]$ and $v(u) > v(t'_1)$ in a neighborhood around t'_2 , then $v(t)$ also has to be strictly greater than $v(t_d(t'_1, l))$ for some $t \in [t_d(t'_1, l), t_d(t'_2, l)]$ for Eq. (A.1) to be satisfied. We ignore a priori whether t'_1 is greater or smaller than $t_d(t'_2, l)$ but we do know that for all $t \in [t'_1, t'_2]$, $v(t) \leq v(t'_2) < (1 + \epsilon)v(t'_1) < v(t_d(t'_1, l))$. Hence, there exists a time $s_2 \in (t_d(t'_1, l), t'_1)$ such that $n(s_2) < n(t_d(t'_1, l))$.

To conclude, one could repeat the same process iteratively with the pair $(s_1 = t_d(t'_1, l), s_2)$ replacing the pair (t_1, t_2) . Since $v(t_d(t'_1, l)) > (1 + \epsilon)v(t'_1) > v(t_1)$, the speed of the left-hand term of the pair tends to infinity. This is not possible with the speed-MFD considered and we can reject the initial assumption.

The reasoning to prove that accumulation cannot be increasing after $\bar{t}^*$ is very similar. Consider an hypothetical pair $\bar{t}^* < t_1 < t_2$, such that $n(t_1) < n(t_2)$. By following the same process as above, one can find an interval $[t'_1, t'_2]$ such that for all t inside it, $n(t)$ belongs to $[n(t'_1), n(t'_2)]$, $n(t_1) < n(t'_1)$, $v(t'_2) > v(t'_1)/(1 + \epsilon)$ and $D(t)$ is not locally constant at t'_2 . We then need to define the function $t_a(t, l)$ which associates to a departure time t and a trip length l the corresponding arrival time. Clearly, $t_a(t_d(t, l), l) = t$ holds for all $t \in \mathbb{R}$. Then, $C(z, t_a(t'_1, l)) \geq C(z, t_a(t'_2, l))$ is equivalent to $\int_{t_a(t'_1, l)}^{t_a(t'_2, l)} f(u) du \leq 0$, where $f(t) = 1 + \frac{\gamma}{\alpha} - \frac{v(u)}{v(t_a(u, l))}$. Again $f(t_a(t'_2, l)) = 0$ and since $v(t_d(u, l)) \geq v(t'_2)$ for all $u \in [t_a(t'_1, l), t_a(t'_2, l)]$ and $v(t_d(u, l)) > v(t'_2)$ in a neighborhood around $t_a(t'_1, l)$, then $v(t)$ also has to be strictly greater than $v(t_a(t'_2, l))$ for some $t \in [t_a(t'_1, l), t_a(t'_2, l)]$. We ignore a priori whether t'_2 is greater or smaller than $t_a(t'_1, l)$ but we do know that for all $t \in [t'_1, t'_2]$, $v(t) \leq v(t'_2) < (1 + \epsilon)v(t'_1) < v(t_a(t'_2, l))$. Hence, there exists a time $s_1 \in (t'_2, t_a(t'_2, l))$ such that $n(s_1) < n(t_a(t'_2, l))$. Note finally that $v(t_a(t'_2, l)) > (1 + \epsilon)v(t'_2) > v(t_2)$. The same process can be repeated iteratively with the pair $(s_1, s_2 = t_a(t'_2, l))$ replacing (t_1, t_2) and the speed of the right-hand term tends to infinity. $\square$

Proposition 6 states essentially that out of the range where desired arrival times are distributed, accumulation does not exhibit local extrema: it is weakly increasing before the lower bound of desired arrival times and weakly decreasing after the upper bound. Intuitively, if desired arrival time is close to being homogeneous, then accumulation would essentially consist of single peak. While this supports the idea that single congestion peaks should be quite common as well with MFD dynamics, this result is weaker than the result of Smith (1984) and Daganzo (1985) for the bottleneck with constant capacity.

A.3. Sorting result

Lemma 3. Consider a system governed by a speed-MFD and a population satisfying Assumption 3. Let $t \in \mathbb{R} \setminus [\underline{t}^*, \bar{t}^*]$.¹⁰ If accumulation is locally constant at time t at equilibrium, then the cumulative departure and cumulative arrival functions are also locally constant at t .

Proof of Lemma 3. This proof follows the same approach as the proof of Proposition 6: we show by contradiction that the existence of a time t that violates Lemma 3 implies the existence of a sequence of speeds that tends to infinity. We only prove the result for the early case as the late case is very similar.

Let us first define $\epsilon > 0$ such that $\frac{\alpha}{\alpha-\beta} > 1 + \epsilon$ and $\frac{\gamma+\alpha}{\alpha} > 1 + \epsilon$ for all users. Assume that there exists $\tilde{t} < \underline{t}^*$ and $\xi > 0$ such that accumulation is constant on $[\tilde{t} - \xi, \tilde{t} + \xi]$. Take ξ small enough to ensure that $\tilde{t} + \xi < \underline{t}^*$. Assume (by contradiction) that the cumulative departure function D or the cumulative arrival function A is not locally constant at $\tilde{t}$. Since $n(t) = D(t) - A(t)$ for all time t , this actually implies that both D and A are not locally constant at $\tilde{t}$.

For any interval I , let $D_I(t)$ denote the cumulative departure function of the users arriving during I . Let M denote the measure of users arriving during $[\tilde{t} - \xi/2, \tilde{t} + \xi/2]$. Choose any $z \in (0, M]$. For any small neighborhood around the time $\tilde{t}' = \min\{t \in \mathbb{R}, D_{[\tilde{t}-\xi/2, \tilde{t}+\xi/2]}(t) = z\}$, there is a positive measure of users that departed in this neighborhood and arrived in $[\tilde{t} - \xi/2, \tilde{t} + \xi/2]$.

We now show that accumulation is locally constant at $\tilde{t}'$. Let $\xi' = \xi v(\tilde{t}) / v_f$, where v_f is the free-flow speed. Let l and t_a denote the trip length and arrival time of some user departing in $[\tilde{t}' - \xi'/3, \tilde{t}' + \xi'/3]$ and arriving in $[\tilde{t} - \xi/2, \tilde{t} + \xi/2]$. On one hand, user equilibrium requires that for all $t \in [\tilde{t} - \xi, \tilde{t} + \xi]$, $C_l(t_a) \leq C_l(t)$. By applying the same reasoning as in the proof of Proposition 6, this is equivalent to $\int_{t_a}^t f(u) du \geq 0$, where the function $f(t)$ is defined as in Eq. (A.1). On the other hand, since $v(u)$ is constant between t_a and t and $v(t_d(u, l))$ is weakly decreasing (Proposition 6), $f(t)$ is also weakly decreasing between t_a and t . Since this is valid for all $t \in [\tilde{t} - \xi, \tilde{t} + \xi]$ and Proposition 2 requires that $f(t_a) = 0$, $f(t)$ is non-negative on $[\tilde{t} - \xi, t_a]$ and non-positive on $[t_a, \tilde{t} + \xi]$.

Together, these two results imply that $f(t) = 0$ for all $t \in [\tilde{t} - \xi, \tilde{t} + \xi]$, i.e. speed is constant on $[t_d(\tilde{t} - \xi, l), t_d(\tilde{t} + \xi, l)]$. By construction, $\tilde{t}'$ belongs to the interior of $[t_d(\tilde{t} - \xi, l), t_d(\tilde{t} + \xi, l)]$. Indeed, as $\int_{t_d(\tilde{t}-\xi, l)}^{\tilde{t}-\xi} v(u) du$ and $\int_{t_d(t_a, l)}^{t_a} v(u) du$ must both be equal to l , $\int_{t_d(\tilde{t}-\xi, l)}^{t_d(t_a, l)} v(u) du = \int_{\tilde{t}-\xi}^{t_a} v(u) du$. Since $\int_{t_d(\tilde{t}-\xi, l)}^{t_d(t_a, l)} v(u) du \leq (t_d(t_a, l) - t_d(\tilde{t} - \xi, l)) v_f$ and $\int_{\tilde{t}-\xi}^{t_a} v(u) du = (t_a - (\tilde{t} - \xi)) v(\tilde{t}) \geq \frac{\xi}{2} v(\tilde{t})$, we obtain $t_d(\tilde{t} - \xi, l) \leq t_d(t_a, l) - \frac{\xi v(\tilde{t})}{2v_f}$. As $t_d(t_a, l) \leq \tilde{t}' + \xi'/3$, we have shown that $t_d(\tilde{t} - \xi, l) \leq \tilde{t}' + \xi'/3 - \xi'/2$, i.e. $t_d(\tilde{t} - \xi, l) < \tilde{t}'$. Similarly, $t_d(\tilde{t} + \xi, l) \geq t_d(t_a, l) + \frac{\xi v(\tilde{t})}{2v_f} \geq \tilde{t}' - \xi'/3 + \xi'/2 > \tilde{t}'$. Hence accumulation is locally constant at $\tilde{t}'$.

To conclude, one could repeat the same process iteratively, going back in time from $\tilde{t}$ to $\tilde{t}'$, then from $\tilde{t}'$ to $\tilde{t}^{(2)}$, etc. Since there are early users that depart close to $\tilde{t}'$ and arrive close to $\tilde{t}$ and accumulation is locally constant in the neighborhoods of $\tilde{t}'$ and $\tilde{t}$, Proposition 2 implies that speed is multiplied by a factor of at least $(1 + \epsilon)$ at every step, and hence tends to infinity.

The proof for the case $\tilde{t} > \bar{t}^*$ is extremely similar, except that one should go forward in time, as in the proof of Proposition 6. $\square$

Proposition 7. Consider a system governed by a speed-MFD and a population of users satisfying Assumption 3. Consider two users with schedule preferences $(\alpha_1, \beta_1, \gamma_1)$ and $(\alpha_2, \beta_2, \gamma_2)$ arriving before $\underline{t}^*$ (resp. after $\bar{t}^*$) at equilibrium and departing at times where the cumulative departure function is not locally constant. If $\frac{\beta_1}{\alpha_1} = \frac{\beta_2}{\alpha_2}$ (resp. $\frac{\gamma_1}{\alpha_1} = \frac{\gamma_2}{\alpha_2}$), then the user that arrives the earliest (resp. latest) has the longest trip length.

The proof is very similar to the proof of Proposition 3 but makes use of the Proposition 6 and Lemma 3 to conclude from Eq. (14).

References

Arnott, R., 2013. A bathtub model of downtown traffic congestion. *Journal of Urban Economics* 76, 110–121.

¹⁰ The notation $A \setminus B$ denotes the complement of B with respect to A , i.e. the set of elements in A but not in B .

- Arnott, R., de Palma, A., Lindsey, R., 1990. Economics of a bottleneck. *Journal of urban Economics* 27, 111–130.
- Arnott, R.J., Kokoza, A., Naji, M., 2016. Equilibrium traffic dynamics in a bathtub model: A special case. *Economics of Transportation* 78, 38–52.
- Buisson, C., Ladier, C., 2009. Exploring the impact of homogeneity of traffic measurements on the existence of macroscopic fundamental diagrams. *Transportation Research Record: Journal of the Transportation Research Board*, 127–136.
- Daganzo, C.F., 1985. The uniqueness of a time-dependent equilibrium distribution of arrivals at a single bottleneck. *Transportation science* 19, 29–37.
- Daganzo, C.F., 2007. Urban gridlock: Macroscopic modeling and mitigation approaches. *Transportation Research Part B: Methodological* 41, 49–62.
- Daganzo, C.F., Lehe, L.J., 2015. Distance-dependent congestion pricing for downtown zones. *Transportation Research Part B: Methodological* 75, 89–99.
- Fosgerau, M., 2015. Congestion in the bathtub. *Economics of Transportation* 4, 241–255.
- Fosgerau, M., Small, K.A., 2013. Hypercongestion in downtown metropolis. *Journal of Urban Economics* 76, 122–134.
- Geroliminis, N., Daganzo, C.F., 2008. Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings. *Transportation Research Part B: Methodological* 42, 759–770.
- Geroliminis, N., Levinson, D., 2009. Cordon pricing consistent with the physics of overcrowding, in: Lam, W.H.K., Wong, S.C., Lo, H.K. (Eds.), *Transportation and Traffic Theory 2009: Golden Jubilee*. Springer US, pp. 219–240.
- Godfrey, J.W., 1969. The mechanism of a road network. *Traffic Engineering and Control* 11, 323–327.
- Henderson, J., 1981. The economics of staggered work hours. *Journal of Urban Economics* 9, 349–364.
- Hendrickson, C., Plank, E., 1984. The flexibility of departure times for work trips. *Transportation Research Part A: General* 18, 25–36.
- Herman, R., Prigogine, I., 1979. A two-fluid approach to town traffic. *Science* 204, 148–151.
- Hjorth, K., Börjesson, M., Engelson, L., Fosgerau, M., 2015. Estimating exponential scheduling preferences. *Transportation Research Part B: Methodological* 81, Part 1, 230–251.
- Ji, Y., Luo, J., Geroliminis, N., 2014. Empirical observations of congestion propagation and dynamic partitioning with probe data for large-scale systems. *Transportation Research Record: Journal of the Transportation Research Board* 2422, 1–11.
- Lamotte, R., Geroliminis, N., 2016. The morning commute in urban areas: Insights from theory and simulation, in: *Transportation Research Board 95th Annual Meeting*.
- Mahmassani, H., Williams, J.C., Herman, R., 1984. Investigation of network-level traffic flow relationships: some simulation results. *Transportation Research Record: Journal of the Transportation Research Board* 971, 121–130.
- Mahmassani, H.S., Saberi, M., Zockaie, A., 2013. Urban network gridlock: Theory, characteristics, and dynamics. *Transportation Research Part C: Emerging Technologies* 36, 480–497.
- Newell, C., 1982. Applications of queueing theory. volume 4.
- Newell, G.F., 1987. The morning commute for nonidentical travelers. *Transportation Science* 21, 74–88.
- de Palma, A., Marchal, F., 2002. Real cases applications of the fully dynamic metropolis tool-box: An advocacy for large-scale mesoscopic transportation systems. *Networks and Spatial Economics* 2, 347–369.
- Parthasarathi, P., Srivastava, A., Geroliminis, N., Levinson, D., 2011. The importance of being early. *Transportation* 38, 227–247.
- Peeta, S., Mahmassani, H.S., 1995. System optimal and user equilibrium time-dependent traffic assignment in congested networks. *Annals of Operations Research* 60, 81–113.
- Ramezani, M., Haddad, J., Geroliminis, N., 2015. Dynamics of heterogeneity in urban networks: aggregated traffic modeling and hierarchical control. *Transportation Research Part B: Methodological* 74, 1–19.
- Small, K.A., 1982. The scheduling of consumer activities: Work trips. *The American Economic Review* 72, 467–479.
- Small, K.A., Chu, X., 2003. Hypercongestion. *Journal of Transport Economics and Policy* 37, 319–352.
- Smith, M.J., 1984. The existence of a time-dependent equilibrium distribution of arrivals at a single bottleneck. *Transportation Science* 18, 385–394.
- Tseng, Y.Y., Verhoef, E.T., 2008. Value of time by time of day: A stated-preference study. *Transportation Research Part B: Methodological* 42, 607 – 618.
- Vickrey, W.S., 1969. Congestion Theory and Transport Investment. *American Economic Review* 59, 251–60.
- Yildirimoglu, M., Geroliminis, N., 2014. Approximating dynamic equilibrium conditions with macroscopic fundamental diagrams. *Transportation Research Part B: Methodological* 70, 186–200.
- Yildirimoglu, M., Limniati, Y., Geroliminis, N., 2015. Investigating empirical implications of hysteresis in day-to-day travel time variability. *Transportation Research Part C: Emerging Technologies* 55, 340–350.