Leaving the tub
Bao, Yue; Verhoef, Erik T.; Koster, Paul

published in
Transportation Research Part E: Logistics and Transportation Review
2021

DOI (link to publisher)
10.1016/j.tre.2021.102389

document version
Publisher's PDF, also known as Version of record
document license
Article 25fa Dutch Copyright Act

Link to publication in VU Research Portal

citation for published version (APA)

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

E-mail address:
vuresearchportal.ub@vu.nl

Download date: 02. Nov. 2023
Leaving the tub: The nature and dynamics of hypercongestion in a bathtub model with a restricted downstream exit

Yue Bao\textsuperscript{a,b,∗}, Erik T. Verhoef\textsuperscript{b,c}, Paul Koster\textsuperscript{b,c}

\textsuperscript{a} School of Traffic and Transportation, Beijing Jiaotong University, Beijing, China
\textsuperscript{b} Department of Spatial Economics, VU University Amsterdam, Amsterdam, the Netherlands
\textsuperscript{c} Tinbergen Institute, Gustav Mahlerplein 117, 1082 MS Amsterdam, the Netherlands

ARTICLE INFO

Keywords:
Road traffic congestion
Flow congestion
Bathtub model
Hypercongestion

ABSTRACT

Hypercongestion is the situation where a certain traffic flow occurs at a combination of low speed and high density, while a more favorable combination of these could produce the same flow. The macroscopic fundamental diagram (MFD) can describe hypercongestion, but does not explicitly explain the dynamic process that leads to hypercongestion. Earlier studies of hypercongestion on single links have, however, confirmed that these processes are important to consider. The bathtub model is a model that can be used to investigate how hypercongestion can arise in urban areas, when drivers choose their departure times optimally. This paper investigates equilibrium outcomes and user costs under the realistic assumption that there is finite capacity to exit the bathtub, without which it would be hard to explain why hypercongestion would not dissolve through shockwaves originating from the bathtub exit. We find that when the exit capacity of the bathtub is lower than the attempted equilibrium exit flow from the bathtub, no additional inefficiencies arise due to hypercongestion in the bathtub. This is because the travel time losses incurred by travelers in the bathtub are exactly offset by the reductions in travel time losses in exit queues, and thus the capacity of the full system is not affected. In contrast, when the exit capacity is higher than the equilibrium exit flows from the bathtub in the central part of the peak period, hypercongestion in the bathtub produces the additional inefficiencies known from the conventional textbook description. Our results thus show that the mere observation of hypercongested speeds does not necessarily mean that there is an efficiency loss from capacity drop at the level of the full system.

1. Introduction

Traffic congestion usually arises because of a combination of limited road capacities and a strong concentration of travel demand in space and time of the day. This leads to travel time losses, scheduling costs, and may increase environmental externalities such as air pollution as well as traffic crashes. When travelers make a trip to an urban region, they may face two types of congestion: queueing behind clearly identifiable bottlenecks like bridges or tunnels, and dynamic flow congestion in urban areas. Bottleneck congestion was first studied from a dynamic equilibrium perspective by Vickrey (1969) and later elaborated upon in papers by Arnott et al. (1990, 1993a). In these models, dynamic equilibrium arises because the capacity of the bottleneck is limited and travelers make trade-offs

∗ Corresponding author.

E-mail addresses: y.bao@vu.nl (Y. Bao), e.t.verhoef@vu.nl (E.T. Verhoef), p.r.koster@vu.nl (P. Koster).

https://doi.org/10.1016/j.tre.2021.102389
Received 31 December 2020; Received in revised form 20 April 2021; Accepted 29 May 2021
Available online 12 June 2021
1366-5545/© 2021 Elsevier Ltd. All rights reserved.
between travel time losses and scheduling disutility from arriving earlier or later at their destination than preferred. In the basic bottleneck model, the only travel time loss, above the free flow travel time, is the waiting time at the queue before the bottleneck, where it is assumed that the capacity outflow of the bottleneck is constant. This feature of the bottleneck model leads to convenient closed-form results for the traffic equilibrium and the social optimum, which makes this stylized bottleneck model attractive for the economic analysis of dynamic traffic equilibrium (Arnott et al., 1993a).

However, also when equilibrium traffic flow is below the maximum flow capacity, additional time losses can already occur, resulting in equilibrium speeds that are below free-flow speed levels. In flow congestion models, the dynamic travel time function provides a structural relationship between capacity, the number of travelers or density at a certain time instant, and speed or travel time. Many of these functions predict delays when equilibrium traffic volumes are below capacity. For example, Henderson (1974) applied a formulation where the travel time for a single road is a smooth function of the flow at the road’s entrance when departing from home, whereas Chu (1995) used a specification where travel time is determined by the flow at the road’s exit when arriving at the destination. Mun (1994) and Mun (1999) divided a one-link road into two parts of endogenous lengths: a part without a queue, followed by a traffic jam part caused by a downstream bottleneck. Others have investigated theoretically the impacts of queue spillback and capacity drops at the bottleneck on the upstream flow (e.g., Daganzo, 1998; Nie and Zhang, 2008; Ma et al., 2017; Yuan et al., 2017; Baer et al., 2019).

More recently, there have been new developments in the modelling of dynamic congestion in cities (e.g., Geroliminis and Levinson, 2009; Arnott, 2013; Fosgerau and Small, 2013; Fosgerau, 2015; Daganzo and Lehe, 2015; Lamotte and Geroliminis, 2018), arising from empirical observations about congestion in urban areas. An important and fascinating empirical insight is the observation of a fairly stable relationship between average equilibrium speed and average equilibrium density at the level of an entire urban area (Geroliminis and Daganzo, 2008). This relationship is called the Macroscopic Fundamental Diagram (MFD), in order to distinguish it from the familiar link-based fundamental diagram of traffic congestion that depicts the inverted U-shape relation between density and flow. The MFD arises as the combined result from speed falling monotonously with density, and flow equaling the product of density and speed. A similar stable relationship between traffic speed, density and flow for an entire urban area also underlies models of “bathtub congestion” proposed in various recent transport economic models (Small and Chu, 2003; Arnott, 2013; Fosgerau, 2015; Li and Huang, 2019). The usual explanation for the parallel with a bathtub is that for an urban area, the process of drivers entering (leaving) the traffic network is like the inflow (outflow) of water into (out of) bathtub, where traffic density would correspond to the level of water in the bathtub (Fosgerau, 2015). In the bathtub model, instantaneous traffic conditions are assumed to be homogeneous over continuous space, making it possible to avoid the explicit modelling of route choices. Furthermore, this setup avoids complications arising from continuous-time-continuous-place congestion modelling, as in for example car-following models (Verhoef, 2001, 2003) and hydrodynamic models (Newell, 1988).

The MFD and the bathtub model describe the trip completion rate, or outflow from the network, in a way that it strongly depends on the average density, and thus the accumulations of vehicles, in the network. Given the spatial homogeneity of congestion and slowly-evolving traffic conditions, it is believed that there exists an approximately constant relation between the outflow from the network and the accumulations in the network (Lamotte and Geroliminis, 2018). Both the MFD and the bathtub model allow for spatially homogeneous hypercongestion, but neither explicitly describes the dynamic process leading up to hypercongestion. From one-directional models such as that in Verhoef (2003), we know that in the absence of a downstream capacity restriction, hypercongestion cannot build up on a single constant-capacity facility, and any initially assumed spatially homogeneous hypercongestion would dissolve from upstream moving shockwaves originating from the exit. For modelling purposes, it is therefore important to include downstream capacity restrictions, even though the earlier bathtub models are constructed in a way that traffic conditions are uniform over space and thus no downstream capacity limitation exists to sustain hypercongestion. That is, in these models, vehicles exiting are, as it were, “pulled back” by vehicles still present in the bathtub and cannot even have the slightest acceleration when no longer having a car in front, which would be enough to start the self-reinforcing dynamic process of dissolution of hypercongestion at the exit.

It is clear that one could readily specify a single speed-flow relation that empirically matches the backward-bending shape also observed in empirical MFD studies, and then use this relationship for policy analyses and welfare assessments. The main potential pitfall of using this approach is that analyses that ignore these exit capacity restrictions might very well produce strongly biased and misleading assessments of traffic efficiencies if in reality it is the existence of capacity restrictions in the exiting of vehicles that causes hypercongested conditions to be observed, and if the existence of such capacity restrictions has an impact on how equilibria would change under differential policies. The simple observation that this would already be true for the two-serial links model in Verhoef (2003), arguably the simplest possible urban network that one could seek to represent by an MFD or bathtub model with a downstream capacity restriction, leaves no reason to hope that the problem would disappear if the network represented by the bathtub or MFD becomes more complex. This motivates our current study, in which we analyze whether, and if so how, the existence of exit capacity restrictions would affect the behavior and insights from MFD and bathtub models in the assessment of traffic efficiencies, with a focus on hypercongestion.

To stay as close as possible to existing MFD and bathtub models we maintain the assumed spatial homogeneity of traffic conditions. Given that these models are in particular designed to model traffic in a downtown area, spatial homogeneity requires that destinations or exits be uniformly and continuously scattered over space. We thus study what would happen in these models if the exits, which we will conveniently refer to as parking facilities, have a finite capacity. It should be noted that for this interpretation of exits from the bathtub we only consider the limited entrance capacity of these parking areas, but do not involve the stock capacity inside the parking areas, if only to simplify the argument of the paper. For a more detailed analysis of drivers’ cruising for parking space with a MFD/bathtub model of travelers we refer to Geroliminis (2015) (without departure time choices) and Liu and Geroliminis (2016) (with...
departure time choices).

The conventional bathtub model incorporates drivers’ behavior by endogenizing commuters’ departure time choices (Arnott, 2013), but it is analytically intractable because network outflow at each point in time is determined by the prior states of traffic over the peak. Amirgholy and Gao (2017) further studied the equilibrium solution with different forms of exogenous departure rate function in MFD by extending Vickrey’s bottleneck theory to the macro-level. In this paper, we aim to investigate how drivers’ dynamic behavior, traffic conditions in the network, and the outflow from the network interact with each other when there exist downstream capacity restrictions. We are in particular concerned with how these restrictions would affect equilibrium travel time and scheduling costs due to hypercongestion, that arises because flow drops below its possible maximum resulting in capacity waste.

The remainder of the paper is organized as follows. The next section will explain the motivations of this paper and its contributions. Section 3 introduces the methodology used this paper. Section 4 extends the model to distinguish different types of hypercongestion. Section 5 presents the numerical analyses. The final section concludes the paper and discusses possible extensions for future research.

2. The bathtub model with restricted downstream capacity

In this paper, we consider commuting trips in which drivers need to drive in a certain area (the bathtub) to their work. To maintain the assumption of spatial homogeneity of traffic conditions in a bathtub, we assume that drivers’ destinations are uniformly distributed in the bathtub. Thus, exits from the bathtub are distributed uniformly over space, in such a fine resolution that they induce no spatial inhomogeneities in traffic conditions. The description most closely representing our model would be a spatial continuum of exits, which we will approximate by letting the number of exits, $E$, go to infinity. Also, to maintain spatial homogeneity, we assume that these exits all have same capacity, so that exit capacity per unit of space in the bathtub is also constant over space.

Denoting the aggregate outflow from the bathtub into all exit queues jointly as $\bar{A}(t)$, the arrival rate of new users at each bottleneck – or at the tail of its queue – is, under spatial homogeneity, $\bar{A}(t)/E$. Denoting the aggregate exit capacity from the bathtub as $\bar{c}_0$, the capacity per bottleneck will be $\bar{c}_0/E$. Now observe that for each of these bottlenecks, the queuing time at any moment is an integral of earlier queue growth rates, $(\bar{A}(t) - \bar{c}_0)/\bar{c}_0$, divided by the bottleneck capacity, $\bar{c}_0/E$, over time. A consequence is that we can multiply any single bottleneck’s queue growth rates and capacity by $E$ and find the same time pattern of queuing delays for that bottleneck. In other words, we can analytically treat the condition of exits as if they form a single bottleneck with capacity $c_0$, with an aggregate inflow of $A(t)$. That is what we will do in what follows.

Concerns over spatial inhomogeneity in traffic conditions, that a localized single exit point of outflow would undoubtedly induce, vanish when $E$ goes to infinity so that exits become continuous over space, while keeping their capacities (which then becomes a density of exit capacity) uniform over space. But we can maintain the relatively simple dynamics that a single bottleneck brings about. The assumed uniform distribution of exit capacities over the bathtub is of course a simplification, but it seems the only description that is consistent with the spatial homogeneity of traffic conditions that characterizes bathtub congestion models in the first place. In particular, spatially homogenous time-varying traffic conditions in the bathtub can only be supported as a dynamic equilibrium if also time-varying exit conditions are homogeneous over space.

In what follows we will thus treat the exit of the bathtub as a single bottleneck with an aggregate attempted time-varying inflow of $A(t)$ and a capacity of $c_0$, but do this keeping in mind that it produces queuing times that would apply for a continuum of spatially homogenous bottlenecks with the same aggregate inflow and capacity. We therefore formulate a mathematical model that considers a spatially homogenous bathtub area with a single bottleneck at its exit. Commuters depart from home, encounter traffic flow congestion in the city/bathtub area. Then they arrive at the exit and face a finite capacity, which results in the build-up of a vertical queue. We also note that the assumption of vertical queues provides a conservative approach in identifying the impact of a downstream capacity restriction on the emergency of hypercongestion: the queue does not itself reduce space available in the bathtub. After passing

For the exits, we thus consider parking facilities as an example. This seems to be the closest approximation to spatially homogenous exits that a model needs in order to remain consistent with the assumption of spatial homogeneity of traffic in the bathtub. But other realistic descriptions of exits from the area with uncongested traffic, like the bridges or tunnels and even traffic lights, are also feasible, as long as their number is high enough to make spatial homogeneity an acceptable approximation, and the desired exit time from the exits is the same for all exits. If bridges/tunnels/lights are very close to the destination, and if the number of bridges/tunnels is large enough that they can lead to a continuum of exits, we believe that our model might also provide a useful approximation of such a setting, and leave the possibilities of the inclusion of other bottlenecks as a direction of future study.

We will not explicitly model spillbacks of queues at the exit into the bathtub. The analytical reason is that doing so would mean that we can no longer maintain the assumption of spatially homogenous traffic conditions in the bathtub, which would greatly complicate the analysis. The economic justification for ignoring these complications is that the qualitative dynamic equilibrium features of the model would not be affected fundamentally. In particular, there would still be the two possible types of dynamic equilibria that we also find without explicit modelling of spillbacks. One in which the exit rate from the bathtub would remain equal to its exit capacity and where there is a perfect substitution between waiting time in the exit queue and travel time in the bathtub. For this equilibrium there is no outflow loss from hypercongestion in terms of reduced arrival rates at the destination. For the other equilibrium the flow in the bathtub is so low in the most congested period, that queuing at the exit disappears close to the most preferred arrival time, and hypercongested speeds in the bathtub do produce the additional outflow loss of arrival rates falling below exit capacity. The reason why the first type of equilibrium could also occur with spillbacks is that also then, the exit flow remains equal to exit capacity when there is a queue. The reason why the second type of equilibrium could also occur with spillbacks is that when we find that the exit queues have disappeared in the most central times of the peak, it no longer matters if queue spillbacks would have been modelled explicitly.
Drivers between the two bottlenecks, in which both bottlenecks again have constant capacities. It was found that a queue may not form at the upstream bottleneck in some cases. Li and Zhang (2015) investigated the tandem-bottleneck system in which the bottleneck capacity can be time-dependent, and showed that the resulting model can be an approximation of a kinematic wave model. Other research on multi-reservoir systems. In their study, an entry supply function and the exit departure rate were explicitly modelled, while drivers’ departure time choices were not considered. The effect of a fixed capacity at the exit of the bathtub was also studied in Ingole et al. (2020) in the context of perimeter gating. As discussed above, what we consider in this paper is that there exists a maximum outflow from the network, which is constant over space to maintain consistency with the spatial homogeneity assumption that is characteristic for the bathtub models. Unlike practice in the bottleneck model, we will not assume that the maximal outflow is achieved throughout the peak, especially not when congestion is very serious. Indeed, we will find equilibria where also outflow from the bathtub decreases with density and falls below the capacity as hypercongestion occurs, in the middle of the peak period. This will actually be shown in Section 4, and relates directly to one of our main contributions: the distinction between different types of hypercongestion.

The set-up of our model makes it closely related to earlier research on so-called tandem or serial bottlenecks, which deals with the interactions in service of an upstream and a downstream bottleneck on a highway. The simplest case would be to have two Vickrey bottlenecks, both with vertical queuing, in series; serving a single origin-destination pair. The implied network configuration would thus be the same as what we are considering: two serial congestible facilities serving the same origin-destination pair. The system would, however, be trivial to analyze, as its equilibrium and optimum are equivalent to that for a single bottleneck with the smallest of the two capacities. If the larger-capacity bottleneck would have the downstream position, it would never become active; neither in the untolled equilibrium nor in the optimum. If it would have the upstream position, it might become active in the untolled equilibrium if its capacity is below the equilibrium departure rate for early travelers. But then, for any arrival moment, the sum of queuing delays that a driver has experienced at the two bottlenecks would be equal to what otherwise – without the upstream bottleneck – would have been experienced at the single smaller-capacity bottleneck. As we will see, when we replace the upstream bottleneck with a bathtub in our model, the analysis is no longer trivial. The difference is that in our model, whether the downstream bottleneck is active or not, there always exists some flow congestion at the upstream facility, and our model can capture hypercongestion in the sense that equilibrium outflow of the downstream bottleneck may fall below its capacity. Still, our founding on the significance of expanding downstream capacity is consistent with earlier findings (Kim, 1999).

Some variants on this simplest tandem-bottleneck problem have been studied. Kuwahara (1990) and Kim (1999) analyzed the dynamic equilibrium queuing patterns for the two-tandem bottleneck network with two origins and one destination, one origin in between the two bottlenecks, in which both bottlenecks again have constant capacities. It was found that a queue may not form at the upstream bottleneck in some cases. Li and Zhang (2015) investigated the tandem-bottleneck system in which the bottleneck capacity can be time-dependent, and showed that the resulting model can be an approximation of a kinematic wave model. Other research on multiple bottlenecks elaborated on more complex network structures, like the Y-shape traffic network with two bottlenecks (Arnott et al. 1993b; Daniel et al. 2009), and the corridor problem with multiple bottlenecks (Shen and Zhang, 2009; Akamatsu et al., 2015). As far as we know, none of these studies deals with the two-serial-congestible facilities that admits hypercongestion.

An interesting and unexpected benefit of equilibria with constant outflow is that they avoid the analytical intractability issue of the standard bathtub model. As stated in Amirgholy and Gao (2017), the equilibrium in the bathtub model can be pinned down when drivers’ departure rates from home are known. Given the constant outflow constraint, an endogenous departure rate from home can be obtained. In fact, even when this constraint is not active, the equilibrium departure rate can be obtained, based on the features of MFD between outflow and flow or the relationship between trip length, speed and arrival time at exit of the bathtub, which in fact corresponds to the two approaches used in this paper to solve the model (see Section 3 below). Therefore, although being somehow unrealistic, an exogenous bottleneck at the exit on one hand helps to explicitly describe the dynamic process leading up to hypercongestion and thus sheds light on the build-up of hypercongestion reflected in the backward-bending segment of MFD, and on the other hand provides the methodology to solve the models in closed form.

Hypercongestion is verified to exist in our model, where the duration of hypercongestion naturally depends both on the properties of the network and the assumed demand parameters. However, we will be more elaborate than this as in the context of our model, two types of hypercongestion can be distinguished. The first is what we will call “bathtub-speed” hypercongestion, which arises when
traffic density in the bathtub exceeds the value consistent with the maximum possible flow in the bathtub (see Fig. 1, B1). The inflow into the queue then naturally falls below the maximum possible inflow, determined by the maximum flow that can occur in the bathtub. The second form will be referred to as “system” hypercongestion, and this arises when a sufficiently extended period of bathtub-speed hypercongestion makes queuing at the bottleneck disappear altogether, and subsequently causes outflow from the bathtub to fall below the exit capacity (see Fig. 1, A1) for a certain period where the bathtub is heavily congested. While bathtub-speed hypercongestion only involves zero-sum tradeoffs between travel time in the bathtub and queuing time at the exit, system hypercongestion results in a longer equilibrium peak period, and therefore corresponds to higher equilibrium travel costs.

Here we first illustrate why hypercongested speeds in equilibrium may not create the conventional efficiency loss associated with flow falling below capacity in a qualitative, intuitive sense. In short, the reason is that whenever the exit capacity is lower than the flow from the bathtub into the exit, the exit flows remain constant, simply because the exit has a fixed capacity and functions as a bottleneck. A higher speed in the network may then simply mean an earlier joining of the queue at exit, so that travel time gains in the bathtub-speed hypercongestion makes queuing at the bottleneck disappear altogether, and subsequently causes outflow from the network to drop below the exit capacity. When departure time choices of travelers are endogenous, the equilibrium duration of the peak period will depend on the downstream capacity in a way identical to the conventional bottleneck model when it operates at full capacity throughout the peak.

Only when the outflow from the network drops below the exit capacity, the equilibrium duration of the peak will be affected by the flow in the bathtub. With reference to the familiar flow-speed and flow-cost relations in Fig. 1: benefits can then be obtained by moving from the hypercongested situation (A2 in Fig. 1) to the normally congested range (A1 in Fig. 1). This implies that if the exit continues to operate at full capacity throughout the peak, it does not matter whether speed in the bathtub varies between $\bar{v}$ and $\bar{v}$ (with hypercongestion), or between $v'$ and $v$ (without hypercongestion). The duration of the peak period is then always equal to the ratio between demand and the downstream capacity: $N/c_{in}$, and travel delays in the bathtub are perfectly substituted by travel delays at the exit queue. Only when speed further increases from $\bar{v}$ (normal congestion) or further decreases from $\bar{v}$ (hypercongestion), exit capacity waste may occur if that deviation is sufficiently long-lived to make an early exit queue fully dissipate, and this would then result in a longer peak period. We will further investigate this intuition analytically in the following sections.

3. Modelling approach

In this section, we will introduce the model to be used in our paper. We first consider equilibria in which the exit is always fully utilized, from the first driver arriving at the bottleneck until the last driver arriving at the bottleneck. In this way, an endogenous departure rate from home can be easily obtained for the whole peak period. However, as mentioned above, this does not mean that in our model the constant outflow is necessarily always achieved. But it may, and the reason that the corresponding equilibria are considered here is that in this way, we can have a clear picture of the methodology used. We will discuss the more general case, where the arrival rate is not constant, in Section 4.

3.1. The bathtub model

The bathtub model assumes that the urban area is perfectly homogeneous, which results in an equilibrium in which traffic speeds, densities, and flows do not vary over space (Fosgerau, 2015). Flow congestion in the bathtub area follows a stable relationship between density and speed. For this paper, this relationship is assumed to be linear and given by (Greenshields et al., 1935):

$$v(t) = v_f(1 - \lambda k(t))$$

where $v_f$ is the free-flow speed and $\lambda = 1/k_v$ is defined as the reciprocal of the jam density, $k_v$, which is the value of traffic density that results in zero traffic speed. We can freely choose units of road capacity such that density, $k(t)$, can be expressed as the total number of cars in the bathtub at time $t$, and is equal to the difference between the cumulative departures from home, $D(t)$, and the cumulative arrivals at exit, $A(t)$:

$$k(t) = D(t) - A(t)$$

Denote $d(t) = D'(t)$ and $a(t) = A'(t)$ as the departure rate from home (inflow into the bathtub) and the arrival rate at exit (outflow from the bathtub into the exit queue), respectively. The evolution of density inside the bathtub is then given by the difference between the departure rate from home and the arrival rate at the exit:

$$\dot{k}(t) = d(t) - a(t)$$

Trip length in the bathtub is logically equal to the integral of speed over time, from the departure time $t$ until the corresponding arrival time at exit queue, $s(t)$:

$$L = \int_t^{s(t)} v(o) do$$

Then, according to the first-in-first-out principle, the cumulative departures from home and the cumulative arrivals at exit are related as:
\[ D(t) = A(s(t)) \]  

\[(5)\]

which means that the cumulative number of drivers having departed from home at time \( t \) is equal to the cumulative number of drivers having arrived at exit at time \( s(t) \).

Unfortunately, this bathtub model cannot be solved without making further assumptions. The difficulty is that, in order to obtain the equilibrium departure and arrival schedules, we need to know drivers’ arrival time at the destination, while this arrival time depends on the traffic speed and density at every instant in the bathtub before arriving at the destination. However, the instantaneous traffic speed and density are determined by the departure and arrival schedules. There are two approaches available in literature to make further progress.

First, Arnott (2013) adopted a proportional relationship between outflow from the bathtub and flow inside the bathtub by assuming that all drivers have the same probability of exiting the network at any time. The relationship is consistent with the “Network Exit Function (NEF)” in Gonzales and Daganzo (2012), which describes “the flow of vehicles exiting the network as a function of the total number of vehicles circulating in the network”. It is demonstrated by Geroliminis and Daganzo (2007) that this function can be estimated using the MFD of the network. This approach is also used by Amirgholy and Gao (2017) and Lamotte and Geroliminis (2018), and is referred as “accumulation-based” approach in Mariotte et al. (2017).

The second approach to obtain an equilibrium solution was introduced by Fosgerau (2015), and makes use of the relationship between trip length and speed in Eq. (4). Equilibrium was obtained based on the assumption of “regular sorting”, which means drivers with a longer trip length will always depart earlier and arrive later in equilibrium. By treating departure time and arrival time as functions of trip length, drivers’ travel time in equilibrium only depends on the speed-density relationship (1), and on the distribution of trip lengths. This approach is also termed as “trip-based” in literature (Mariotte et al., 2017).

We will adopt and compare both approaches in our paper. The first one, we will refer to as the CPI (constant proportion between flow in the bathtub and outflow from the bathtub) model in this paper, while the second method will be referred to as the FTL (fixed trip length) model as we assume drivers’ trip length are fixed in this case. The key difference of our model with earlier bathtub models in literature is that we consider a stylized bathtub area with an exit capacity restriction. As a result, an analytical solution for the trip-length model can be obtained, which makes it much easier to solve the model. For the CPI model, we do not need the “regular-sorting” assumption. In the following, we will first introduce how an endogenous departure rate is obtained, and then how the two approaches solve our models.

### 3.2. Bathtub model with a bottleneck at the exit

Given the constant outflow constraint, the queuing time for a traveler who departs from home at time \( t \) and arrives at exit at time \( s(t) \) is given by:

\[ T_q(s(t)) = \frac{A(s(t)) - c_a(s(t) - t_q)}{c_a} \]  

\[(6)\]

where \( c_a \) is the capacity of the exit and \( t_q \) is the arrival time at exit (and also at the destination as no queue exists) of the very first driver departed. For a driver departing at time \( t \), the arrival time at the destination is given by:

\[ \psi(t) = s(t) + T_q(s(t)) = \frac{A(s(t))}{c_a} + t_q = \frac{D(t)}{c_a} + t_q \]  

\[(7)\]

which satisfies:

\[ \psi'(t) = s'(t) + T_q'(s(t)) = \frac{d(t)}{c_a} \]  

\[(8)\]

Denote \( t' \) as drivers’ preferred arrival time at destination. Drivers’ user cost in monetary units is assumed to be given by the conventional expressions for “\( \alpha - \beta - \gamma \)” preferences:

\[ p(t) = \alpha(s(t) + T_q(s(t)) - t') + \max \{ \beta(t' - s(t) - T_q(s(t))) , \gamma(s(t) + T_q(s(t)) - t') \} \]  

\[(9)\]

where \( \alpha \) denotes drivers’ value of time, \( \beta \) and \( \gamma \) denote the unit schedule-early cost and the unit schedule-late cost, respectively. The first-order condition for drivers’ optimal departure time choices is given by:

\[ p'(t) = \begin{cases} 
(\alpha - \beta) \frac{d(t)}{c_a} - \alpha, t_e \leq t \leq \tilde{t} \\
(\alpha + \gamma) \frac{d(t)}{c_a} - \alpha, \tilde{t} \leq t \leq t_e 
\end{cases} = 0 \]  

\[(10)\]

where \( t_e \) and \( t_e \) denote the times of the first departure and the last departure from home, and \( \tilde{t} \) is the departure time from home of the drivers who arrive at destination at \( t' \). Solving Eq. (10) for the departure rate yields:
\[
d(t) = \begin{cases} 
\phi_1, t_s \leq t \leq \tilde{t} \\
\phi_2, \tilde{t} < t \leq t_c 
\end{cases}
\]
(11)

where \(\phi_1 = \frac{\alpha}{\gamma} c_A\) and \(\phi_2 = \frac{\alpha}{\gamma} c_B\), which means the equilibrium departure rate is the same as that in the conventional bottleneck model (Vickrey, 1969). The reason is as follows. Because the capacity of the exit is always fully utilized, and given the first-in-first-out principle that should hold in equilibrium, drivers must queue at the exit in the order of their departure time. Thus, their arrival time at the destination is determined by how many drivers have departed before them, in combination with the exit capacity and the arrival time of the first departure. It is therefore not determined by their travel time inside the bathtub, but instead by the sum of that travel time and the queuing time. Thus, we basically have a standard bottleneck model, where for an individual’s total travel delay cost, it is only the sum of bathtub travel time and bottleneck queuing time that matters. The cumulative departures are thus given by:

\[
D(t) = \begin{cases} 
\phi_1(t - t_c), t_s \leq t \leq \tilde{t} \\
\phi_2(t - t_c) + \phi_s, \tilde{t} < t \leq t_c 
\end{cases}
\]
(12)

where \(\phi_s = \frac{\gamma + \alpha}{\gamma} c_A (t' - t_q)\).

In what follows, we will consider two ways to model the traffic situation in the bathtub, given this knowledge of the equilibrium departure pattern.

### 3.3. Model I: Constant proportion of inflow into the bottleneck (CPI)

To allow us to consider a conventional bathtub area but with a downstream capacity restriction, and maintain the standard assumption that density, speed, and flow are homogeneous over space in the bathtub, we assume that queuing involves a standard spaceless, vertical “Vickrey” queue (Vickrey, 1969). For the CPI model, the arrival rate at the exit queue (the instantaneous outflow of the bathtub) should be proportional to the flow in the bathtub, as for the empirical NEF function (Gonzales and Daganzo, 2012), resulting in:

\[
a(t) = \eta k(t) v(k(t)), t \geq t_q
\]
(13)

where \(\eta\) is the ratio between outflow from the bathtub (or arrival rate at exit) and flow in the bathtub, which reflects the average trip length experienced by drivers in the bathtub (Daganzo, 2007) and depends on the availability and size of the ubiquitous exit facilities. The arrival rate at exit before \(t_q\) is zero, while after \(t_q\), it is equal to \(\eta\) times of flow in the bathtub.\(^3\)

Substituting condition (13) into condition (3) yields:

\[
k'(t) = d(t) - a(t) = \begin{cases} 
\frac{d(t), t_s \leq t \leq t_q}{d(t) - \eta k(t) v(k(t)), t_q \leq t \leq t_c}
\end{cases}
\]
(14)

Given the departure rate \(d(t)\) in Eq. (11), the above ordinary differential equation (14) can be solved for equilibrium density, speed and flow with the boundary condition \(k(t_q) = 0\). Appendix A provides more details on these derivations. With the equilibrium departure rates and arrival rates, the arrival time at the exit can be obtained using Eq. (5) (see Appendix A).

### 3.4. Model II: Fixed trip length (FTL)

The FTL model builds on the relationship between trip length and speed in the bathtub in Eq. (4), an approach adopted in Fosgerau (2015). Different from Fosgerau (2015), here we assume a non-stochastic fixed trip length, in our case identical across drivers. From Eq. (4) it follows that for a given \(L\), the derivative of \(s(t)\) with respect to \(t\) is given by:

\[
s'(t) = \frac{v(k(t))}{v(k(s(t)))}
\]
(15)

implying that the change in the arrival time at exit is determined by the ratio between speed at the moment of departure from home and speed at the moment of arrival at exit. Combining this equation with conditions (2) and (5), we have:

\[
s'(t) = \frac{v(D(t) - A(t))}{v[D(s(t)) - A(s(t))]} 
\]
(16)

\(^3\) We assume that the very first group of drivers’ travel time is positive, and there are no arrivals before they arrive; i.e., there is a small time window where drivers travel the trip length \(L\) in the bathtub without anyone exiting. For the FTL model, the positive travel time of the first group of drivers reflects that we are looking at cases with a positive trip length in the bathtub. We also adopt this assumption for the CPI model, for which it means that the first group of drivers’ realizations of trip lengths is deterministic and equal to the expected trip length. This allows a straightforward determination of their arrival time given their departure time which greatly simplifies the determination of equilibrium, while it seems a very small sacrifice also given that this only concerns a very small fraction of the drivers, and given that they still have the same expected trip length as all other drivers.
which is an ordinary differential equation (ODE). It should be noted that the departure rate at time \( t_s \) is \( \phi_1 = \frac{\alpha_1 \beta_1 c_1}{\alpha_1 + \beta_1} \), which is larger than the exit capacity. Therefore, a queue builds up as soon as the first group of drivers arrive at exit. Given the cumulative departures from home, \( D(t) \), in Eq. (12), the above ODE (16) can be solved if \( A(t) \) is known. Notice that the first driver’s arrival time at the bottleneck is \( t_q \) and there are no arrivals at the bottleneck before that moment, we have \( A(t) = 0 \) for \( [t_s, t_q) \). Thus, ODE (16) can be rewritten as:

\[
s'(t) = \frac{v(D(t) - A(t))}{v(D(s(t)) - A(s(t)))} \cdot A(s(t)) = D(t)
\]

(17)

Given the traffic speed-density relationship (1), \( s(t) \) can be obtained by solving the above ODE (17) with \( s(t_s) = t_q \). Based on Eq. (5), the cumulative arrivals \( A(t) \) for \( t \in [t_q, s(t_q)) \) is given by:

\[
A(t) = D(s^{-1}(t))
\]

(18)

With Eqs. (12) and (18), ODE (16) can be numerically solved for \( t \in [t_q, s(t_q)) \) to yield the arrival time at exit, \( s(t) \). Again, the cumulative arrivals, \( A(t) \), for the next time interval can be obtained based on Eq. (18), and then we can further obtain \( s(t) \) in this interval. Fig. 2 summarizes the method for the calculation of the cumulative arrivals and arrival time at exit in this model. Basically, it is an iterative procedure, that links solutions of earlier arrival intervals to later arrival intervals. In this way, we can calculate \( s(t) \) and \( A(t) \) for the whole peak period. Then, equilibrium density results from Eq. (2) and equilibrium speed results from Eq. (1).

Although both model formulations, CPI and FTL, look reasonable, it is in general impossible for them to hold at the same time; otherwise, condition (5) is not satisfied. For the CPI model, the arrival rate at the exit is proportional to flow in the bathtub, which is attractive from the perspective of logical relations between flows of vehicles. But the CPI model results in time varying and stochastic trip lengths. Here we follow that in Arnott (2013), and assume that drivers’ expect trip lengths are constant and they make their departure time choices based on their travel cost and their expected trip length. For the FTL model, trip length is fixed and deterministic, but the arrival rate at the exit is not deterministically related to the flow in the bathtub. This has the disadvantage that the ratio between flow in the bathtub and outflow from the bathtub, \( \eta \), in fact becomes endogenous, and time varying in this formulation. Depending on the empirical plausibility, both sets of assumptions might be more or less unattractive for the reasons mentioned above, but both can guarantee that an equilibrium can be found. Section 5 will assess numerically whether the two formulations lead to strong differences in predicted dynamic congestion patterns in the bathtub, therewith providing insight into the importance of choosing between these two model formulations. But before that, we will first distinguish two different types of hypercongestion, by allowing for capacity waste at the exit.
4. Hypercongestion

4.1. Two types of hypercongestion

This section discusses the qualitative insights on hypercongestion that we can derive from our models. Hypercongestion refers to the situation where flow decreases with density, because the impact of reduced speed dominates the direct effect of increased density on flow. For our model, this implies that one way to identify hypercongestion is to identify conditions where equilibrium speed increases with equilibrium flow in the bathtub. This is fully in line with the empirical findings of Geroliminis and Daganzo (2008), who established a backward-bending part of the macroscopic speed-flow curve (the part in red oval in Fig. 3(a)). This type of hypercongestion affects travelers when they are in the bathtub area. However, it should be noted that for these equilibria, the bathtub exit operates at full capacity for the whole hypercongestion period. The lower speed in the bathtub makes travelers join the exit queue later, but since the outflow from exit is not affected, the arrival time at destination does not change, either. Both inflow into the bathtub and outflow from the exit are therefore unaffected by this hypercongestion in the bathtub, and although hypercongestion exists in the bathtub area, the performance of the full system of the bathtub-exit in terms of arrival rates at the final destination is not affected. This happens because there is perfect substitution between the two types of travel delays: an increase of travel time in the bathtub leads to a shrinking queueing length at the exit, and thus a decrease of queueing time perfectly substitutes the increase of travel time in the bathtub. Indeed, in dynamic equilibrium, it is the sum of the two types of delay that compensates for dynamic variation in schedule delay costs, so this sum will not change when hypercongested speeds occur in the bathtub.

However, there is a second type of hypercongestion that can occur: the increase of travel time in the bathtub may be so large, and the drop in the outflow from the bathtub into the exit can be so strong, that the arrival rate at the exit at some moments falls below its capacity, and if it does so for a sufficiently long time, the exit maybe no longer be used at full capacity throughout the peak (see Fig. 3(c)). For this type of hypercongestion, inefficiency caused by hypercongestion in the bathtub will affect the full system’s arrival rate. As a result, the equilibrium duration of the peak increases, and so does the equilibrium user cost (recall that the average user cost is constant over time in a dynamic equilibrium, so that a longer peak duration, with higher schedule delay cost for the very first traveler who faces no travel delay, implies a higher average cost level for all drivers).

In single-facility models, these two types of hypercongestion coincide: a speed below the one that maximizes the flow, implies that the facility operates below its capacity. The same is true for “single-facility” MFD or bathtub model. In our setting, in contrast, hypercongested speed in the bathtub may or may not imply an arrival rate below the exit capacity for part of the peak. It is therefore important to distinguish these two types of hypercongestion. We call the former type of hypercongestion “bathtub-speed hypercongestion”, as it only involves change of flow-speed relationship in the bathtub; and call the latter type of hypercongestion “system hypercongestion”, because it causes an additional inefficiency in the performance of the full bathtub-bottleneck system.

Different from the bathtub-speed hypercongestion, system hypercongestion means that the exit capacity is no longer fully utilized during the whole peak period, and thus the equilibrium departure rate in Eq. (11) does not hold for some time. In the following, we will analyze how to obtain the equilibrium solution in this analytically more challenging case.

4.2. Equilibrium with system hypercongestion

This section explores equilibrium with system hypercongestion. Here we only consider the CPI model, which allows for the derivation of closed-form expressions of equilibrium density in the bathtub (Appendix A).

Bathtub-speed hypercongestion will occur whenever density exceeds half of the jam density, $\frac{1}{2}$, in equilibrium, density is obtained by solving ODE (14) which is given in Appendix A. It can be observed that the equilibrium density pattern between $[t, \bar{t}]$ depends on the sign of $\eta \phi - 4\lambda \phi_1$. If $\eta \phi - 4\lambda \phi_1 > 0$, density decreases over time from $t$, and is given by

$$k(t) = \frac{1 - \rho_1}{\lambda t} + \frac{\rho_1}{\lambda (1 + \exp(\psi_1 t + c_1))} \quad t \leq t \leq \bar{t}$$

where $c_1$ is a parameter such that $k(t)$ is continuous at $t$, and $\rho_1 = \sqrt{\eta \phi - 4\lambda \phi_1}$, $\psi_1 = \sqrt{\eta \phi (\eta \phi - 4\lambda \phi_1)}$. Thus, bathtub-speed hypercongestion exists only if density at $t$ is higher than $\frac{1}{2}$. As there are no arrivals at the bottleneck before the first arrival, density is equal to the cumulative departures for $t < t$, i.e.,

$$k(t) = D(t) = \psi_1 (t - t), t, t \leq t < t$$

From the speed-density relationship in Eq. (1) and the above density function (20), the first driver’s travel time is given by:

Eq (21) implies that the first driver already chooses the same speed as his/her followers when he/she travels. Again, the assumption has the benefit of consistency of assumptions, but the disadvantage of being unrealistic: the very first driver is slowed down because of upstream density, i.e., because of traffic conditions behind. Although being unrealistic for a physical queue on a single link, it is not necessarily an issue in a whole network with an instantaneous spread of homogeneous traffic conditions. Daganzo (2007) and Geroliminis and Daganzo (2008) observe that this assumption should be acceptable under slow-varying conditions, and is indeed a typical assumption in the literature.
where \( \sigma = \sqrt{1 - 2\lambda \phi_1 L} \). Substituting the above Eq. (21) into Eq. (20), we find:

\[
k(t_e) = \frac{1 - \sigma}{\lambda}
\]

As bathtub-speed hypercongestion exists only if \( k(t_e) > \frac{1}{2} \), we find:

\[
\lambda \phi_1 L > \frac{3}{8}
\]

Therefore, if \( \eta \gamma - 4\lambda \phi_1 > 0 \), bathtub-speed hypercongestion exists if and only if \( \lambda \phi_1 L > \frac{3}{8} \). If \( \eta \gamma - 4\lambda \phi_1 = 0 \), density increases over time between \([t_0, \bar{t}]\), and is given by:

\[
k(t) = \frac{1}{2\lambda} - \frac{1}{4\lambda^2 (\phi_1 t + c_1)}, \quad t_0 \leq t \leq \bar{t}
\]

Thus, bathtub-speed hypercongestion exists as long as \( k(\bar{t}) > \frac{1}{2} \), which can be simplified as:

\[
c_a (\bar{t} - t_0) (\frac{1}{2\lambda (1 - \sigma)} + \frac{1 - \sigma}{\lambda \phi_1})
\]

If \( \eta \gamma < 4\lambda \phi_1 \), density is given by:

\[
k(t) = \frac{1}{2\lambda} + \frac{1}{2\lambda} \arctan \left( \frac{2\sigma - 1}{\nu_1} \right) + \frac{1 - \sigma}{\lambda \phi_1}
\]

which also increases over time between \([t_0, \bar{t}]\). Therefore, bathtub-speed hypercongestion exists if \( k(\bar{t}) > \frac{1}{2} \), which can be simplified as:

\[
\bar{t} - t_0 > 2\phi_1 \arctan \left( \frac{2\sigma - 1}{\nu_1} \right) + \frac{1 - \sigma}{\lambda \phi_1}
\]

Compared to the conditions for bathtub-speed hypercongestion, the establishment of the conditions for the existence of system hypercongestion is much more complicated. This type of hypercongestion occurs when there exists at least one moment, \( \hat{t} \), in the middle of the peak period such that 1) the arrival rate at the exit is lower than the capacity of the exit; and 2) the cumulative arrivals at the exit fall below of the potential maximum of cumulative arrivals at destination after \( t_0 \). These two conditions can be written as

\[
a(\hat{t}) = \eta k(\hat{t}) v(\hat{t}) < c_a
\]

\[
A(\hat{t}) - c_a (\bar{t} - t_0) \leq 0
\]

Unfortunately, it is not easy to pinpoint when the above conditions (28)–(29) hold for general cases. Instead, we consider an alternative condition here. Let \( \hat{t} \) denotes the moment that the arrival rate at the bottleneck is equal to the bottleneck capacity, i.e., \( a(\hat{t}) = c_a \). Then, system hypercongestion exists if there exists \( \hat{t} \) such that

\[
A(\hat{t}) \leq c_a (\bar{t} - t_0)
\]

Thus, the above Eq. (30) is equivalent to conditions (28)–(29). Without loss of generality, here we assume that \( \hat{t} > \bar{t} \). In Appendix A, the density after \( \hat{t} \) is derived which is given by:

\[
k(t) = \frac{1 - \rho_2}{2\lambda} + \frac{\rho_2}{\lambda (1 + \exp(\phi_2 t + c_2))}, \quad \hat{t} < t \leq t_e,
\]

where \( \rho_2 = \sqrt{\frac{\eta \gamma - 4 \lambda \phi_2}{\eta \gamma}}, \phi_2 = \sqrt{\eta \gamma (\eta \gamma - 4 \lambda \phi_2)}, \) and \( c_2 \) is a parameter such that \( k(t) \) is continuous at \( \hat{t} \). Then, the arrival rate and the cumulative arrivals at exit are given by

\[
a(t) = \frac{\eta \gamma (\rho_2 - 1) \rho_2 (2\lambda + 1 + \exp(\phi_2 t + c_2)) + (\rho_2 - 1) \rho_2}{4 \lambda^2 (1 + \exp(\phi_2 t + c_2))}, \quad \hat{t} < t \leq t_e,
\]

\[
A(t) = \frac{\eta \gamma (1 - \rho_2^2 t - \frac{4 \rho_2^2}{\phi_2 (1 + \exp(\phi_2 t + c_2))}) + c_A}{\phi_2}, \quad \hat{t} < t \leq t_e,
\]

where \( c_A \) is a parameter such that \( A(t) \) is continuous at \( \hat{t} \). Solving \( a(\hat{t}) = c_a \) yields
\[
\bar{\xi} = \frac{1}{\varphi_2} \ln \left( \frac{\eta \gamma (1 + \rho_2^2)}{4 \lambda c_0 + \eta \gamma (\rho_2^2 - 1)} \right) - \frac{c_2}{\varphi_2}.
\]

By substituting the above Eqs. (33)–(34) into Eq. (30), we find:
\[
2\lambda c_0 \varphi_2 (t' - t_2) - 2 \frac{c_2}{\varphi_1} \varphi_2 (1 - \sigma) - v_2 + \sqrt{v_0 v_2} + 2\lambda (\varphi_2 - c_0) \ln \left( \frac{v_2 + \sqrt{v_0 v_2}}{\lambda (c_0 - \varphi_2)} - 2 \right)
- 2\lambda (\varphi_2 - c_0) \ln \left( \frac{\varphi_2 - v_2}{\lambda^2 \eta \gamma k} - 2 \right) + \lambda v_2 \frac{\eta \gamma}{\varphi_2} (\eta \gamma v_2 - 2 v_2 + \varphi_2) \tilde{k} < 0,
\]

where \( v_0 = \eta \gamma - 4\lambda c_0, v_2 = \eta \gamma - 4\lambda \varphi_2, \) and \( \tilde{k} = k(\bar{t}) \) denotes the value of density at \( \bar{t} \) which depends on the density pattern between \([t_\xi, \bar{t}]\). Although no closed-form results can be obtained, it is not hard to test Eq. (35) numerically. Apparently, the existence of bathtub-speed hypercongestion cannot ensure that condition (30) holds, which means that, quite intuitively, the existence of bathtub-speed hypercongestion is a necessary but not a sufficient condition for system hypercongestion to occur.

When system hypercongestion occurs, the capacity of the exit is no longer fully used, and thus the departure rate function (11) does not hold for a certain period, here we will further introduce how to obtain an equilibrium in this case, which use the methods from both the CPI model and the FTL model. Denote \( \bar{t} \) and \( \bar{t} \) as the departure times of the first driver and the last driver who pass the bottleneck without queue in the middle part of the peak, respectively. For drivers departing at \( t \in [\bar{t}, \bar{t}] \), their arrival time at the bottleneck, and therefore also at the destination as no queue exists for them, is given by:
\[
s(t) = \begin{cases} 
\alpha / (\alpha - \beta) (t - \bar{t}) + s(\bar{t}), & \text{if } \bar{t} \leq t < \bar{t} \\alpha / (\alpha + \gamma) (t - \bar{t}) + t', & \text{if } \bar{t} \leq t < \bar{t} \\alpha / (\alpha + \gamma) (t - \bar{t}) + s(\bar{t}), & \text{if } \bar{t} \leq t < \bar{t} \end{cases} \]

However, the departure rate from home during this period does not satisfy Eq. (11) for \( t \in [\bar{t}, \bar{t}] \), and the “constant proportion of inflow into the bottleneck”, i.e., condition (13), does not hold for \([s(\bar{t}), s(\bar{t})]\). Therefore, for \( t \in [\bar{t}, \bar{t}] \), conditions (13) - (14) hold but the departure rate is unknown; for \([s(\bar{t}), s(\bar{t})]\), the departure rate is given by Eq. (11) but conditions (13) - (14) do not hold. Without further assumptions on the departure rate, the traffic density and speed in the bathtub for \( t \in [\bar{t}, \bar{t}] \) cannot be obtained by solving Eq. (14). In fact, multiple solutions of the departure rate, and thus different traffic speeds in the bathtub during \( t \in [\bar{t}, \bar{t}] \) can lead to an equilibrium. This is because drivers’ trip length is not required to be fixed, so as long as the travel time in the bathtub follows Eq. (5), different combinations of departure patterns and traffic densities satisfying condition (14) can lead to equilibrium (see also Fosgerau, 2015). For example, if we in addition assume that the trip length for drivers departing during \([\bar{t}, \bar{t}]\) is fixed and equal to the average trip length \( L \), based on condition (4), we have:
\[
s' = \frac{\nu(k(t))}{\nu(k(s(t)))} = \frac{1 - \lambda (D(t) - A(t))}{1 - \lambda (D(s(t)) - D(t))}, \quad t \in [\bar{t}, \bar{t}]
\]

Here we give only a special case that \( \bar{t} \leq \bar{t} \) and \( s(\bar{t}) > \bar{t} \). As conditions (13)–(14) hold for \([\bar{t}, \bar{t}]\), we can get \( A(t) \), and as the departure rate follows Eq. (11) for \([s(\bar{t}), s(\bar{t})]\), we can get \( D(s(t)) \). Combining Eqs. (36) – (37) yields:
\[
D(t) = \frac{\alpha}{2a - \beta} + (A_2 (s(t) - t_2) + C) + (a - \beta) A(t)
\]

and hence:
\[
d(t) = \frac{a}{2a - \beta} \frac{a}{a - \beta} \phi_2 + \frac{a - \beta}{2a - \beta} \eta k(t) v(k(t))
\]

Substituting the above Eq. (39) into conditions (13) · (14) yields:
\[
k'(t) = \frac{a}{2a - \beta} \frac{a}{a - \beta} \phi_2 - \frac{a - \beta}{2a - \beta} \eta k(t) v(k(t))
\]

Solving this differential equation gives the equilibrium traffic density in the bathtub for \( t \in [\bar{t}, \bar{t}] \). But other assumptions on \( L \) during \([\bar{t}, \bar{t}]\) can also be consistent with equilibrium (although a different one in terms of departure patterns).

From the above analyses, bathtub-speed hypercongestion is a necessary but not a sufficient condition for system hypercongestion to occur. Bathtub-speed hypercongestion occurs as long as density exceeds half of the jam density. We will know whether it occurs given the equilibrium departure time of the first driver. Whether system hypercongestion occurs depends on the cumulative arrivals at exit. To determine its existence, the density at \( \bar{t} \) is also needed. In the next section, we will use a numerical example to analyze the models.
Hypercgesion occurs if the capacity of the bathtub (represented by the jam density, and analytical notion that bathtub-speed hypercongestion is a necessary but not a sufficient condition for system hypercongestion to occur. However, if the capacity of the bathtub (represented by the jam density, \( k_j = 1/\lambda \)) further decreases, the assumption of no capacity waste at the bottleneck will not hold anymore, and some drivers’ equilibrium travel times in the bathtub are so long that the arrival rates at the bottleneck are less than the exit capacity for some period. Therefore, system hypercongestion may arise, and will occur if the period of low arrival rates is long enough to make the queues disappear in the central period in the peak, so that outflow from the bottleneck falls below capacity. To investigate this kind of hypercongestion, the assumptions on parameters to secure that no capacity waste at the bottleneck exists will be changed.

5. Numerical results

In the following, we will first compare the two alternative modelling approaches described in Section 3, i.e., the CPI and FTL models; and then illustrate the two types of hypercongestion just discussed.

Example 1. This example is used to illustrate the CPI model and the FTL model. The value of time and the shadow prices for early and late arrivals follows that in Verhoef (2005): \( \alpha = 7.5 ($/h) \), \( \beta = 3.75 ($/h) \), \( \gamma = 15 ($/h) \). The total travel demand is set to be \( N = 10000 \) cars, while the capacity of the bottleneck is \( c_a = 5000 \) (cars/h) so that the minimum time needed for all drivers to pass the bottleneck is two hours. The length of the trip in the bathtub area is assumed to be \( L = 10 \) km, and the free-flow traffic speed is \( v_f = 40 \) (km/h). Let \( \lambda = 0.00012 \), so that the jam density is about \( k_j = 8333.3 \) (cars/km).\(^5\) Drivers’ preferred arrival time at the destination is set to be \( \tau' = 9 \). The ratio between the flow of the bathtub and the inflow to the bottleneck is assumed to be equal to \( \eta = 0.07982 \), which is chosen in such a way that the maximum density in the CPI model is the same as in the FTL model. To demonstrate that the proposed models converge to equilibrium, plots of the resulting generalized travel costs are shown in Appendix C.

5.1. Comparing models: CPI versus FTL

Table 1 presents the numerical results of the CPI model and the FTL model. As the benchmark model we use closed-form approximations given in Appendix B, assuming that the last driver’s travel time is free-flow travel time, which leads to the equilibrium generalized travel cost below (Appendix B):

\[
\eta(t) = \frac{L}{v_f} + \frac{c_a}{c_a} \frac{N}{\beta + \gamma} + \frac{\alpha \gamma}{\beta + \gamma} \tau'
\]

(41)

where the last part of this equation is the extra travel time the first driver needs above free-flow travel time multiplied by \( \frac{\alpha \gamma}{\beta + \gamma} \). It can be observed that the divergences caused by the different analytical assumptions are very small. The generalized travel costs for the two models increase by 0.30% and 0.11%, respectively, compared to that in the approximately closed-form solution. Moreover, the generalized travel cost in the FTL model is closer to the benchmark than that in the CPI model. Similar conclusions can be obtained for the equilibrium peak period. In fact, compared to that in the benchmark, the equilibrium peak periods in both models are merely

---

\(^5\) Density is measured in cars/km\(^2\) (rather than cars/km) as it is defined as the total number of cars in the bathtub (Arnott, 2013). Thus, flow is measured in cars/km-h. For uni-directional traffic, the cars/km in the numerator of flow should be interpreted as care per kilometre of “width” of urban space; i.e., perpendicular on the driving direction.
earlier/later less than 1 min. Therefore, given the different assumptions on the relationship between the outflow and flow (e.g., the NEF in Gonzales and Daganzo, 2012) or the relationship between the travel time and speed (e.g., Fosgerau, 2015), no substantial quantitative differences result for the equilibrium generalized travel costs and the equilibrium peak periods. Both models have a hypercongested period in the bathtub that lasts about 16 min. The last driver’s travel time is also affected by the traffic situation in the bathtub. We find that the last driver’s travel times in both cases are slightly higher than the free-flow travel time (less than 2 min). This means that when the last driver travels in the bathtub, speed is already very close to free-flow speed.

To better understand the traffic situation in the bathtub, graphical representations of the various relevant traffic variables, the equilibrium cumulative departures and arrivals, and the resulting equilibrium travel time are given in this section (Figs. 4a and 4b and Fig. 5). First, Figs. 4a and 4b present the equilibrium cumulative departures and arrivals, traffic density, speed and flow in the bathtub over the peak period for the CPI model and the FTL model, respectively. From Figs. 4a and 4b, it can be observed that the shapes of the cumulative arrivals at exit in the CPI model and the FTL model are somewhat different. In Fig. 4a, the curve of the cumulative arrivals at exit is approximately linear at the early part of the peak period. Moreover, no queue exists in the last few minutes of the peak period in the CPI model, and thus drivers pass the exit as soon as they arrive. In contrast, the curve of the cumulative arrivals at exit in Fig. 4b is clearly convex at the early part of the peak period and is approximately linear for the late part of the peak period.
Interestingly, the traffic variables in the bathtub are quite different in the late part of the peak period for the two model formulations. When we compare Figs. 4a and 4b, we observe that the curve of density between $\tilde{t}$ and $t_e$ is convex in the CPI model ([7.9016, 9.1492]) while it is first concave and then convex in the FTL model ([7.9037, 9.1638]). In particular, for the FTL model (Fig. 4b), density decreases rather sharply after $\tilde{t}$, reaching a relatively low value in a short time. Then it becomes very flat for a long time. As a result, the corresponding speed (flow) in the FTL model also rises (declines) very sharply over a short time and then stays flat over a substantially long time. This can also be verified by the horizontal distances between the cumulative departures from home and the cumulative arrivals at exit, which are very small and only have a minor decrease for the late part of the peak period. In contrast, density decreases gradually after $\tilde{t}$ until $t_e$ in the CPI model (Fig. 4a). As a result, the horizontal distances between the cumulative departures from home and the cumulative arrivals at exit are relatively large and decrease over time.

Fig. 5 shows drivers’ equilibrium travel times in the bathtub for the two approaches. The CPI model shows that drivers’ equilibrium travel times in the bathtub both increase and decrease over a relatively long period. In contrast, the FTL model shows that drivers’ equilibrium travel time in the bathtub increases to its maximal value in a relatively short time and is flat over an extended period in the late part of the peak period, which is consistent with the findings in Mariotte et al. (2017). As a result, drivers departing at $\tilde{t}$ (7.9016) have a maximum travel time in the bathtub in the CPI model, while the maximum travel time in the bathtub occurs for drivers departing much earlier than $\tilde{t}$ (around 7.6). Moreover, the maximum travel time in the bathtub is higher for the CPI model compared to
that for the FTL model. This is because for the CPI model, the actual trip length is not constant. This means that drivers may have a longer trip length, and thus a longer trip time.

5.2. Bathtub-speed hypercongestion

Table 1 shows that bathtub-speed hypercongestion exists in both models. Based on the speed-density relationship in Eq. (1), the critical value of density for bathtub-speed hypercongestion to occur is 4166.7. As the value of $\eta$ is set such that the maximal density in the FTL model and the CPI model is the same, we have a maximum density of 4744 for both models. The durations of hypercongestion in both models are also similar, both last about 16 min. Whether bathtub-speed hypercongestion exists depends on the values of various parameters, such as the total number of drivers, the preference parameters, and $\eta$ in the CPI model (see Section 4.2). In order to show the impact of $\eta$ on the occurrence of bathtub-speed hypercongestion, Fig. 6 shows the equilibrium density and flow in the CPI model for different values of $\eta$. The figure confirms the intuition that bathtub-speed hypercongestion is more likely to occur when $\eta$ is smaller ($\eta < 0.1$); i.e., when the capacity of the exit is smaller. Only then ($\eta = 0.07, 0.08, 0.09$) we observe that equilibrium density exceeds the critical value of 4166.7 in the upper left panel in combination with a drop of flow in the central peak in the upper right panel and longer equilibrium travel times in the lower left panel as a result. Consequently, the generalized travel costs decrease with $\eta$.

![Fig. 7a. The CPI model with different values of $\lambda$.](image1)

![Fig. 7b. The FTL model with different values of $\lambda$.](image2)
The value of $\lambda$ also plays an important role in the occurrence of hypercongestion, as it is the reciprocal of the jam density. Figs. 7a and 7b give the equilibrium density, flow and travel time in the bathtub and the equilibrium generalized travel costs with different values of $\lambda$ in the CPI model and the FTL model, respectively. The value of $\lambda$ greatly affects the existence and the duration of hypercongestion in both models. The impact for the FTL model is more pronounced than that for the CPI model, but the qualitative effects in both models are consistent. Density during the whole peak period is higher when $\lambda$ increases (and hence the jam density declines). Equilibrium travel times in the bathtub therefore also increase with $\lambda$. The beginning time of the peak period is naturally earlier when $\lambda$ increases and therefore the generalized travel cost also increases with $\lambda$.

5.3. System hypercongestion

Fig. 8 presents the CPI model with system hypercongestion, and hence capacity waste, in the middle of the peak. We set $\lambda = 0.000145$, i.e., the jam density is equal to $k_j = 6896.6$, while the values of the other parameters follow those above. Fig. 8a confirms that the cumulative arrivals at the exit between $[s(t), s(T)]$ are equal to the cumulative arrivals at the destination, and the arrival rate is

(lower right panel).

The value of $\lambda$ also plays an important role in the occurrence of hypercongestion, as it is the reciprocal of the jam density. Figs. 7a and 7b give the equilibrium density, flow and travel time in the bathtub and the equilibrium generalized travel costs with different values of $\lambda$ in the CPI model and the FTL model, respectively. The value of $\lambda$ greatly affects the existence and the duration of hypercongestion in both models. The impact for the FTL model is more pronounced than that for the CPI model, but the qualitative effects in both models are consistent. Density during the whole peak period is higher when $\lambda$ increases (and hence the jam density declines). Equilibrium travel times in the bathtub therefore also increase with $\lambda$. The beginning time of the peak period is naturally earlier when $\lambda$ increases and therefore the generalized travel cost also increases with $\lambda$.
less than the exit capacity. At the same time, the departure rate from home between $[t, \tilde{t}]$ follows Eq. (26), and is less than the departure rate at other moments in the early departure period (before $t$ or after $\tilde{t}$), which implies fewer drivers departing between $[t, \tilde{t}]$ to avoid the jam in the bathtub. Fig. 8b confirms that the equilibrium generalized travel costs during the peak period are identical for all drivers departing between the entire equilibrium departure period $[t_s, t_e]$, and would be higher outside. The duration of the equilibrium departure period (2.1520 h in this case) is longer than that without capacity waste (2 h), and the equilibrium generalized travel cost (8.6911) is higher (increased by 5.50%). Consistent with this, the total travel time cost is higher (55390 vs 52365), and so is the total schedule delay cost (31521 vs 30012), underlining the inefficiency of system hypercongestion. Fig. 8c and 8d show the corresponding density and flow in the bathtub. The hypercongested period with bathtub-speed hypercongestion, is given by $[\tilde{t}_s, \tilde{t}_e)$. However, system hypercongestion only exists between $[\tilde{t}_s, \tilde{t}_e)$ when the queuing time at the exit is zero. In summary, the numerical results confirm that bathtub-speed hypercongestion can exist without the specific loss of efficiency associated with a drop in the outflow of the whole system. This happens when flow decreases with density in the bathtub, but the exit capacity nevertheless remains fully utilized. In this case, drivers’ total travel time does not change and thus it is merely a tradeoff between longer travel time in the bathtub and longer queuing time at the bottleneck. On the other hand, when the capacity of the bathtub (represented by the jam density) further decreases (or, the exit capacity further increases), system hypercongestion can occur. While flow decreases with density in the bathtub, the exit capacity is then not fully utilized resulting in an efficiency loss at the level of the full system.

6. Conclusions and future work

This paper investigated the impacts of introducing finite exit capacity in the bathtub model on equilibrium outcomes and user costs, using two model formulations that were proposed in the earlier literature. The motivation for making this extension is that earlier studies of one-directional facilities, like the one in Verhoef (2003), have shown that without the existence of a downstream capacity restriction, hypercongestion cannot build up on a single facility, and initially assumed hypercongestion would dissolve following upstream moving shockwaves from the exit as traffic there speeds up in the absence of a capacity restriction. Especially because a capacity restriction at the exit significantly affects efficiency assessments for congested facilities, and even more so in the face of hypercongestion, it is a crucial addition to a modelling framework based on the MFD or bathtub methodology. The same issue was also pointed in Mariotte and Leclercq (2019), in which an outflow model was proposed to account for the outflow limitation and to overcome the unrealistic very long duration of hypercongestion.

A first main finding from our analysis is that the occurrence of hypercongested speed in the bathtub does not necessary lead to the particular type of efficiency losses that would occur at the level of the full system when hypercongestion causes the outflow to fall below its maximum achievable level. Two types of hypercongestion can thus be distinguished: “bathtub-speed hypercongestion”, and “system hypercongestion”. When there is bathtub-speed hypercongestion, but the exit capacity remains fully utilized, longer travel time in the bathtub caused by hypercongested speed results in shorter equilibrium queuing time at exit, and therefore the total equilibrium travel time of the trip does not change. On the contrary, when the drop in equilibrium bathtub flow is sufficiently large and long-lasting, the exit capacity may no longer be fully used in the most central time window of the peak period, and system hypercongestion arises. Hypercongestion in the bathtub then results in a longer equilibrium peak period, and affects the full system’s...
performance and thereby causes a specific type of efficiency losses on top of the conventional losses due to the existence of congestion externalities, namely a drop in the system’s effective capacity.

It was demonstrated that bathtub-speed hypercongestion is a necessary but not a sufficient condition for system hypercongestion to occur. This implies that with a restricted exit capacity, the possible gains from managing hypercongestion in an urban area partly depends on the performance of the exit capacity, which is consistent with the conclusions in Ingole et al. (2020). Expanding the downstream capacity may be a better way to deal with hypercongestion than expanding flow capacity in the bathtub; a finding which is consistent with the one-directional model of Verhoef (2003), where an expansion of upstream capacity that holds the hypercongested queue would make queues shorter, wider and slower; while trip completion rates remain capped by downstream capacity.

A second issue addressed was the comparison of two alternative model formulations for bathtub congestion: the FTL versus CPI models. Consistent with other findings in the literature (e.g., Amirgholy and Gao, 2017), we find that the numerical differences in the equilibrium generalized travel cost and peak period for the two approaches are almost negligible, and are particularly negligible compared to an approximate model where it is assumed that the last driver’s travel time is the free-flow travel time. The resulting analytically closed-form expressions for the generalized travel costs and the peak period of our approximate model can give policy analysts a quick estimation of these variables, without having to know the specific traffic situation in the bathtub and at exit.

However, the generated dynamic patterns of the traffic variables in the bathtub and the queue between the two models can differ substantially. In fact, one of the main differences between the numerical results of the CPI model and the FTL model lies in the tradeoff between the equilibrium travel time in the bathtub and the equilibrium queueing time at the exit. For the CPI model, the travel time in the bathtub accounts for the majority of the travel time cost whereas in the FTL model, the queuing time at the exit accounts for the majority of the travel time costs. However, the sum of the travel time in the bathtub and the queuing time at the exit at every instant in the CPI model and the FTL model are similar, as well as the quantitative estimate of total travel cost. It is interesting to empirically investigate in future research which of these two models will describe urban traffic patterns more accurately. However, these questions are not straightforward to address. For now, our analysis suggests that observationally diverging predictions in travel conditions may be consistent with rather similar predictions of equilibrium travel costs.

Further development of our research will include pricing strategies, the spillback of the exit queue and heterogeneous trip lengths as well as heterogeneous preferences and preferred arrival times. Pricing strategies will be designed to alleviate hypercongestion, especially the system-hypercongestion in the facilities. Congestion from the queue will propagate and extend to the whole network, and the size of the network will shrink with the increase of the exit queue. Heterogeneous trip lengths may change the FIFO principle, and we will investigate its impact on the trip completion rate. Furthermore, empirical investigation of the exit flow during peak hours can avoid exploring the relationship between the trip completion rate, the flow inside the bathtub, and the downstream capacity.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work is supported by the Fundamental Research Funds for the Central Universities 2021RC243, the 111 Project B20071, and the funding of the Netherlands Organization for Scientific Research (NWO) as part of the U-SMILE project 438-15-176, which is gratefully acknowledged.

Appendix A. The traffic density in the bathtub for the CPI model

Here we assume $t_q \leq \bar{t}$, and the density function with $t_q > \bar{t}$ can be obtained in a similar way. As mentioned before, density in the bathtub is obtained by solving ODE. (14). When $\eta \lambda - 4\lambda \phi_1 > 0$, we find:

$$k(t) = \begin{cases} \frac{1 - \rho_1}{2\lambda} + \frac{\rho_1}{\lambda(1 + \exp(\phi t + c_1))} & t_q \leq t \leq \bar{t} \\ \frac{1 - \rho_2}{2\lambda} + \frac{\rho_2}{\lambda(1 + \exp(\phi t + c_1))} & \bar{t} < t \leq t_e \end{cases}$$

(A1)

When $\eta \lambda - 4\lambda \phi_1 = 0$, we find:

$$k(t) = \begin{cases} \frac{1}{2\lambda} - \frac{1}{4\lambda^2(\phi t + c_1)} & t_q \leq t \leq \bar{t} \\ \frac{1 - \rho_2}{2\lambda} + \frac{\rho_2}{\lambda(1 + \exp(\phi t + c_1))} & \bar{t} < t \leq t_e \end{cases}$$

(A2)

When $4\lambda \phi_2 < \eta \lambda < 4\lambda \phi_1$, we find:
The departure rate should be no higher than the exit capacity, which means 
\[ k(t) = \begin{cases} 
\frac{1}{\omega} + \frac{1}{\omega} \phi_1 \tan \left( \frac{1}{2} \phi_1 (t + c_1) \right), & t_q \leq t < \tilde{t} \\
1 - \frac{\rho_1}{\omega} + \frac{\rho_2}{\omega (1 + \exp(\rho_2 + c_2))}, & \tilde{t} \leq t \leq t_e 
\end{cases} \tag{A3} \]

where \( c_1 \) and \( c_2 \) are parameters such that \( k(t) \) is continuous at \( t_q \) and \( \tilde{t} \), and for convenience we denote: 
\[ \phi_1 = \sqrt{\frac{\eta \gamma + 4 \lambda \phi_1}{\eta}}, \quad \phi_2 = \sqrt{\frac{\eta \gamma (\eta \gamma - 4 \lambda \phi_2)}{\eta}}. \]

As there are no departures after \( t_e \) while arrivals at exit still occur, for \( t_e \leq t < s(t_e) \), we have:

\[ k(t) = N - A(t) = N - \eta \gamma \int_{t_e}^s k(\omega)(1 - 2k(\omega))d\omega \tag{A4} \]

which yields:

\[ k(t) = \frac{1}{\exp(\eta \gamma(t + c_1)) + \lambda}, \quad t_e \leq t < s(t_e) \tag{A5} \]

where \( c_3 \) is parameter such that \( k(t) \) is continuous at \( t_e \). Based on condition (2), the cumulative arrivals are equal to the difference between the cumulative departures and density. Therefore, we find:

\[ A(s(t_e)) = N - \frac{1}{\exp(\eta \gamma(t + c_1)) + \lambda}, \quad t_e \leq t < s(t_e) \tag{A6} \]

which implies the last driver departing will approach the exit, but will never arrive. This is caused by the homogenous density assumption, as the last driver needs to circuit in the bathtub to keep homogenous density over space. It means that the peak period will last forever, which is impossible and implausible.

Fortunately, it also implies capacity waste may occur at exit for the very last few minutes of the peak period, which means the equilibrium condition (10) cannot apply. Denote \( s(\tilde{t}) \) as the instant when the length of queue at exit becomes zero, namely the cumulative arrivals at exit is equal to the cumulative arrivals at destination:

\[ A(s(\tilde{t})) = c_a(s(\tilde{t}) - t_q) \tag{A7} \]

For drivers departing after \( \tilde{t} \), they can pass the exit as soon as they arrive there. Thus, their generalized travel cost is given by:

\[ u(t) = a(s(t) - t) + \gamma(s(t) - s(\tilde{t})), \quad t \in [\tilde{t}, t_e] \tag{A8} \]

The first-order condition of equilibrium is:

\[ u'(t) = (\alpha + \gamma) s'(t) - \alpha = 0 \tag{A9} \]

which yields:

\[ s'(t) = \frac{\alpha}{\alpha + \gamma}, \quad t \in [\tilde{t}, t_e] \tag{A10} \]

From condition (5), the departure rate satisfies:

\[ d(t) = a(s(t)) \frac{\alpha}{\alpha + \gamma} a(s(t)), \quad t \in [\tilde{t}, t_e] \tag{A11} \]

As there is no queue at the exit after \( s(\tilde{t}) \), the arrival rate at the exit, \( a(s(t)) \), should be no higher than the exit capacity, which means the departure rate should be no higher than \( \phi_2 \). For simplicity, here we further assume the departure rate between \( [\tilde{t}, t_e] \) is still equal to \( \phi_2 \), and thus, the arrival rate at the exit is equal to the exit capacity. In this way, the exit capacity is still fully used and no queue exists. Therefore, the cumulative departures still follow Eq. (12) between \( [\tilde{t}, t_e] \), and the cumulative arrivals at the exit between \([s(\tilde{t}), s(t_e)] \) are given by:

\[ A(t) = c_a(t - t_q) \tag{A12} \]

Thus, the density between \([s(\tilde{t}), s(t_e)] \) will be given by:

\[ k(t) = N - c_a(t - t_e) \tag{A13} \]

and all drivers arrive at their destination at \( s(t_e) = t_e + N/c_a \). Although drivers can also choose earlier departure rates in this period to reach equilibria, they all lead to longer equilibrium peak period and higher equilibrium travel cost. For peaks in which the duration of these light travel conditions in the very last few minutes is relatively short-lived compared to the full peak duration, and are of little importance because of their short duration and the low flow levels in those intervals, the assumption seems acceptable, definitely in the light of the analytical advantages it brings.
Appendix B. Approximate closed-form solutions: assuming free-flow travel time for the last driver

Here we first give the peak period and the generalized travel cost when the travel time of the last driver is assumed to be free-flow travel time. This case is useful as it leads to closed-form solutions for the start and the end of the peak, as well as for the equilibrium travel costs, which can serve as approximations for true equilibrium costs. Obviously, in reality, the last driver’s travel time cannot be far from the free-flow value, as otherwise it would become attractive to depart later from home and drive at higher speed. As the exit is always fully utilized between \( t_s \) and \( t_e \), the total number of drivers passing the bottleneck is equal to the total demand, \( N \), i.e.:

\[
c_a(s(t_e) - t_e) = N \tag{B1}
\]

Thus, the arrival time of the last driver at the destination is given by:

\[
s(t_e) = t_e + \frac{N}{c_a} = t_e + \frac{L}{v_f} \tag{B2}
\]

Furthermore, the total number of drivers departed should also be equal to the travel demand, i.e., \( D(t_e) = N \). Based on Eq. (12), the departure time of the last driver should satisfy the condition below:

\[
D(t_e) = \frac{\alpha}{\alpha + \gamma} c_a(t_e - t_s) + \frac{\beta + \gamma}{\alpha + \gamma} c_a(t_e - t_s) = N \tag{B3}
\]

Combining the above Eqs. (B2)–(B3) with Eq. (21) yields:

\[
t_s = t_e - \frac{\gamma}{\gamma + \beta} \frac{N}{c_a} \frac{L}{v_f} (1 - \frac{\alpha}{\gamma + \beta})^\epsilon \tag{B4}
\]

\[
t_e = t_e + \frac{\beta}{\gamma + \beta} \frac{N}{c_a} \frac{L}{v_f} (1 - \frac{\alpha}{\gamma + \beta})^\epsilon \tag{B5}
\]

where \( \epsilon = (1 - \sigma)/\lambda f_1 - L/v_f \) is the extra travel time above the free-flow travel time which will always be positive. The resulting equilibrium generalized travel cost is then given by:

\[
u'(t) = \frac{\alpha L}{v_f} + \frac{\gamma \beta}{\beta + \gamma} \frac{N}{c_a} \frac{L}{v_f} + \frac{\alpha \gamma}{\beta + \gamma} \epsilon \tag{B6}
\]

where the last part of this equation is the extra travel time the first driver needs above free-flow travel time multiplied by \( \frac{\alpha \gamma}{\beta + \gamma} \epsilon \). Compared to the conventional bottlenecks model, it therefore has extra equilibrium travel cost caused by the increase of the first driver’s travel time which is equal to the extra travel time cost multiplied by \( \frac{\alpha \gamma}{\beta + \gamma} \epsilon \). In other words, it implies that when the first driver’s travel time increases by \( \epsilon \), the equilibrium generalized travel cost increases by \( \frac{\alpha \gamma}{\beta + \gamma} \epsilon \). From Eqs. (B4)–(B5), we observe that the departure peak period starts earlier with \( \left(1 - \frac{\alpha}{\gamma + \beta}\right) \epsilon \).

The above results are all based on the assumption that the travel time of the last driver is free-flow travel time. However, the travel times of the last driver in the CPI model and in the FTL model are both endogenously determined by the condition that the total number of drivers arriving at the destination should be equal to the travel demand. The example in Section 5.1 shows that the last driver’s travel time is numerically very close to the free-flow travel time. This means the resulting equilibrium generalized travel cost and equilibrium peak period will also be very close to those in Eqs. (B4)–(B6). In fact, the differences turn out to be almost negligible, and therefore policy analysts can use the above closed-form results to get a quick and fairly precise estimation of the equilibrium generalized travel costs, as well as the equilibrium peak period.

Appendix C. The equilibrium generalized travel costs

To demonstrate that the proposed models converge to equilibria, the resulting generalized travel costs in equilibrium are plotted below. Figs. 9 and 10 confirm that the equilibrium generalized travel costs during the peak period are identical for all drivers departing between the entire equilibrium departure period, and would be higher outside.

References


