

Priority queue formulation of agent-based bathtub model for network trip flows in the relative space[☆]

Irene Martínez^{a,*}, Wen-Long Jin^b

^a Transport & Planning, Delft University of Technology, Stevinweg 1, Delft, 2628 CN, The Netherlands

^b Department of Civil and Environmental Engineering, University of California Irvine., 4000 Anteater Instruction and Research Bldg., Irvine, 92697, CA, United States

ARTICLE INFO

Keywords:

Agent-based bathtub model
Efficient simulation model
Priority queue
Relative space
Trip travel time distribution

ABSTRACT

Agent-based models have been extensively used to simulate the behavior of travelers in transportation systems because they allow for realistic and versatile modeling of interactions. However, traditional agent-based models suffer from high computational costs and rely on tracking physical locations, raising privacy concerns. This paper proposes an efficient formulation for the agent-based bathtub model (AB²M) in the relative space, where each agent's trajectory is represented by a time series of the remaining distance to its destination. The AB²M can be understood as a microscopic model that tracks individual trips' initiation, progression, and completion and is an exact numerical solution of the bathtub model for generic (time-dependent) trip distance distributions. The model can be solved for a deterministic set of trips with a given demand pattern (defined by the start time of each trip and its distance), or it can be used to run Monte Carlo simulations to capture the average behavior and variations of stochastic demand patterns. To enhance the computational efficiency, we introduce a priority queue formulation for AB²M, eliminating the need to update trip positions at each time step and allowing us to run large-scale scenarios with millions of individual trips in seconds. We systematically explore the scaling properties of AB²M and discuss the introduction of biases and numerical errors. Finally, we analyze the upper bound of the computational complexity of the AB²M and the benefits of the priority queue formulation and downscaling on the computational cost. The systematic exploration of scaling properties of the modeling of individual agents in the relative space with the AB²M further enhances its applicability to large-scale transportation systems and opens up opportunities for studying travel time reliability, scheduling, and mode choices.

1. Introduction

Transportation systems are defined by complex interactions between system supply and demand for goods and travelers. The heterogeneous behavior of travelers navigating dynamic environments results in complex dynamics of collective systems. To simulate this, agent-based modeling tools (originally developed for the field of computing) are suitable (Crooks et al., 2018). This approach decomposes complex systems into individual agents with specific rules like activity constraints and travel preferences (Bonabeau, 2002). Each agent is a unique, autonomous entity that adapts to changes. Agent-based models (ABM) effectively represent various

[☆] This article belongs to the Virtual Special Issue on IG005590: VSI:ISTTT25.

* Corresponding author.

E-mail addresses: I.Martinez@tudelft.nl (I. Martínez), wjin@uci.edu (W.-L. Jin).

interactions, such as between vehicles and passengers, and their adaptable nature allows for easy modification of assumptions and requirements.

In the transportation domain, ABMs have been traditionally used for microscopic traffic and macroparticles simulations, e.g., with MATSim (Balmer et al., 2008) or POLARIS (Auld et al., 2016), and have been used to study the system dynamics and to evaluate the impact of management strategies. The readers are referred to a recent review by Bastariento et al. (2023). However, ABMs also come with some challenges, including data collection and accuracy and very high computational cost (Kagho et al., 2020). To reduce the computational cost of simulations, traditional ABM relies on downscaling methodologies (Nicolai, 2012; Ben-Dor et al., 2021). For example, the downscaling approach in MATSim software relies on a method that scales the network characteristics based on the reduced number of agents modeled. Although this methodology has been used for years, there is limited systematic understanding as to whether a downsized model can reproduce the same exact results as a full-scale model. Recent studies show that downscaling “too much” might introduce bias into the results (Llorca and Moeckel, 2019; Ben-Dor et al., 2021). Furthermore, even relying on downscaling, the computational complexity of traditional ABM is still high; for example, a MATSim simulation for the city of Paris (with 10% of the population) takes 5 hours to run on a modern cluster (Hörl et al., 2019).

Traditionally, ABMs (as well as other traditional traffic flow simulation software) require the generation of a network with individual links that compose the physical configuration. We refer to this traditional perspective as the absolute space. This approach involves significant computational efforts, both for calibration and simulation. Moreover, using ABM to track the movement of people in the absolute space leads to personally identifiable location information being traced, which can lead to privacy concerns. In contrast, one can take the relative space approach, where the network is an undifferentiated unit, and disregard the physical locations of vehicle-trips inside the network. In this relative space dimension, the trip flow dynamics are modeled based on their remaining trip distances to their respective destinations, preserving personally identifiable location information.

The bathtub model (a.k.a. reservoir model) is an aggregated model used to describe the trip flow dynamics in the relative space and has been garnering interest among the transportation research community (Vickrey, 2020; Small and Chu, 2003; Arnott, 2013; Johari et al., 2021). It can be viewed as a network queuing system, and it relies on three main assumptions¹: (A1) the network is treated as an undifferentiated unit, where links and individual trips’ origins, destinations, and routes are implicit; (A2) the demand is described by the trip distance and the trip initiation rate; and (A3) there exists a network-level speed-density relation, which is nowadays commonly referred to as network fundamental diagram (NFD) or (speed) macroscopic fundamental diagram (MFD). This network-wide relation between speed and density was proposed and calibrated by Godfrey (1969). Later, the NFD was also studied by Mahmassani et al. (1987) and Daganzo (2007) and gained significant interest in the research community. The existence of such relation at the network level with a large-scale loop detector data collection was first verified in downtown Yokohama by Geroliminis and Daganzo (2008) and, since then, NFDs have been experimentally calibrated for multiple cities (Johari et al., 2021). These models, applicable to various transportation modes, essentially function as network queues in a relative space dimension. Note that bathtub models can also be viewed as compartmental models (Jin et al., 2021), where all the trips in the same network are lumped together in a bathtub.

Most of the bathtub models in the literature assume continuum travelers, similar to water in a bathtub, and are thus called *continuum* bathtub models. In the last decade, there has been interest in using agent-based approaches to model trip flow dynamics at the network level (Arnott, 2013; Daganzo and Lehe, 2015; Mariotte et al., 2017; Lamotte et al., 2018). Mariotte et al. (2017) introduced the so-called trip-based model (TBM) to numerically track individual trips. However, existing formulations do not fully take advantage of the characteristics of the relative space for an efficient algorithm and often consider some restricting assumptions about the demand, e.g., time-independent trip distances and/or deterministic demand. Consequently, there is a need to develop an efficient model in the relative space that accurately and efficiently captures the individual agents’ trajectories for any demand assumption.

In this paper, we propose a highly efficient (simulation) model to capture the agent’s dynamics in the relative space. We refer to the proposed model as *agent-based* bathtub model (AB²M) for two reasons. First, to highlight that the travelers are people with decision capabilities and can be modeled as independent agents with corresponding socio-economic characteristics. A single agent could perform multiple trips (trip-chain) or use multiple modes in a single trip. Second, to highlight that the model is based on the same three assumptions of relative space as the bathtub model. Unlike continuum bathtub models that rely on aggregated variables, the agent-based formulation provides detailed information on individuals, including travel times. This enables the introduction of heterogeneity and studying higher order moments such as travel time variation to assess travel time reliability and investigate scheduling (Noland and Polak, 2002) and route choices (Lam and Small, 2001). The proposed AB²M is an exact numerical solution that extends the TBM (Arnott, 2013; Fosgerau, 2015; Mariotte et al., 2017) by accommodating agents with heterogeneous trip distances and time-dependent trip distance distributions. Moreover, the AB²M can model stochastic demands through Monte Carlo simulations. To enhance the efficiency of AB²M, we propose a priority queue formulation, which eliminates the need to update the position of trips at each time step, resulting in significant computational performance improvements. For instance, simulations involving one million individual trips over a one-hour period can be processed in less than two seconds. Further, we explore its scaling properties systematically, similar to the downscaling strategies used in traditional ABM models.

The rest of the paper is structured as follows. Section 2 reviews existing models in the literature and proposes the AB²M, which tracks the trajectories of agents in the relative space. Then, Section 3 presents a new formulation that is computationally more efficient. Section 4 discusses the scalability property of the bathtub model, and Section 5 reviews the numerical complexity of the proposed models. Then, some numerical results are presented in Section 6. Finally, Section 7 concludes the paper and discusses future directions. For the readers’ convenience Table 1 presents a list of variables.

¹ Read the foreword of Vickrey (2020) for details.

Table 1
List of notations.

Variable	Description
$\tilde{D}(t)$	Average trip distance of entering trips at time t
$E(t)$	Cumulative number of starting trips at time t
$G(t)$	Cumulative number of finished trips at time t
I	Total number of trips
$K(t, x)$	Number of active trips at time t with remaining distance not smaller than x
L_N	Total network length
$N(t, i)$	Position (order) of trip i at time t regarding its characteristic trip distance.
$T(i)$	Trip i start time on simulation
$\hat{T}(i)$	Time when trip i exits the network
$V(\rho)$	Network level speed-density relation
$X(i)$	Trip distance of trip i
$e(t)$	Trip initiation rate at time t
$g(t)$	Completion rate of trips at time t
i	Trip ID, which refers to the trip position regarding its initiation time
$m(t)$	Total remaining distance to be traveled by active trips
t_f	Total simulation time
u_f	Free-flow speed
$v(t)$	Average travel speed in the system at time t
w	Shock wave speed
$x(t, i)$	Remaining distance of trip i at time t
$y(t, n)$	Remaining distance of n -shortest trip at time t (ordered remaining trip distance)
$z(t)$	Characteristic network traveled distance
Δt	Time interval or time step of the simulation. Note that it is not necessarily fixed.
Δv	Speed variation in a time-step considered.
$\Theta(t, n)$	Ordered characteristic trip distance of active trips at time t
$\delta(t)$	Number of active trips at time t
$\gamma(i)$	Travel time of trip i .
$\rho(t)$	Per-lane density in the system at time t
ρ_j	Per-lane jam density
$\theta(i)$	Characteristic trip distance of trip i
$\varphi(t, x)$	Distribution of remaining trip distances x at time t for active trips.
$\tilde{\varphi}(t, x)$	Joint distribution of trip distance, x , and departure times, t

2. Model formulation

2.1. Demand definition in the relative space

In the relative space, the demand is defined with the joint distribution of trip distance and departure time (Jin, 2020), which is described by $\tilde{\varphi}(t, x)$ and is defined for $x \geq 0$ and $t \geq 0$. This joint distribution can be continuous or discrete; and deterministic or probabilistic, both in time t and trip distance x . Therefore, there are eight possible types of demand in the relative space. If the demand is stochastic, the joint distribution $\tilde{\varphi}(t, x)$ becomes a probabilistic function. By definition of joint probability, the joint distribution can be written as the probability of departure time multiplied by the conditional distribution of trip distances, $\tilde{\varphi}(t, x) = \tilde{\varphi}(t) \cdot \tilde{\varphi}(x|t)$ where $\tilde{\varphi}(t)$ is the marginal probability density (or mass) function (Martínez and Jin, 2021). The conditional distribution $\tilde{\varphi}(x|t)$ is the most generic expression for capturing trip distance distributions as a function of time. If the trip distance distribution is a continuous variable, the trip initiation rate is defined by

$$e(t) = I \int_0^{\infty} \tilde{\varphi}(t, x) dx, \quad (1)$$

where I is the total number of trips in the study period. The completion rate of trips, $g(t)$, is determined by the interaction of supply and demand. The supply is defined by the fundamental relation of speed and density, $V(\rho)$, at the network level, i.e., the NFD. Several modeling formulations have been proposed in the literature and described in the next subsection.

2.2. Review of existing models

The existence of a uni-modal, low-scatter speed density relationship between the speed and average density of the region (Godfrey, 1969; Geroliminis and Daganzo, 2008; Johari et al., 2021), has allowed the development of continuum bathtub models. These aggregated models capture the network traffic dynamics in an urban region by modeling the inflow and outflow of trips in a system (or region).

The continuum bathtub models generally describe the rate of change of active trips $\delta(t)$. The first derivation was built on a particular case of assumption A2, i.e., that trip distances follow a time-independent negative exponential (NE) distribution, leading to an ordinary differential equation (ODE) (Vickrey, 1991, 2020). We will refer to this version as Vickrey's bathtub model (VBM). Similar ODEs were derived independently in the early 2000s by other authors (Small and Chu, 2003; Daganzo, 2007) but without

Table 2
Characteristics of demand assumption in the relative space based on model formulations.

Characteristic	BBM	VBM/ PL	GBM	TBM	M-model	AB ² M
Travelers	C	C	C	D	C	D
Time-dependent TDD	No	No	Yes	(Generally) No	No	Yes
Trip distance	D	C	D or C	D or C	D or C	D or C

TDD: Trip distance distribution, C: Continuous, D: Discrete.

building on the assumption of NE trip distance distribution. For example, [Daganzo \(2007\)](#) assumed uniform average trip lengths, i.e., “the average trip length is the same for all origins, d ”. This ODE formulation became widely accepted in the literature ([Johari et al., 2021](#)), and is also known as accumulation-based model ([Mariotte et al., 2017](#)) or PL model ([Sirmatel et al., 2021](#)). In the 2010s’ researchers proposed the basic bathtub model (BBM), which assumes that all travelers have the same (homogeneous) trip distance ([Arnott, 2013](#); [Arnott et al., 2016](#); [Arnott and Buli, 2018](#)). Several authors noted that using the VBM to model the dynamics of a system with homogeneous trip distances would lead to inconsistencies, such as an instantaneous increase in the completion rate of trips with an increased inflow to the system ([Arnott et al., 1993](#); [Mariotte et al., 2017](#); [Johari et al., 2021](#)).

Later developments include the generalized bathtub model (GBM) by [Jin \(2020\)](#), which introduced the (relative) spatial dimension explicitly and used partial differential equations (PDEs) to model more generic demand patterns through

$$\frac{\partial K(t, x)}{\partial t} - V \left(\frac{K(t, 0)}{L_N} \right) \frac{\partial K(t, x)}{\partial x} = e(t) \tilde{\Phi}(t, x), \quad (2)$$

where $K(t, x)$ is the number of active trips at time t with remaining distance not smaller than x , $K(t, 0) = \delta(t)$, and $\tilde{\Phi}(t, x) = \int_x^\infty \tilde{\varphi}(t, y) dy$ is the cumulative density function of entering trips with distances not smaller than x . The M-model, proposed by [Sirmatel et al. \(2021\)](#) and based on [Lamotte et al. \(2018\)](#), uses the conservation of trip distance, i.e.,

$$\dot{m}(t) = e(t) \tilde{D}(t) - \delta(t) V \left(\frac{\delta(t)}{L_N} \right), \quad (3)$$

and propose to use the total remaining distance to be traveled $m(t)$ to approximate the completion rate of trips outside steady-state conditions. Note that the M-model during steady states is simplified to VBM.

Instead of relying on continuum models, an alternative is to use agents (discrete travelers) to model the trip flow dynamics. [Daganzo and Lehe \(2015\)](#) were the first to do so. They indirectly considered the relative space, claiming that “since there are no routes in the macroscopic theory [...] a network governed by an MFD [...] may be modeled as a simple, aspatial queuing system [...]”. In other words, they abstracted the network into a queuing system and explicitly disregarded the (traditional) absolute space dimension. They outlined the idea that the trip dynamics can be modeled as a non- first-in-first-out (non-FIFO), multi-channel queuing system, where the trip distance is the customers’ workloads, and the servers process the customers at a common rate (the speed determined by the NFD). They considered the departure time and trip distance, which they assumed follow a uniform distribution, to be input to the simulation. Another well-established formulation is the TBM by [Mariotte et al. \(2017\)](#), where they proposed two numerical resolutions. First, a formulation where the trips are also considered a continuum variable and the cumulative curves of trip initiation and trip completion are calculated. To do so, they estimate the time difference between the exit time of vehicle n and vehicle $n + \Delta n$. They proposed a resolution of the method at higher orders and concluded that a first-order approximation is accurate enough. Second, an event-based formulation inspired by an unpublished work by [Lamotte and Geroliminis \(2016\)](#), where the progression of each trip towards the destination is updated every event, i.e., when a vehicle enters or exists the system. The algorithm proposed updates the traveled distance for all the circulating vehicles during each event. Later, [Lamotte et al. \(2018\)](#) proposed a complete agent-based simulation, and to limit the computational complexity they did not consider that each agent may have a different trip distance but considered 1000 deterministic trip distances associated each to 1000 batches of agents. [Lamotte et al. \(2018\)](#) also consider multiple trip distance distributions, including uniform and mixtures of uniform distributions by varying the standard deviation and all their trip distance distributions had the same (time-independent) average trip distance. On top of the discrete formulation, [Lamotte et al. \(2018\)](#) presented a continuum function to calculate the trip completion rate (a.k.a. outflow function) to capture more generic time-independent trip distance distributions. This continuum formulation has also been referred to as TBM ([Sirmatel et al., 2021](#)). A more general formulation the trip completion rate was recently proposed by [Laval \(2023\)](#), where the trip distance distribution is explicitly considered time dependent. Note that all the above formulations disregard the space dimension ([Johari et al., 2021](#)) with the exception of the GBM ([Jin, 2020](#)).

The reviewed models categorize travelers as either continuous or discrete entities and have different assumptions about the demand, characterized by the trip distance distribution. This trip distance can be either time-dependent or independent, as well as continuous or discrete. These assumptions are summarized in [Table 2](#).

Some researchers have concluded that the VBM, GBM, and TBM formulations are equivalent when the trip distance distributions follow a time-independent NE distribution or under steady states ([Lamotte et al., 2018](#); [Jin, 2020](#); [Laval, 2023](#)). However, the definition of steady (or stationary) state is not consistent in the literature. It can be generally interpreted as vehicle inflow equals vehicle outflow, i.e., $\dot{\delta}(t) = 0$. [Laval \(2023\)](#) defines the steady state as “[...] the circulating flow over the total network distance, has to match the incoming production [of trip distance]”, i.e., $\dot{m}(t) = 0$, while ([Jin, 2020](#)) defines stationary states when the number of active trips with remaining distance not smaller than x is time independent for any x .

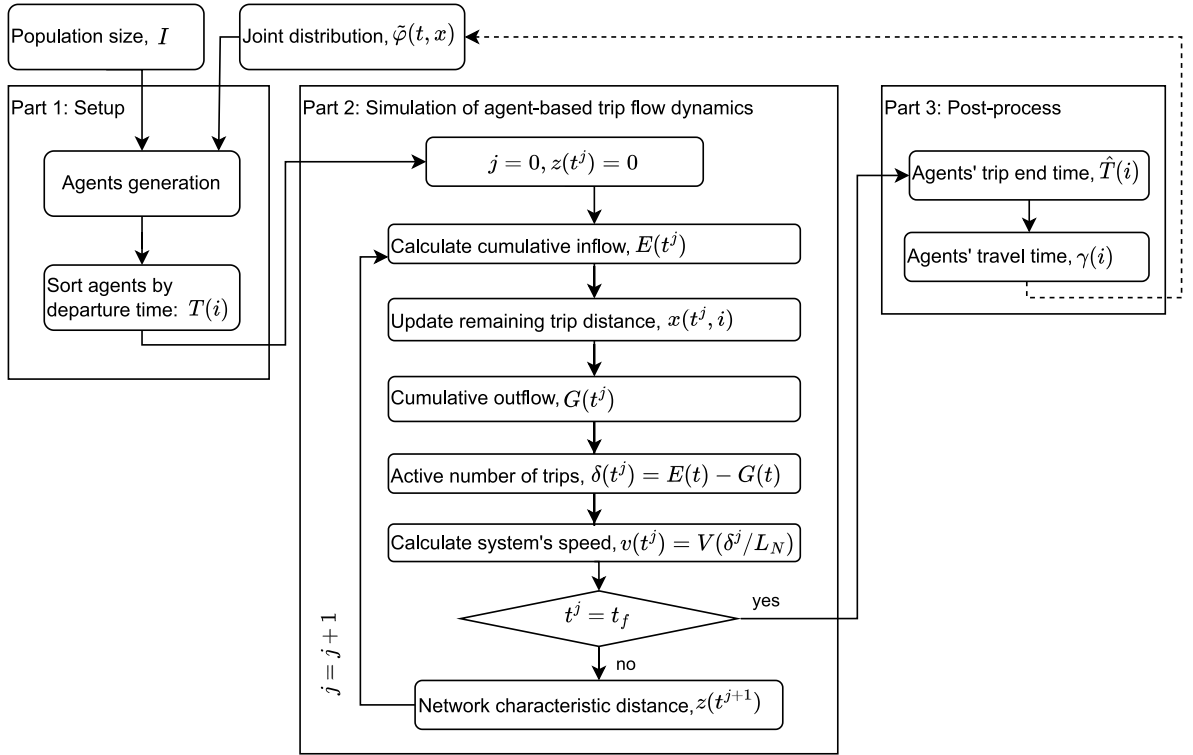


Fig. 1. Overview of the main steps of the AB²M.

Many of these models have been used to study multiple transportation research questions. For example, to study the departure time problem different researchers have used the TBM (Lamotte and Geroliminis, 2018), the BBM (Arnott and Buli, 2018) or the GBM (Ameli et al., 2022). Moreover, the VBM has been extensively used to propose management strategies at the network level, such as gating strategies (Keyvan-Ekbatani et al., 2012) or pricing (Zheng et al., 2012; Zheng and Geroliminis, 2016).

2.3. Naive formulation of AB²M

Most of the models reviewed in Section 2.2 consider continuum travelers, and we refer to them as *continuum bathtub models*. Instead, in this paper, we are interested in proposing a new formulation to efficiently model the agent-based version of the bathtub model, i.e., considering discrete travelers. To the best of our knowledge, only three of the aforementioned studies proposed such a discrete formulation and study the model characteristics (Daganzo and Lehe, 2015; Mariotte et al., 2017; Lamotte et al., 2018). Daganzo and Lehe (2015) proposed a discrete time step simulation and both Mariotte et al. (2017) and Lamotte et al. (2018) proposed event-based simulations. In all three studies, the authors track the commuters' traveled distance and the trips completed when the traveled distance equals or exceeds their trip distance. In this section, we will present a formalized agent-based bathtub algorithm with a fixed time step that is conceptually equivalent to existing models in the literature.

First, we introduce the characteristic network traveled distance as

$$z(t) = \int_0^t v(s) ds, \quad (4)$$

which as defined by Lamotte et al. (2018) as “cumulative distance traveled” and by Jin (2020) as the “characteristic travel distance”. This characteristic network traveled distance is useful to simplify the notation of the remaining trip distance. From the agent-based perspective, one can interpret the characteristic network traveled distance as the cumulative traveled distance by a reference vehicle trip that initiates its trip at $t = 0$ and never exits the system (Lamotte et al., 2018).

In the AB²M, the number of active agents in the system are not tracked by differential or integral equations as presented in Section 2.2. Instead, the trip progression of each individual agent i (towards their destination) is tracked by reducing her remaining trip distance over time. Each vehicle trip represents an agent moving in the relative space, and the coordinates will be Lagrangian coordinates. Conceptually, the AB²M is similar to a traditional microscopic simulation model, but (i) the trajectory is in the relative space, and the personally identifiable location information is preserved, and (ii) the speed is determined by a global speed-density relation (NFD) instead of a local speed-density relation.

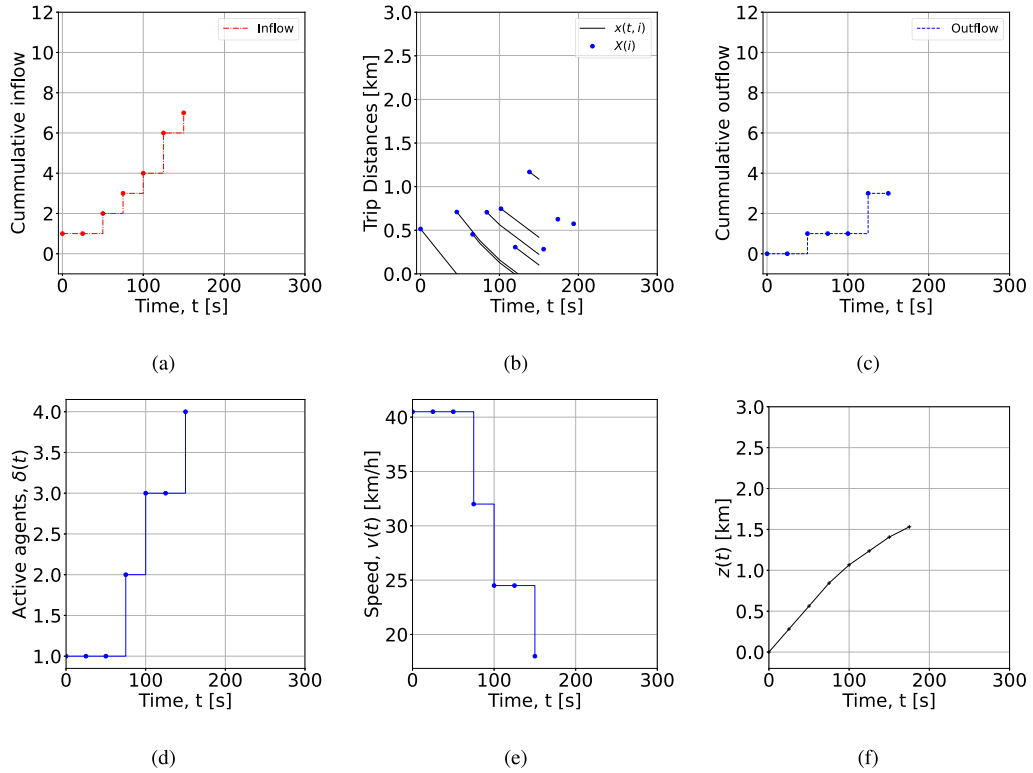


Fig. 2. Example with 10 agents in 200 s, NFD as $V(\rho) = 50(1 - \rho/10)^2$, $L_N = 1$ km, $\Delta t = 25$ s, simulation stopped at $t = 150$ s.

The AB²M consists of three main parts (see Fig. 1): Part 1 is the setup, where the agents are generated with their characteristics; Part 2 is the simulation process updating the speed of the system and the positions of the agents (i.e., the discrete trajectories of the agents in the $x - t$ plane, see Fig. 2(b)); and Part 3 is the post-process, used to determine the travel times of the agents. As in any dynamic system, the AB²M needs boundary conditions and initial conditions. The boundary conditions are characterized by a sample of I agents, that is generated by the joint distribution. The travel time experience of each agent may influence the demand, indicated with a dashed arrow in Fig. 1 and should be studied as departure time choice (Lamotte and Geroliminis, 2018; Ameli et al., 2022). However, the influence of the travel time on the demand (e.g., departure time choice) is out of the scope of this paper, and we assume that the population size and joint distribution are exogenous and given.

For Part 1, we consider simple agents that only have two characteristics: the trip distance, $X(i)$, and the trip start time, $T(i)$. Thus, the model demand input can be a joint trip distance and departure time distribution, $\tilde{\varphi}(t, x)$, (Jin, 2020; Martínez and Jin, 2021). In the case that this joint distribution is a probability distribution, the demand of the AB²M can be obtained as a sample of I agents from the joint probability distribution. In the AB²M, the demand can be defined by agents (dots) in the $X - T$ plane. In the future, the socio-economic characteristics or trip purpose could be added to the agents' characteristics to endogenously capture the departure time and the mode choice, for example. In Part 1, the agents are sorted by departure time as a fixed priority queue and are given an index $i = 1, \dots, I$, which can be interpreted as the trip ID, based on their start time, i.e., $T(i) \leq T(i + 1)$.

In Part 2, the actual trip flow dynamics are modeled. If the study period starts at $t = 0$, w.l.o.g., we will assume that no agents have entered the system before that time, i.e., $E(t < 0), z(0) = 0$. For all time steps considered ($j = 1, \dots, J$) the step size is fixed, i.e., $\Delta t^j = \Delta t$ and we divide the total simulation period t_f into J intervals of $\Delta t = \frac{t_f}{J}$. The positions of the agents and the speed of the system are updated for each time step following the six steps as shown in Fig. 1, and presented in Fig. 2 for a numerical example. The AB²M can be easily extended to be an event-based simulation.

First, $E(t)$ is obtained from calculating the number of agents that have already started $T(i) < t$. A new discrete variable $x(t, i)$ stores the remaining trip distance to the destination for i 's agent, which is updated. For each new trip $x(t, i) = X(i)$, and for agents that were already in the system their remaining trip distance is $x(t, i) = X(i) - \int_{T(i)}^t v(\tau) d\tau$. From Eq. (4) we have

$$x(t, i) = X(i) + z(T(i)) - z(t). \quad (5)$$

The trip i is completed when $x(t, i) = 0$. Then, the total number of trips completed is

$$G(t) = \sum_{i=1}^{E(t)} \beta_{x(t,i)}, \quad (6a)$$

where $\beta_{x(t,i)}$ is defined as

$$\beta_{x(t,i)} = \begin{cases} 1 & \text{if } x(t,i) \leq 0 \\ 0 & \text{if } x(t,i) > 0. \end{cases} \quad (6b)$$

By conservation of vehicles, we have $\delta(t) = E(t) - G(t)$, and the speed is obtained from the NFD, i.e., $v(t) = V(\rho(t))$, where $\rho(t) = \frac{\delta(t)}{L_N}$ is the density. Finally, one updates the characteristic network traveled distance as

$$z(t + \Delta t) = z(t) + v(t)\Delta t.$$

In Part 3 (the post-process), the completion time of the agents ($\hat{T}(i)$) is calculated by solving

$$X(i) = z(\hat{T}(i)) - z(T(i)). \quad (7)$$

Then, the travel time of each user is determined as

$$\gamma(i) = \hat{T}(i) - T(i).$$

The native formulation presented does not take advantage of the fact that once a trip has been completed, we do not need to store its information anymore. A way to reduce the computational cost of the summation in Eq. (6), would be to use a new variable $\hat{x}(t, j)$, where only the active trips are sorted. Thus, the size of the variable is time-dependent, i.e., $E(t) - G(t)$. However, this is still computationally expensive because all the remaining trip distances need to be updated at each time step.

The AB²M can capture the dynamics in an exact way for a deterministic set of trips by running the simulation once. Further, it can be used to study a probabilistic joint distribution, $\tilde{\varphi}(t, x)$, including the assumption of a stochastic exogenous trip initiation rate and/or stochastic trip distance distribution. In that case, the expected behavior of the system and the standard deviation of the main variables should be studied through Monte Carlo simulations. This allows for accounting for and studying the variability of those variables, such as travel time. To perform Monte Carlo simulations, it is important to ensure a low computational complexity of the AB²M. In the following, we will discuss how the AB²M can become computationally efficient by treating the active agents as a priority queue (Section 3) and through its scalability property (Section 4).

3. Priority queue formulation of AB²M

Updating and tracking the remaining trip distance of circulating trips requires high memory storage and computational cost. In response, we propose a new formulation of the AB²M, designed for greater efficiency. We define the sorted remaining trip distance as $y(t, n)$, where $n = \{1, \dots, E(t)\}$ represents the trips ordered by their remaining trip distance at time t . We refer to $n = N(t, i)$ to the position of trip i at time t in this ordered list. Note that $y(t, n)$ increases with n and is only guaranteed to decrease in t for $n \leq G(t)$, since active trips could change position in the ordered list when new trips are added to the system. Then, Eq. (6a) is adapted to count only the active trips

$$g(t) = \sum_{n=G(t)}^{E(t)} \hat{\beta}_{y(t,n)},$$

where $\hat{\beta}_{y(t,i)}$ is defined as

$$\hat{\beta}_{y(t,i)} = \begin{cases} 1 & \text{if } y(t, n) \leq 0 \\ 0 & \text{if } y(t, n) > 0. \end{cases}$$

However, FIFO does not hold for bathtub models in general.² Thus, sorting the whole set of agents every time step becomes computationally very expensive. In the following, we show how the “shorter-(characteristic)-distance-first-out” (SCDFO) principle (Jin, 2020) enables an efficient sorting of trips.

3.1. Shorter-characteristic-distance-first-out principle

We extend the concept of the characteristic distance proposed by Jin (2020) for individual agents explicitly, and we define the “characteristic trip distance” for each agent i as

$$\theta(i) = X(i) + z(T(i)), \quad (8)$$

which is a time-independent characteristic of the agent. From Eqs. (5) and (8), the remaining trip distance at each time instant $t > T(i)$, is related to the characteristic trip distance as $x(t, i) = \theta(i) - z(t)$. Although FIFO does not hold for bathtub models in general, trips with shorter $\theta(i)$ will exit the network earlier, i.e., trip i will be completed earlier than trip j if $\theta(i) < \theta(j)$, which is referred to as the SCDFO principle (Jin, 2020). In other words, ordering trips based on their remaining trip distance is equivalent to ordering trips by their characteristic trip distance, $\theta(i)$. Using this property, which is the counterpart of FIFO in the absolute space, one is not required to update the remaining trip distance at each time step. Instead, the characteristic trip distance is calculated

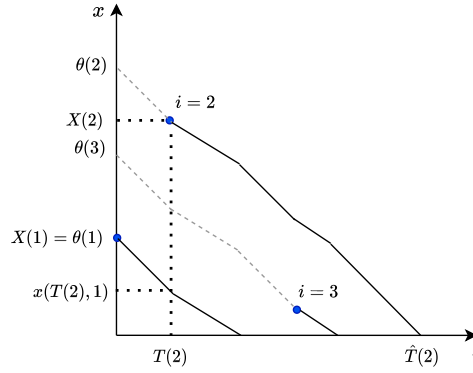


Fig. 3. Representation of characteristic trip distance $\theta(i)$ from the trajectories in relative space.

only once (when the agent enters the network) by tracing back in the relative space, see the dashed lines in Fig. 3, where the solid lines in Fig. 3 represent the trajectories of the agents in the relative space.

As mentioned earlier, the AB²M formulation is equivalent to a microscopic model (in Lagrangian coordinates) of the trip flow dynamics in the relative space. However, the vertical distance between the trajectories has no physical meaning. Because the speed of all agents is the same, the trajectories will not cross. However, two agents can have overlapping trajectories if they have the same $\theta(i) = \theta(j)$ even if they start at different times $T(i) \neq T(j)$ and have different trip distances $X(i) \neq X(j)$.

The advantage of SCDFO is that the agents can be sorted by increasing $\theta(i)$ to facilitate the calculation of $G(t)$. Moreover, from Eqs. (7) and (8) the post-process can be simplified to solve $\theta(i) = z(\hat{T}(i))$.

3.2. Priority queue definition and efficient algorithm

To take advantage of the SCDFO principle, we propose a new main variable, which is an ordered list of characteristic trip distances of the circulating agents $\Theta(t, n)$. This vector satisfies $\Theta(t, N(t, i)) = \theta(i)$, where $N(t, i)$ is the position in this ordered list of the trip i at time t . Thus, $\Theta(t, N(t, i) = 1)$ is the active trip at time t with the shortest characteristic trip distance. In contrast, $\Theta(t, \delta(t))$ is the active trip with the longest characteristic trip distance.

In this efficient formulation, Steps 2 and 3 of Part 2 (Fig. 1) are modified. Instead of updating the agents' remaining trip distance, we update $\Theta(t, n)$ in Step 2. Since the characteristic trip distance is a time-invariant feature of each agent, the modeler only needs to sort the new arriving trips into an already existing ordered list, which represents a significant computational complexity reduction. Then, the completion of trips, $g(t)$, corresponds to the sum of all agents with $\Theta(t, n) \leq z(t)$, i.e.,

$$g(t) = \max_n \{\Theta(t, n) \leq z(t)\}. \quad (9)$$

The position of trip i in the ordered list by characteristic trip distance may change with time, since trips with start time $T(j) > t$ might have shorter characteristic trip distance than existing trips, i.e., $\theta(j) < \Theta(t, \delta(t))$. Thus, $N(t, i)$ is time-dependent for a given active trip i .

By definition of an ordered list, $\Theta(t, n)$ is monotonically non-decreasing with n . Moreover, $\Theta(t, n)$ is also monotonically increasing in time for a given n , since the characteristic trip distance on a given position in the priority queue is only replaced by a longer one. The discrete interpretation is presented in Fig. 4, which shows the agents that initiate their trip between t and $t + \Delta t$ in blue, and the agents that have completed their trip outlined.

In the following, we discuss the details of the fixed time-step leveraging the priority queue formulation, i.e., Algorithm 1. Since $\Theta(t, n)$ is a priority queue, we propose to use an ordered list using min-heaps to store $\Theta(t, n)$ in the algorithm. Heaps are a type of data structure, and a min-heap is a binary tree where the root parent node is smaller than the children. Heaps are usually implemented to handle priority queues and are common in shortest path algorithms (Johnson, 1975). Min-heaps also allow an easy restructure when new elements are added or when certain elements are removed from the list. Min-heaps are defined by two characteristics:

² For the BBM, i.e., when all trips have the same distance, FIFO holds for the bathtub models.

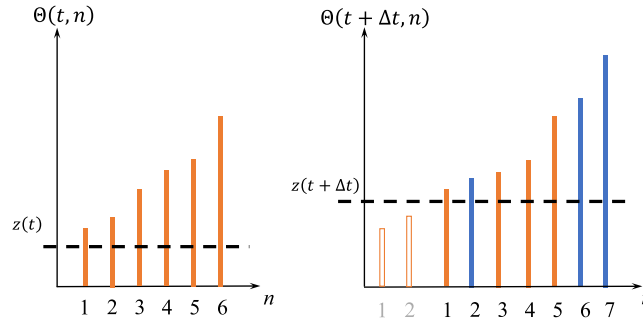


Fig. 4. Evolution of $\Theta(t, n)$. When the characteristic trip distance of the new trips is the largest of all the current circulating agents, it is added at the end of the queue. If a new trip $\theta(i)$ is not the longest, it is added between active trips.

(i) only elements on the top (i.e., shorter characteristic trip distance) are removed, and (ii) new trips can be inserted in any place since they can have a shorter or longer effective distance than active trips in the network.

Algorithm 1: Pseudo-code of the proposed algorithm with fixed time step.

```

% Initialization of variables
t = 0, E(t) = 0, z(t) = 0;
% Simulation
for t = 0 to t = t_f do
    Obtain E(t) from T(i);
    if E(t) > E(t - Δt) then
        for j = 1 to j = E(t) - E(t - Δt) do
            z(T(j)) = z(t) - v(t - Δt)(t - T(j)); % Characteristic network travel distance at trip start time
            θ(E(t) + j) = X(E(t) + j) + z(T(j)); % Eq. (8)
            Insert θ(E(t) + j) in an ordered way in heap Θ(n); % Cost O(log n)
        end
    end
    g(t) = 0;
    while Θ(n = 1) ≤ z(t) do
        Drop Θ(n = 1) from the heap; % Cost O(log n)
        g(t) = g(t) + 1;
    end
    G(t) = G(t - Δt) + g(t); % Calculate cumulative outflow
    v(t) = V((E(t) - G(t))/L_N); % Update of speed
    z(t + Δt) = z(t) + Δt v(t);
    t = t + Δt
end

```

In Algorithm 1, in each time step the cumulative inflow is calculated. If $E(t) > E(t - \Delta t)$, new trips are entering the system, and their characteristic trip distance should be calculated from Eq. (8). Then, the entering trips' $\theta(i)$ are introduced into $\Theta(t, n)$. The completion rate of trips Eq. (9) can be determined by comparing the first element of $\Theta(t, n)$ with the characteristic network traveled distance. If $\Theta(t, 1) \leq z(t)$, the trip is removed and $\Theta(t, n)$ is updated. This process is iterated until all active trips have a characteristic trip distance larger than $z(t)$, i.e., $\Theta(t, 1) > z(t)$. This is only efficient if there are very few active trips. Otherwise, it would be worth to use a bisection method to find the $\Theta(t, n)$ closest to $z(t)$.

4. Scalability property of the bathtub models

A promising direction to make AB²M more efficient is to explore the concept of “downscaling” already used to reduce the computational cost in traditional ABM in the absolute space (Nicolai, 2012). There are two approaches to downscaling: (i) use so-called “super-agents” (SA) that represent several agents, although it requires a re-programming of the agents and their behaviors, and (ii) the so-called “one-represents-several” (ORS) methods, where the network characteristics (such as capacity and jam density) are reduced based on scaling ratios (usually proportional to the reduction of the number of agents). The downscaling of agents in MATSim has been widely implemented, using the ORS method by downscaling linearly the capacity of the links as $C_r = Cr$ and nonlinearly the jam density of the links as $\rho_{j,r} = \rho_j r^{0.75}$, see details in Nicolai (2012). The scaling factor, r , ranges from 1% in Munich; see Kickhöfer and Kern (2015) to 25% in Dublin; see Mcardle et al. (2014). Despite the long-term implementation of this approach, there are recent studies that conclude that such downscaling might introduce bias into the results (Llorca and Moeckel, 2019; Ben-Dor et al., 2021). Llorca and Moeckel (2019) performed simulations on the Munich network and showed how

the average travel time is highly dependent on the scaling factor. Further, Ben-Dor et al. (2021) analyzed the effect of downscaling on the accuracy of the results at a simple network (Sioux Falls). Both studies concluded that downscaling below $r = 0.1$ might introduce bias into the results. Moreover, there is a physical limit for downscaling ABM in absolute space. Since the characteristics of the individual links are scaled, the links cannot be reduced to fit less than one vehicle. The theoretical lowest jam density that can be considered is $\rho_{j,r} = 1 \text{ veh/km/lane}$, which limits the scaling factor r .

The downscaling for bathtub models is a natural idea that has been used by several researchers but not systematically discussed in the literature. For example, Lamotte et al. (2018) propose to use weights for trips with different characteristics and then calculate the speed as $v = V \left(\frac{\sum_i \omega_i \delta_i}{L_N} \right)$. Note that their approach does not involve downscaling the system (supply), and the network lane miles are not modified.

This section fills this gap by methodically studying the scalability (including both downscaling and up-scaling) in the relative space. First, identify two types of scaling strategies for bathtub models in 4.1 and later discuss the numerical errors that can be introduced by downscaling in Section 4.2.

4.1. Scaling of bathtub models

When considering the GBM Eq. (2), we can observe that the dynamic system can be scaled by r as

$$\frac{\partial K(t, x) \cdot r}{\partial t} - V \left(\frac{K(t, 0) \cdot r}{L_N \cdot r} \right) \frac{\partial K(t, x) \cdot r}{\partial x} = e(t) \tilde{\Phi}(t, x) \cdot r,$$

where r is the scaling factor. If $r < 1$, the system is down-scaled, and if $r > 1$, the system is up-scaled. In the scaled system, the time and speed are invariant. Since the speed depends on the density through the NFD, the density in the scaled system should be invariant too. Therefore, a good way to look at the GBM to assess its scalability is to look at the density evolution, which in turn defines the speed evolution. From Eq. (1) the GBM (Jin, 2020) can be written as

$$\dot{\rho}(t) = \frac{I}{L_N} \int_0^\infty \tilde{\varphi}(t, x) dx - \rho(t) \varphi(t, 0) V(\rho(t)), \quad (10)$$

where $\varphi(t, x)$ is the distribution of remaining trip distances x at time t . In particular, $\varphi(t, 0)$ represents the proportion of active trips that are completed at time t since their remaining distance $x(t, i) = 0$.

A first and natural way to downscaling is to consider that the scaling factor r is used to modify the trip initiation rate $e(t)$, the length of the network, L_N , and the number of active trips with remaining distance not smaller than x , i.e., $K(t, x)$. We refer to this strategy as flow-based scaling. Clearly, the traffic dynamics described by Eq. (10) on multiple neighborhoods or cities will be the same as long these have the same NFD, $\tilde{\varphi}(t, x)$, and ratio $\frac{I}{L_N}$. Therefore, the flow-based scaling modifies the supply and demand without introducing biases. On one hand, the demand is scaled through the number of agents rI , without changing $\tilde{\varphi}(t, x)$. On the other hand, the supply is scaled by changing the total network lane distance, i.e., rL_N , without changing the parameters of the NFD. Note that in this flow-based scaling, the trip distances and the characteristic trip distances are not modified. Therefore, the comparison of $z(t)$ with $\Theta(n)$ is consistent in the scaled system, which ensures that $\delta(t)$ is invariant to the scaling. This, in turn, guarantees that the speed and $z(t)$ are also invariant to the scaling.

Since the demand is defined by total vehicle-distance units, it may be interesting to explore another (not so natural) type of scaling, by modifying the distance traveled instead of the number of vehicles. In this case, the supply is scaled in the same way, i.e., modifying the network length, $L_{N,r} = rL_N$. However, the demand is scaled, maintaining the number of agents and modifying the average distance of those agents, as $\tilde{D}_r(t) = r\tilde{D}(t)$. This will be referred as distance-based scaling, and may introduce some bias because the comparison of $z(t)$ to the characteristic trip distances (which are scaled as $r\theta$) will lead to different outflows and, subsequently, different speeds. The bias introduced by distance-based scaling cannot be mathematically quantified but will be explored through numerical simulations in Section 6.2.

Although the impact of distance-based scaling cannot be neatly represented in Eq. (10), it can be shown that during steady states,³ the systems are equivalent, with both flow-based and distance-based scalings. From Eq. (3) and $\dot{m}(t) = 0$ we have

$$\frac{e(t)\tilde{D}(t)}{L_N} = \rho(t)V \left(\frac{\delta(t)}{L_N} \right).$$

Whether the same steady state is stable, reachable, etc. for different demand patterns is a relevant research question out of the scope of this paper. In Section 6.2, the numerical results will show that large values of r modify the trip flow dynamics in a way where the steady state is not reached.

As discussed earlier, there is a physical limit to downscaling the model in the absolute space. In contrast, in relative space, there is no such limit, and one can simulate 1.25 million agents on a network of $L_N = 1000 \text{ km}$ (computational cost 2.2 s) with 1250 agents on a network of $L_{N,r} = 1 \text{ km}$ (0.12 s), which represents a downscaling of 0.1% in the number of agents and a computational cost reduction to ~5%. A detailed discussion on the computational complexity of the algorithm and the benefits of downscaling will be presented in Section 5. However, the scaling in the relative space demand should be done carefully, since there is a risk that downscaling the network too much introduces numerical errors, as will be discussed in Section 4.2.

³ In this case we refer to the definition by Laval (2023): “In steady-state [...] the circulating flow over the total network distance, has to match the incoming production [of trip distance]”, i.e., $\dot{m}(t) = 0$.

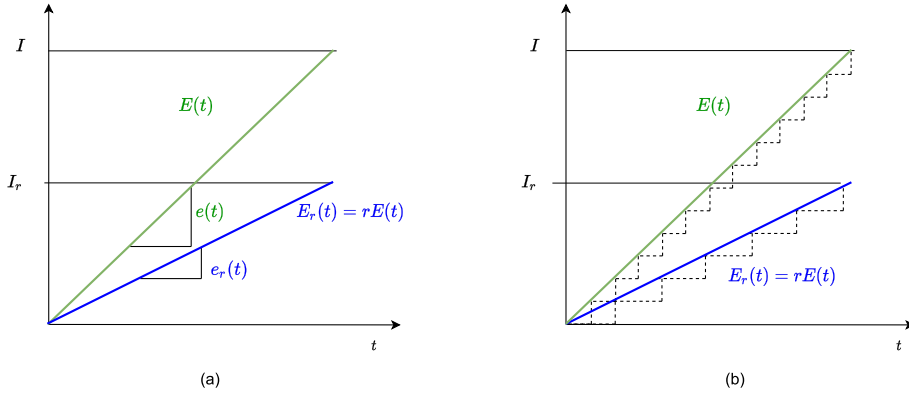


Fig. 5. Downscaling of demand: (a) For continuous $E(t)$. (b) For piece-wise constant $E(t)$.

This unbiased flow-scaling property of the bathtub model allows the modeler to study the traffic dynamics of a large system with a lower number of agents and a smaller network with $r < 1$. For example, the modeler can create a normalized twin city with $L_{N,r} = 1$ and total demand $I_r = \frac{I}{L_N}$ to study traffic patterns. This scalable property of the bathtub model will allow modelers to use non-dimensional analysis to study the traffic congestion patterns in very large cities. Alternatively, if the modeler has information on the characteristics of a reduced number of agents in the system, I_r , instead of the total number of trips during a study period in a study network and mode (I) on a network of L_N lane-distance, the traffic dynamics in the system can be modeled with agents I_r in an equivalent city setting the network lane distance to $\frac{I}{I_r} \cdot L_N$.

4.2. Numerical errors induced by flow-based scaling in the AB²M

From Eq. (1), the scaling of the total number of agents influences the inflow of trips as $e_r(t) = I_r \tilde{\varphi}(t)$. For continuous demand, the cumulative trip initiation rate is $E_r(t) = rE(t)$. However, since the AB²M is by nature discrete, the modeler should first approximate the piecewise constant $E(t)$ to a piecewise linear function, scale the approximation down, and discretize it again as a piecewise constant function; see Fig. 5.

There are two different types of error that can arise when downscaling the AB²M: First, the downscaling can lead to a non-integer number of agents in a discrete demand, as discussed in Section 4.2.1. Second, the downscaling can reduce the network in a way where a single agent starting or finishing their trip leads to large speed variations (and non-representative results of the original system), as will be discussed in Section 4.2.2.

4.2.1. Discrete and deterministic demands

In the case of a discrete demand, where $I\tilde{\varphi}(t, x)$ is defined by an integer number of agents that start their trip at certain discrete times, $T(i)$, with certain trip distances $X(i)$, the downscaling can introduce numerical errors if r is not chosen carefully. The reason is that not all agents might be represented if downscaling requires rounding. Let us consider a discrete demand, where n_{km} is the number of agents with departure times T_k and trip distances X_m . In that case, the lowest scaling ratio r_{min} , needs to guarantee that $r_{min} \cdot n_{km}$ are integers $\forall m, k$. Therefore, the lowest scaling ratio can be obtained from $\frac{1}{r_{min}} = GCD$, where GCD is the highest common divisor of all n_{km} . For example, a network with two trip distances $\tilde{D}_1$ and $\tilde{D}_2$ and two departure times T_1 and $T_2 > T_1$. As depicted in Fig. 6(a), the demand is defined as follows: at time T_1 , 550 agents start their trip, with 250 agents having $\tilde{D}_1$, and 300 having $\tilde{D}_2$. Then, 180 agents with $\tilde{D}_1$ and 50 agents with $\tilde{D}_2$ depart at T_2 . Choosing $r = 1/50$ leads to a non-integer number of agents, see Fig. 6(b). Therefore, choosing 3 or 4 agents with trip distance $\tilde{D}_1$ departing at T_2 would lead to numerical errors. Instead, the lowest scaling ratio should be $r_{min} = 1/10$, because 10 is GCD of n_{km} , see Fig. 6(c).

Note that when the GCD value is significantly small, e.g., 2, the reduction in computational complexity achieved through scaling may be marginal, e.g., only halving the computational time. In those cases, choosing a smaller scaling ratio than r_{min} requires careful evaluation of possible numerical errors.

In contrast, when the demand is stochastic the scaling ratio r will influence the sample size of agents obtained from the stochastic joint distribution of the demand. If the trip distance distribution is a continuous stochastic function, e.g., a NE distribution, a large sample size will represent the distribution well, but a very small sample size may not. Thus, the accuracy of the model may be compromised. To solve this, multiple Monte Carlo simulations can be performed. In theory, the scaling ratio r does not influence the expected behavior as long as the reduced number of agents does not induce numerical errors through excessive speed variation as we will discuss in the next section.

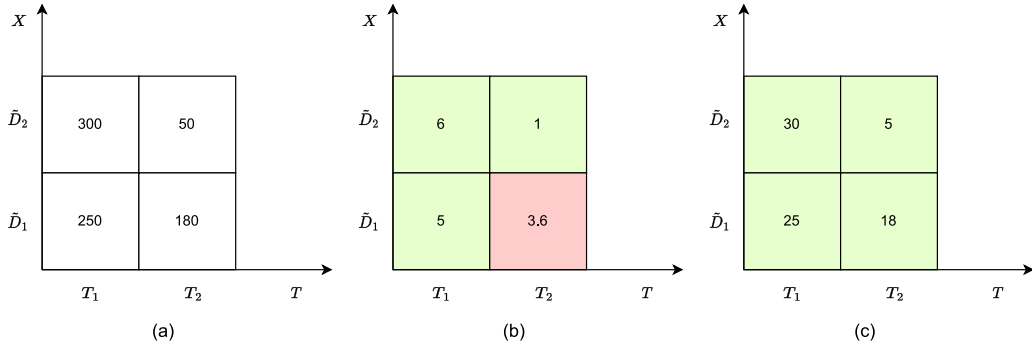


Fig. 6. Example of ensuring integer number of agents with deterministic discrete demands. (a) Original sample of agents, (b) Scaled demand with $r = 1/50$, (c) Scaled demand with $r = 1/10$.

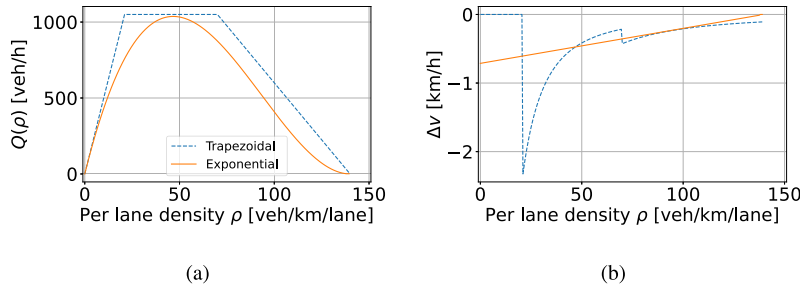


Fig. 7. Numerical examples with free-flow speed of $u_f = 50$ km/h and per lane jam density of $\rho_j = 140$ veh/km/lane. (a) Speed density relation for the trapezoidal NFD with $C = 1050$ veh/h and shock wave speed $w = 15$ km/h, and the exponential NFD. (b) Speed variation for the two NFDs at different densities corresponds to a network with $L_N = 1$ and a single vehicle entering or leaving the system, i.e., $\Delta\delta = \Delta\rho = 1$.

4.2.2. Speed variation

If r is too small, a single vehicle entering or leaving the system could cause significant changes in density and speed, introducing numerical errors. For example, modeling a network as a non-dimensional system with $r = \frac{1}{L_N}$ might not allow capturing a smooth change in the speed over time even if the time step is very small and ensures a low number of entering agents each time step. This is because a single vehicle entering or leaving the network can substantially modify speed. In fact, the rate of change in speed depends on the change in the accumulation of active trips $\delta(t)$ as

$$\frac{dV(\delta)}{d\delta} = \frac{dV(\rho)}{d\rho} \frac{1}{rL_N}, \quad (11)$$

which is inversely proportional to the total network length for a given density.

As an example, let us consider two different NFDs presented in Fig. 7(a) with the same free-flow speed and per lane jam density: the trapezoidal NFD $V(\rho) = \min\{u_f; \frac{C}{\rho}\}$; $w(\frac{\rho_j}{\rho} - 1)$; and the exponential NFD $V(\rho) = u_f(1 - \frac{\rho}{\rho_j})^2$. The derivative of the speed with respect to the density is depicted in Fig. 7(b) for the NFDs considered. For trapezoidal NFD, the largest $|\frac{dV}{d\rho}|$ is observed at the critical density; while in exponential NFD the largest variation is for $\rho = 0$. From Eq. (11) a downscaled city with $L_{N,r} < 1$ will lead to higher speed variations than those presented in Fig. 7(b). Therefore, if the modeler wants to guarantee a smooth speed (low Δv), she can determine the minimum reduced network lane distance from Fig. 7 as $r_{min}L_N = \frac{1}{\Delta v} \max_{\rho} \{ \frac{dV(\rho)}{d\rho} \}$. For example, for the above NFDs with $\Delta v = 0.1$ we would have $r_{min}L_N = 10$ km for the exponential NFD, and $r_{min}L_N = 25$ km for the Trapezoidal NFD.

In summary, although there are significant computational benefits in reducing r , this also leads to an increase in maximum speed variation for each unit vehicle that enters or exits the system Eq. (11). Therefore, the modeler needs to evaluate this type of trade-off to choose r to obtain accurate results in a computationally efficient manner. This trade-off analysis is presented in Section 6 with numerical results.

5. Complexity analysis

In this section, we compare the computational complexity of the algorithms proposed and analyze the impact of downscaling. We are interested in analyzing the complexity of a system with a certain simulation period of t_f with fixed time step Δt . To do so, we define the average inflow rate per unit distance during the simulation period as

$$\bar{e} = \frac{I}{L_N t_f}. \quad (12)$$

As discussed earlier, Part 1 and Part 3 of Fig. 1 are the same for the naive AB²M formulation and the efficient priority queue formulation proposed in Section 3. In the setup, sorting all the agents by departure time can be done with efficient algorithms relying on tree-sorting, where the cost is $n \log(n)$ for a list of n elements (Hetland, 2010). The post-process, i.e., finding the completion time for each trip, can be done numerically with any zero of functions algorithm search Eq. (7). For example, a binary search algorithm can be used, where the cost is logarithmic with the number of time steps (Hetland, 2010).

Lemma 1. *The upper bound complexity of Part 1 and Part 3 of the algorithms to model AB²M are $\mathcal{O}(I \log(I))$ for Part 1, and $\mathcal{O}\left(I \log\left(\frac{I}{\bar{e} \Delta t L_N}\right)\right)$ for Part 3, where $\bar{e}$ is the average inflow.*

Proof. The generation of I agents has cost I and sorting the agents by departure time can be done in $\mathcal{O}(I \log(I))$, with the most efficient algorithms relying on tree-sorting (Hetland, 2010). Thus, overall, the complexity of Part 1 is $\mathcal{O}(I \log(I))$. For Part 3, the time of completion for each trip has computational upper bound complexity of $\mathcal{O}(\log \frac{t_f}{\Delta t})$, if the time step is fixed Δt . Then, the total post-process cost is $\mathcal{O}(I \log \frac{t_f}{\Delta t})$. From Eq. (12), we have the upper bound complexity for Part 3 is $\mathcal{O}\left(I \log\left(\frac{I}{\bar{e} \Delta t L_N}\right)\right)$. $\square$

From Lemma 1, the computational cost increases log-linearly with increasing the number of agents. Instead, it decreases logarithmically with larger Δt , $\bar{e}$, and networks' lane distance, L_N . Although the last two seem counter-intuitive, with a fixed number of agents I lowering the average trip initiation rate will lead to a longer simulation period and thus will increase the computational cost.

In the following, we discuss the computational complexity for the simulation dynamics (i.e., Part 2 in Fig. 1).

Theorem 1. *The naive algorithm proposed in Fig. 1 has the following upper bound complexity for Part 2 of $\mathcal{O}\left(\frac{I^2}{\bar{e} \Delta t L_N}\right)$. This quadratic complexity from Part 2 dominates the overall complexity of the naive algorithm.*

Proof. The simulation runs for $\frac{t_f}{\Delta t}$ time steps. In each time step, the cumulative inflow $E(t)$ is obtained from a binary search among all the sorted agents, i.e., $\mathcal{O}(\log(I))$. In each time step, there is an update of the remaining trip distance as well as a search over all agents that have left their origin, $E(t)$, see Eq. (6). Thus, the total cost is $\mathcal{O}\left(\frac{t_f}{\Delta t} (\log(I) + I)\right)$, and from Eq. (12), the upper bound is $\mathcal{O}\left(\frac{I^2}{\bar{e} \Delta t L_N}\right)$. $\square$

Theorem 2. *The priority queue Algorithm 1 upper bound complexity for Part 2 is $\mathcal{O}\left(\frac{I}{\Delta t \bar{e} L_N} \log(I) + I \cdot \log(\rho_j L_N)\right)$. Thus, the Algorithm is dominated by Part 1 computational complexity unless $\Delta t \ll \frac{1}{\bar{e} L_N}$, which would make $\mathcal{O}\left(\frac{I \log(I)}{\Delta t \bar{e} L_N}\right)$ dominate the algorithm.*

Proof. Similar to in the naive formulation, there is a cost of defining $E(t)$ for all time steps, i.e., from Eq. (6) $\mathcal{O}\left(\frac{I}{\Delta t \bar{e} L_N} (\log(I))\right)$. The complexity from inserting an agent to and removing it from the priority queue of size $\delta(t)$ is $\mathcal{O}(\log(\delta(t)))$ for a heap structure (Hetland, 2010). Thus, the computational cost of incorporating $e(t)$ elements into a sorted heap is

$$\sum_{k=0}^{e(t)-1} \mathcal{O}(\log(\delta(t - \Delta t) + k)).$$

If $e(t) < \delta(t - \Delta t)$, it can be approximated by $\mathcal{O}\left(e(t) \log(\delta(t - \Delta t) + \frac{e(t)}{2})\right)$, which is dominated by $\mathcal{O}(e(t) \log(\delta(t - \Delta t)))$. Since the number of active trips is always $\delta(t) \leq \rho_j L_N$, the upper bound complexity of inserting and removing all agents is $\mathcal{O}(I \cdot \log(\rho_j L_N))$. $\square$

Corollary 2.1. *Adding and removing agents to the priority queue is dominated by defining $E(t)$, when $I \gg \rho_j L_N$, e.g., for longer periods of time or high $\bar{e}$. Note that if $\bar{e}$ is very small or the time step is very large, i.e., $\Delta t \gg \frac{1}{\bar{e} L_N}$, adding and removing agents will dominate.*

From Theorems 1 and 2, the computational complexity of Algorithm 1 is significantly lower than a naive algorithm without exploiting the SCDFO principle of the bathtub model. The main difference is caused due to the method of calculating the completion rate for trips. In the naive algorithm, Eq. (6) is used each time step to compare $E(t)$ times whether $x(t, i) < z(t)$. Fig. 8 presents a comparison of computational costs. For example, for 10 million agents, Part 1 is less than 5 s, Algorithm 1 has a cost of 20 s, and the Naive formulation is more than 200 s.

As discussed in Section 4.2.2, when using a fixed time-step algorithm, one wants to ensure that the Δt is small enough to eliminate numerical errors and captures adequately the changes in density in the system. Note that there is no requirement on the time step Δt for the AB²M to be well defined. Δt should be only small enough to eliminate numerical errors. From Theorem 2, it is clear that choosing a very small Δt can increase the computational complexity of Part 2 of the algorithm. A first rough estimate for a time step can be obtained from $\Delta t \leq \frac{t_f}{I}$, which aims to capture on average the start of an agent. From Eq. (12), we have

$$\Delta t \leq \frac{1}{\bar{e} L_N}, \quad (13)$$

Thus, scenarios with (or periods of) higher average inflow will require smaller Δt . In fact, for a fixed simulation period t_f , it is reasonable to decrease hyperbolically Δt with an increasing number of agents, I , to maintain the same accuracy. However, from

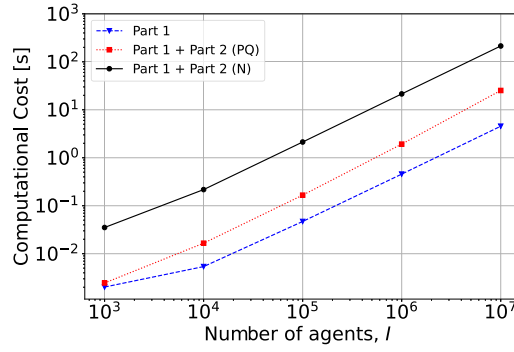


Fig. 8. Comparing Naive (N) and Priority Queue (PQ) formulations for 30 min of simulation, with $\Delta t = 20$ s.

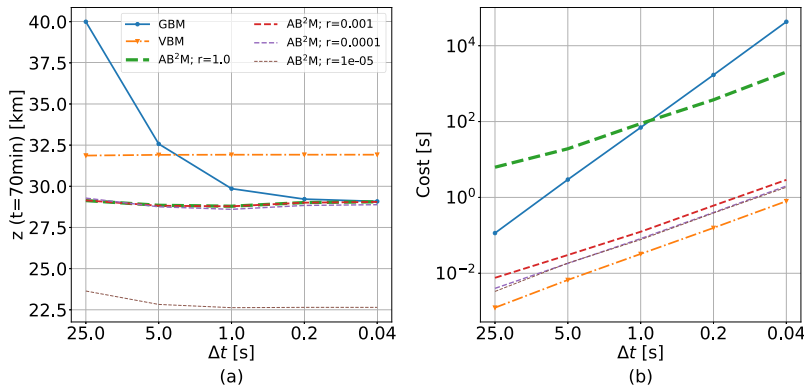


Fig. 9. Comparison for three different models. Example with 1 953 125 agents in 2h with trip distance $\bar{D} = 2$ km. (a) Convergence of $z(t = 70 \text{ min})$ for decreasing Δt . (b) The computational cost for decreasing Δt .

Section 4.2.2, larger networks can allow more agents to enter or leave the system by maintaining the same accuracy (when using Δv as a proxy). In all, there is a non-trivial relation on how to set up Δt given the demand and supply characteristics. This complex relation goes beyond a trade-off between accuracy and computational cost when considering multiple parameters, such as the network length. Note that a reasonable lowest threshold of Δt would be in the order of seconds. The exact selection of Δt should be established after performing a convergence analysis for each scenario.

The flow-based scaling can further reduce the computational cost. From Theorem 1, the algorithm for a single simulation run in the scaled system has the complexity $\mathcal{O}(\frac{I \log(rI)}{\epsilon \Delta t L_N})$.⁴ Moreover, scaling the system can have implications for the time step that should be chosen. In particular, downscaling the system allows one to have a larger time step and still capture accurately the inflow of each individual vehicle following Eq. (13). Thus, considering $\Delta t_r = \frac{1}{\epsilon L_N r}$, can further reduce the computational complexity as $\mathcal{O}(rI \log(rI))$. Thus, a downscaled network $r < 1$ can have a significant computational benefit.

6. Numerical results

6.1. Comparison to existing models

In this section, we aim to compare the trip flow dynamics by using different models and under different demand patterns. We also compare the computational cost of the priority-queue formulation of AB^2M with other existing models, such as the GBM,⁵ discretized with the finite difference method (see Section 3.3 in Jin, 2020) and the VBM, also called accumulation-based model (Mariotte et al., 2017) or PL model (Sirmatel et al., 2021). In this section, the speed-density relation assumed is the exponential NFD with the free-flow speed of 50 km/h and per lane jam density of 140 veh/km/lane, as in Fig. 7.

First, a convergence analysis for the selection of Δt is required. To do so, we compare the value of $z(t)$ at a given time for multiple models, see Fig. 9(a). In this case, a constant trip distance for all trips is assumed with constant inflow over 2 h of simulation, and we observe that the error converges with smaller Δt for all models. While VBM and AB^2M errors are not sensitive to Δt , GBM

⁴ Note that some of the scaling effects are canceled out $\mathcal{O}(\frac{rI \log(rI)}{\epsilon L_N \Delta t})$.

⁵ Comparing the AB^2M to the GBM allows us to compare it to the TBM since they are equivalent for time-independent trip distance distributions.

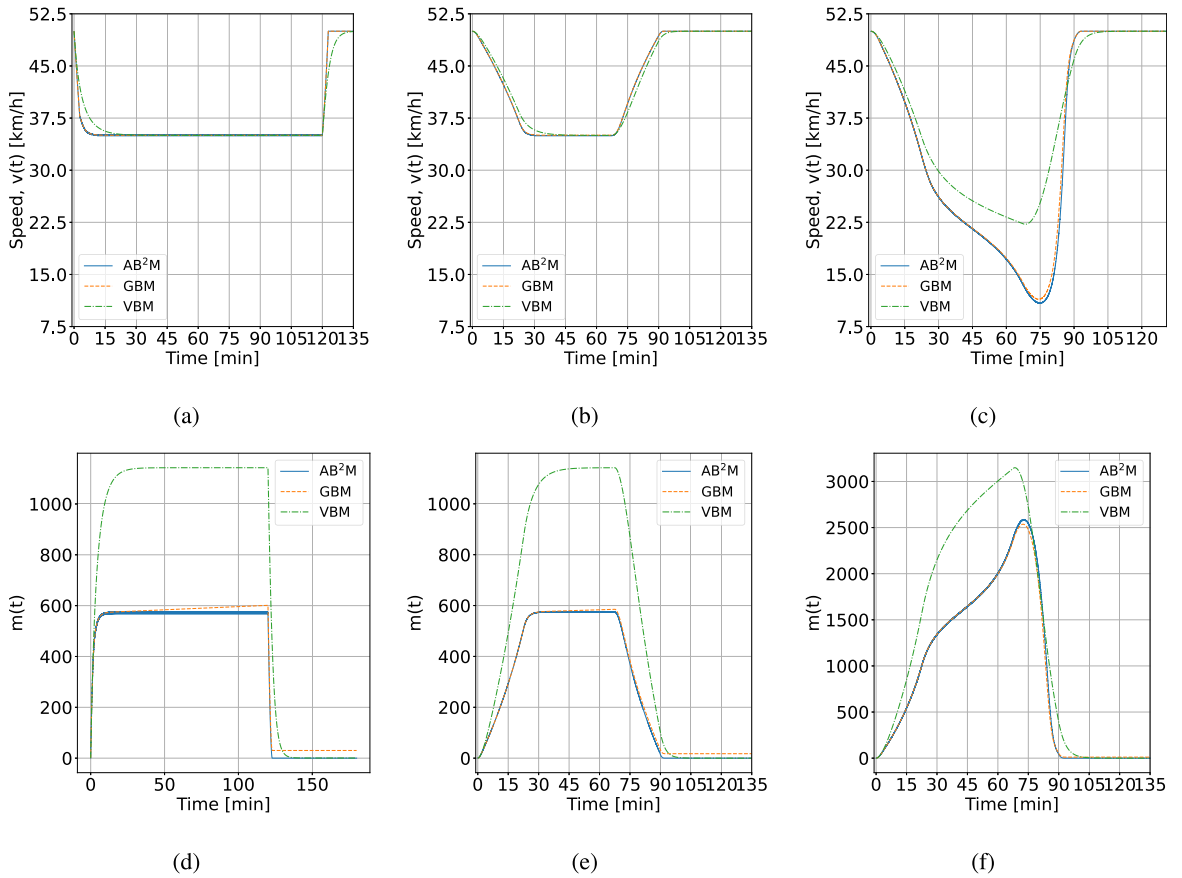


Fig. 10. Dynamics in a city of $L_N = 25$ km with time step $\Delta t = 0.1$ s. The upper figures present the speed dynamics, and the lower figures present the total remaining trip distance dynamics. (a, d) Constant inflow over 2 hours and trip distance of $\bar{D} = 2$ km. Total number of agents $I = 20000$. (b, e) Trapezoidal inflow over 1.5 hours and constant trip distance of $\bar{D} = 2$ km. Total number of agents $I = 11250$. (d, f) Trapezoidal inflow over 1.5 hours and constant trip distance of $\bar{D} = 2.7$ km. Total number of agents $I = 11250$.

requires very small Δt to converge. Moreover, AB^2M and GBM converge to the same value of $z(t) \sim 29$ km, while VBM converges to a different value. This is expected, since the underlying assumption of VBM is a NE distribution of trip distances instead of constant trip distances. The results suggest that VBM underestimates congestion since it leads to higher $z(t)$. In Fig. 9(b) the computational costs of the three models are compared. For $\Delta t = 0.2$ s the cost for the AB^2M is around 6 min, while the VBM is below 1 s, and the GBM is above 25 min.

Although the impact of scaled systems for constant, deterministic trip distances⁶ will be discussed in detail in Section 6.2, we observe in Fig. 9 that the AB^2M with natural scaling of $r = 0.01\%$ (equivalent to using 1953 agents) still leads to the same results with a cost below 1 s. However, an excessive downscaling to 19 agents causes the AB^2M to deviate from accurate convergence, underscoring the importance of selecting the scaling factor r .

In the following, we will consider the traffic dynamics under three different demand scenarios, with constant trip distance for all agents. In two out of three scenarios, a steady state is reached, and both the GBM and ABM lead to the same results, exact traffic dynamics in Fig. 10. Instead, the VBM leads to some errors in the transition period due to the underlying assumption of NE distribution of trip distances. Moreover, these numerical simulations validate the conclusion that the speed for GBM and VBM are equivalent during steady states (i.e., $m(t) = 0$), already suggested by Mariotte et al. (2017), Lamotte et al. (2018) and Laval (2023). However, we show that the $m(t)$ dynamics are clearly different between VBM and the other two models, where $m_{VBM}(t) \gg m_{AB^2M}(t)$ during the steady state. In the third scenario, where the trip distance $\bar{D}$ is significantly larger, the demand is larger than the supply until the end of the peak period. Thus, no steady state is observed. In this case, the VBM significantly underestimates the congestion level.

⁶ The impact of scaling for stochastic demands would require the study of Monte Carlo simulations, which is out of the scope of this paper.

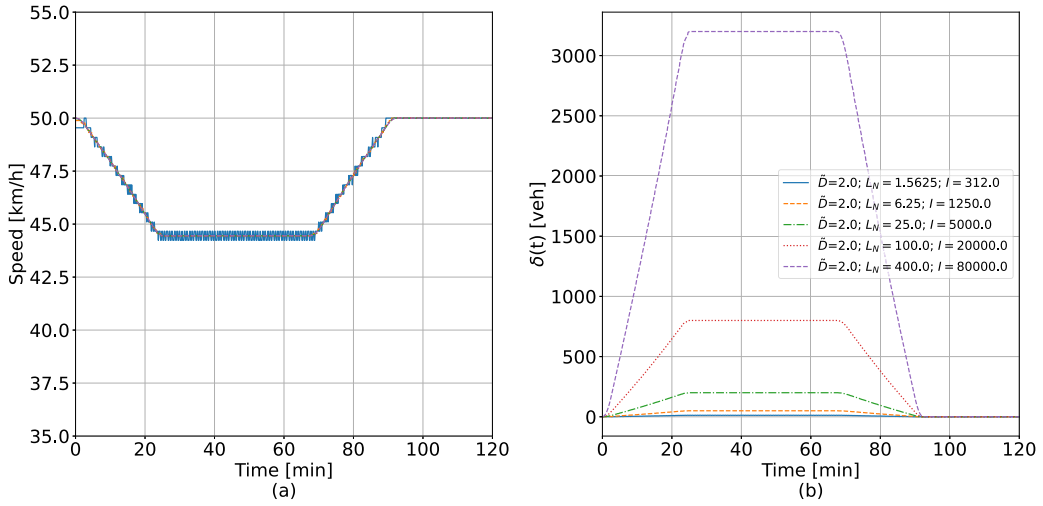


Fig. 11. Constant trip distances and trapezoidal inflow. Flow-based scaling modifying I and L_N with $\Delta t = 10$ s.

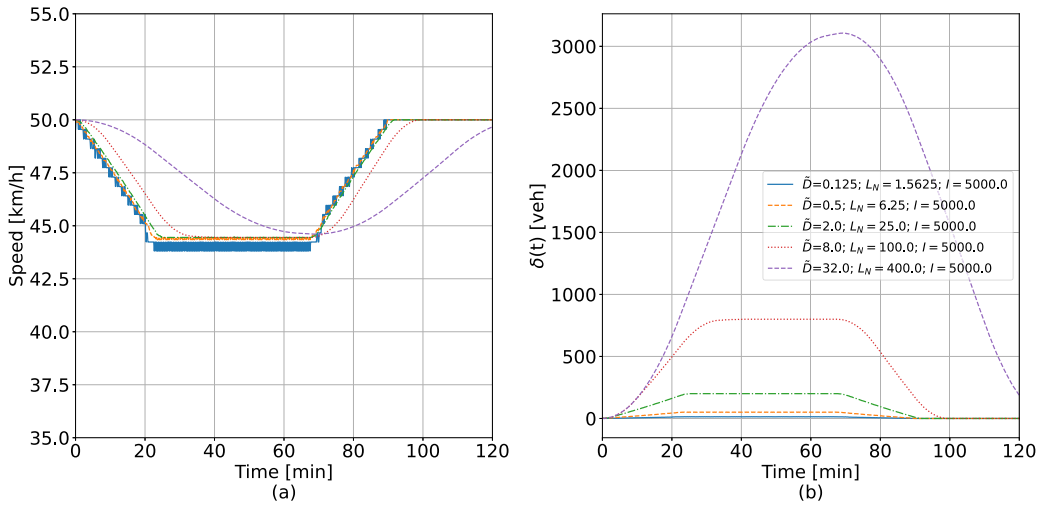


Fig. 12. Constant trip distances and trapezoidal inflow. Distance-based scaling modifying $\bar{D}$ and L_N with $\Delta t = 10$ s.

6.2. Comparison of scaled systems

In this section, we want to compare the impact of scaling on the trip flow dynamics and the computational cost. The basic scenario that we consider is a network with $L_N = 25$ km and 5 000 agents traveling a distance of 2 km. We will consider two down-scaled and two up-scaled scenarios, with $r = [6.25\%, 25\%, 400\%, 1600\%]$. Taking into account an exponential NFD and a trapezoidal demand over 90 min with a peak duration of 45 min, we evaluate the congestion pattern (i.e., speed evolution) for different r .

Considering the (natural) flow-based scaling that modifies the network lane distance and the number of agents, the results are presented in Fig. 11. Clearly, the results in the speed profile are the same for all systems, and only the total number of circulating trips changes. As predicted from the analysis in Section 4.1 no biases are introduced with flow-based scaling. However, downscaling too much introduces some numerical errors.

In Fig. 12, we present the simulation results of the distance-based scaling. As explained in our analytical discussion, the steady state remains consistent across scaling factors between 25% to 400%, and the transition period varies more for the up-scaled system. In larger cities characterized by greater trip distances, the simulations show extended transition phases. This is particularly evident in the case of $L_N = 400$, where our model predicts that the system fails to reach a steady state with the assumed trapezoidal inflow. Moreover, downscaling too much seems to also lead to another steady state but the difference is small and could be attributed to numerical errors. These findings validate our earlier hypotheses and offer new insights into the scalability of bathtub models, suggesting that only flow-based scaling should be used.

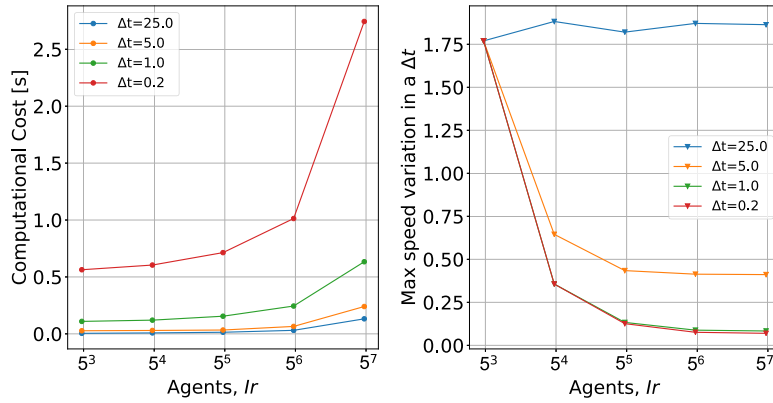


Fig. 13. (a) Evaluation of computational cost, (b) Evaluation of maximum speed variation.

Table 3

Six cases considered. Note that the parameters of the log-normal (LN) distribution for case 1c are $\mu \sim 0.648$ and $\sigma = 0.3$, and for case 2c are $\mu \sim 0.871$ and $\sigma = 0.3$.

	Constant	NE distribution	LN distribution
$\bar{D} = 2$ km	Case 1a	Case 1b	Case 1c
$\bar{D} = 2.5$ km	Case 2a	Case 2b	Case 2c

Choosing r to ensure low computational cost but high accuracy might not be a trivial task. Fig. 13 presents a numerical analysis of the computational cost and the maximum speed variation measured in a time step (used as a proxy of the numerical error). The numerical error is eliminated by reducing Δt for several r . In contrast, for $r = 0.08$, the same speed variation is observed for all time steps. Thus, only for certain rI convergence can be achieved with smaller Δt , emphasising the importance of carefully choosing r .

6.3. Trip travel time distribution

In this section, we will discuss the trip travel time distribution (TTTD). The TTTD is one of the higher-order moments that can be analyzed with the AB²M but cannot be obtained with aggregated continuum models. To study the TTTD, we will consider different demand assumptions. Under stochastic demand patterns, i.e., when the assumption of the trip distance distributions is a distribution, the AB²M results will depend on the sample of trips from the trip distance distributions. Studying the expected behavior of systems with stochastic demand requires the implementation of Monte Carlo simulations. Because this is out of the scope of this paper, we will consider a single sample of trips with different assumptions of trip distance distributions. The traffic dynamics under the three assumptions of trip distance distributions for the same trip initiation rate $e(r)$ will be compared. All cases considered here correspond to a peak period of one hour with trapezoidal trip initiation rate. We will consider six demand cases, with the same total number of agents ($I = 10\,000$) and same network length ($L_N = 30$ km) but different trip distance distributions. In particular, two mean trip distances and three different assumptions will be considered: (i) A deterministic, discrete demand through the BBM with constant trip distance $\bar{D}$, (ii) a sample of trips from a stochastic time-independent NE distribution, with a mean trip distance of $\bar{D}$, and (iii) a sample of trips from a stochastic time-independent log-normal (LN) distribution with mean trip distance $\bar{D}$, since several empirical studies suggest that is a better assumption (Colak et al., 2016; Yang et al., 2018; Martínez and Jin, 2021). The six cases analyzed are summarized in Table 3.

First, we will have a look at the aggregated information. The average speed of the system, as well as the cumulative departure and arrival curves, are presented in Fig. 14. Note that these results could be obtained through continuum bathtub models. The average speed of the system presented in Fig. 14(a, b) shows that in all cases 1a-c, the system will reach the same steady state for a certain period of time. In contrast, cases 2a-c with $\bar{D} = 2.5$ km will experience increasing congestion until the demand is reduced again. Clearly, the mean aggregated behavior for low demand (cases 1a-c) is not significantly different for the constant trip distances and the LN distribution. However, with the assumption of NE, the steady state is reached about 10 min later and is shorter. The congestion period is slightly longer overall for the NE assumption than in the other two cases. The differences in the congestion dynamics caused by different demand patterns are more evident under high demand (cases 2a-c), where the case with NE leads to significantly less severe congestion, even if the congestion lasts longer. This can be explained by the fact that under the NE distribution, most trips have very short trip distances, which stay less time in the system and are completed faster. However, the probability of having very long trips is larger than the LN distribution, and these few and very long trips stay longer in the network.

In the following, we take advantage of the information on the microscopic level, which is enabled through the proposed AB²M formulation. We can look at the trip initiation and completion times of each individual trip and the agent's travel time. The results are presented in Fig. 15, where the differences in the trip distance distribution assumptions are much more evident. For the case

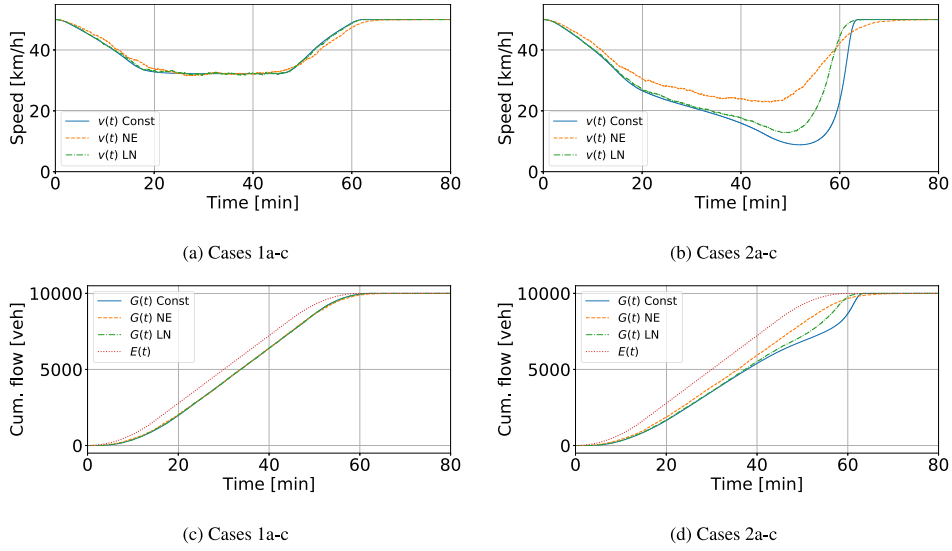


Fig. 14. Comparison of macroscopic information under six different demand scenarios described in Table 3. (a, b) Speed profiles and (c, d) Cumulative curves (inflow $E(t)$ and outflow $G(t)$).

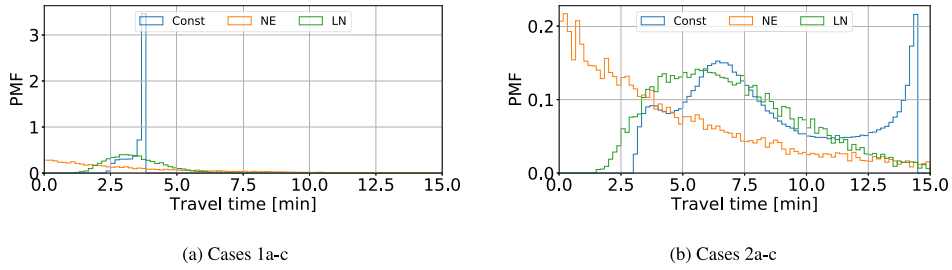


Fig. 15. Comparison of microscopic information for simulation cases presented in Table 3. (a) Trip travel time distribution for $\tilde{D} = 2.0$ km, (b) Trip travel time distribution for $\tilde{D} = 2.5$ km.

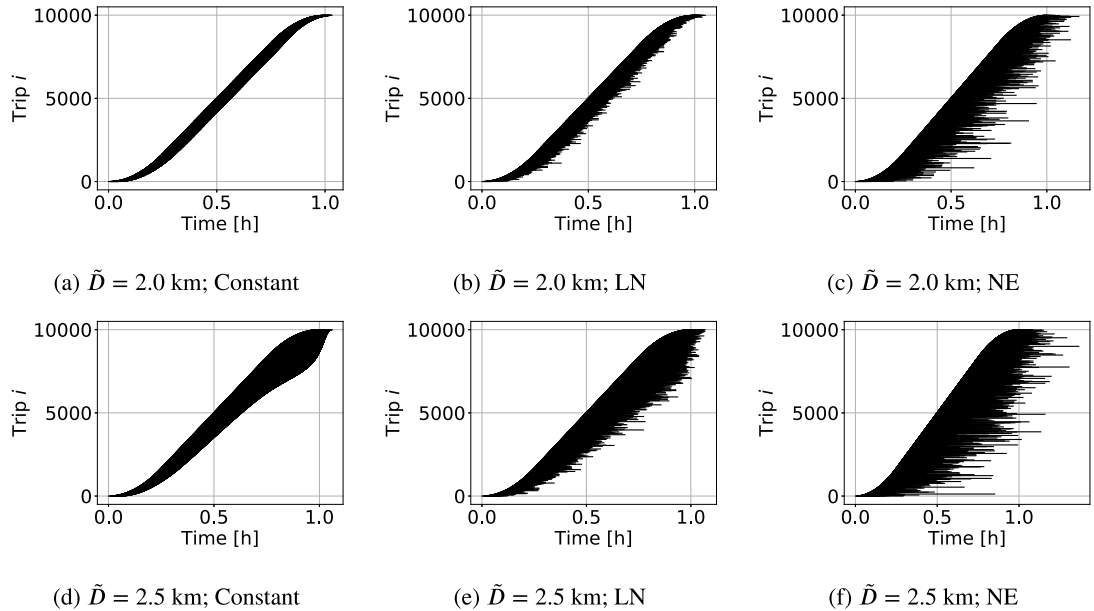


Fig. 16. Departure ($T(i)$) and arrival ($\hat{T}(i)$) for 10 000 individual trips, where the width of each black bar is the agent's travel time; each figure corresponds to a different trip distance distribution assumption.

with constant trip distances (i.e., equivalent to the BBM), the distribution of travel times has a maximum travel time, which is the most likely. While the NE assumption clearly has the shortest travel times, and the LN distribution has a bell-shaped travel time distribution. Further, the AB²M information on individual trips allows us to visually present the travel time of individuals with an analogous diagram to the cumulative curves, where the horizontal black lines represent the travel time for each individual. For this example, we present in Fig. 16 an analogous diagram to the cumulative curves, where $E(t)$ is effectively the same, but $G(t)$ is replaced by the departure times of individuals. These diagrams show the nature of the bathtub model queuing system, where the horizontal distance represents the travel time for each agent i but the vertical distance between the curves is not the number of active trips in general. Note that the BBM follows FIFO. Thus, $\Theta(t, n)$ does not require sorting, and the height of the black area at any given time represents $\delta(t)$. On the other hand, both NE and LN assumptions lead to FIFO violations. That means that a shorter trip that enters the network at t may exit earlier than a longer trip that entered earlier than t . This evidences the importance of sorting $\Theta(t, n)$ for trip distance distributions different from the constant trip distance for all users.

7. Conclusion

The use of bathtub models, a.k.a. reservoir models, is becoming increasingly popular among researchers. The main difference from the traditional transportation models is that the bathtub model does not require setting up the physical network or tracking the agents' location within the network. Instead, the bathtub models capture the network flow dynamics in a relative space with respect to the trips' destinations. These bathtub models simplify the calibration and improve computational efficiency by eliminating vehicle position tracking and utilizing a global speed assumption. Additionally, bathtub models preserve privacy by avoiding the collection of personally identifiable location information. In this study, we presented the agent-based bathtub model (AB²M), which can be understood as a microscopic model that tracks individual trips' initiation, progression, and completion in Lagrangian coordinates on relative space. This discrete version of the bathtub model has been overlooked in the literature because it is claimed to have a high computational cost, similar to traditional agent-based models in the absolute space (Kagho et al., 2020). However, using an agent-based formulation to describe the bathtub model traffic dynamics has the advantage that agent heterogeneity can be easily incorporated. Based on the higher-order moments obtained from Monte Carlo simulations, one could study the travel time reliability and other variables.

We made two contributions to enhance the efficiency of the AB²M: First, in Section 3, we leverage the “shorter-(characteristic)-distance-first-out” (SCDFO) principle (Jin, 2020) to introduce $\Theta(t, n)$ as a sorted collection of active trips based on their characteristic trip distance. This variable plays a vital role in proposing an efficient algorithm that uses $\Theta(t, n)$ as a priority queue with binary trees. Secondly, in Section 4, we examine the scalability of bathtub models and propose two downsizing approaches: the flow-based downsizing, which reduces the number of agents without introducing biases into the trip flow dynamics, and the distance-based downsizing, which is proven to be equivalent to the original system only in steady states. Numerical results show that if the distance-based downsizing scaling ratio is very low or very large, the dynamics will differ significantly. Therefore, we conclude that only the flow-based scaling should be used to reduce computational complexity without introducing biases in non-stationary situations. Then, a systematic discussion on the computation complexity of the algorithms is presented in Section 5. Finally, through numerical simulations, we discussed the differences between the AB²M and two of the most established continuum bathtub models, i.e., the Vickrey's bathtub model (VBM) (Vickrey, 2020), sometimes referred to as accumulation-based model (Mariotte et al., 2017) or PL model (Sirmatel et al., 2021), and the Generalized bathtub model (Jin, 2020). Through numerical examples, we also study the higher-order moments, such as trip travel time distribution, for three different common assumptions on the trip distance distribution, i.e., the negative exponential distribution (Vickrey, 2020), the constant trip distance (Arnott et al., 1993; Arnott and Buli, 2018), the log-normal distribution (Martínez and Jin, 2021). We showed that for the same average trip distance, the negative exponential distribution leads to less severe congestion and shorter travel times than systems with constant trip distance or a log-normal distribution. This highlights the importance of not relying on VBM to develop management strategies at the network level since they could underestimate the travel time of users. Further, the numerical simulations also shed some light on the criteria to select Δt and r for a down-scaled city.

An AB²M has several advantages over traditional ABM in the absolute space: First, it is computationally more efficient because one does not need to track the position of vehicle trips in the real network, which ensures the preservation of personally identifiable location information, addressing privacy concerns associated with individual data collection in ABM in absolute space. Secondly, the speed of all vehicles is the same (global speed), determined by a single equation based on the NFD assumption. In contrast, in the absolute space, the speed is local, usually determined at the link level. Thirdly, the flow-based downsizing employed in AB²M introduces no biases and exhibits no physical limitations on the extent of downsizing the system.

In summary, the main contributions of this paper are threefold:

1. A methodological/modeling contribution through the definition of $\Theta(t, n)$ as a priority queue (as binary trees) of active agents sorted by characteristic trip distance, based on the SCDFO principle.
2. Although the downsizing of agents is a natural idea and has been implemented in the past, this paper thoroughly discusses two scaling implementations as well as their impacts both analytically and numerically. We show that the flow-based downsizing of AB²M allows us to reduce the computational cost without introducing biases. Conversely, our analysis demonstrates that distance-based scaling is not desirable due to its tendency to introduce bias into traffic states. This bias is especially notable in non-stationary states.
3. We rigorously established the upper bound of the computational complexity of the two fixed time step formulations of the AB²M, and the benefits of downsizing.

This study is the first step towards a better understanding of agent-based modeling in the relative space, the impacts of downscaling, and time-step discretization on the computational efficiency and numerical errors. From the insights gained in this paper, an event-based formulation of the AB²M, as the ones proposed by Mariotte et al. (2017) and Lamotte et al. (2018), may consider longer event periods than the start or end time of each trip and still eliminate the numerical errors. In the future, we are interested in a detailed comparison of accuracy and computational complexity between a discrete time step and an event-based formulation of the AB²M.

The numerical examples in this study are based on deterministic demand patterns. The insights provided in this paper on selecting the scaling factor may not apply to stochastic demands. Understanding how the scaling factor and the choice of the number of agents can accurately reflect stochastic demand and traffic congestion patterns is crucial. In the future, we will use the AB²M to model stochastic demands through Monte Carlo simulations. Ongoing preliminary research indicates that trip distance distributions with smaller standard deviations are less sensitive to downscaling. A higher standard deviation is associated with more violations on first-in-first-out (FIFO) for a given sample size. Thus, the relation between the FIFO violations and the impact of flow-based scaling should be studied in the future.

The proposed AB²M considered a single bathtub and single transportation mode (and thus a single speed for the whole set of trips at a given time). In the future, the AB²M should also be extended to study multi-region transportation systems with multiple connected bathtubs. To model several connected bathtubs, e.g., modeling the downtown and periphery of a city with different bathtubs, the extension of the AB²M should define how to handle the trip moving between the connected bathtubs.

Moreover, due to its computational efficiency, the AB²M has the potential to be extended to model multi-modal transportation systems and incorporate the decision making of independent agents based on behavioral rules. Due to the nature of agent-based modeling, AB²M can be extended to systems where the person trips and vehicle-trips are not the same. While the number of vehicles in the system determines the speed (i.e., rate of trip progression), the passenger trips could differ from the vehicle trips. For example, considering transit services where vehicles generally have high occupancy and users' trips are also composed of access and waiting time. In this case, the agents' trips can be modeled as a chain of trips, i.e., the trip consists of several stages or "states". These agents' states can be defined as a set or subset of attributes (e.g., the agent can be at her origin/destination, walking to the bus station, waiting for a bus, or traveling inside the bus). Each state of its trip has a different speed progression. The possible agent states should be defined depending on the transportation system to be modeled. An example of these states is presented in the compartmental model, where trips are planned, traveling, or completed for privately owned vehicle trips or can be planned, waiting, traveling, and completed for users of a shared mobility system (Jin et al., 2021). The trip chains can be modeled as different compartments or bathtubs.

Note that even if the downscaling (or up-scaling) in the relative space framework has been shown to be mathematically rigorous with a single-mode and single bathtub, the introduction of other modes and multiple bathtubs may require more discussions and extensions of the scaling properties. How this multi-modal system should be downscaled has not been systematically studied yet; neither in the absolute (Ben-Dor et al., 2021) nor relative spaces. In fact, multi-modal ABM presents a significant problem for downscaling in the absolute space. Future studies on downscaling should try to address these issues.

CRedit authorship contribution statement

Irene Martínez: Writing – review & editing, Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Conceptualization. **Wen-Long Jin:** Writing – review & editing, Supervision, Methodology, Investigation, Funding acquisition.

Acknowledgments

The authors would like to thank the support of the California Statewide Transportation Research Program (SB1), 2021–2022, and the SCC-IRG Track 1 from NSF-SCC CMMI 2125560. The authors would also like to thank the anonymous reviewers who helped improve the original version of the paper.

References

- Ameli, M., Faradonbeh, M.S.S., Lebacque, J.-P., Abouee-Mehrzi, H., Leclercq, L., 2022. Departure time choice models in urban transportation systems based on mean field games. *Transp. Sci.* 56 (6), 1483–1504.
- Arnott, R., 2013. A bathtub model of downtown traffic congestion. *J. Urban Econ.* 76, 110–121.
- Arnott, R., Buli, J., 2018. Solving for equilibrium in the basic bathtub model. *Transp. Res. B* 109, 150–175.
- Arnott, R., Kokoza, A., Naji, M., 2016. Equilibrium traffic dynamics in a bathtub model: A special case. *Econ. Transp.* 7–8 (August), 38–52.
- Arnott, R., de Palma, A., Lindsey, R., 1993. A structural model of peak-period congestion: A traffic bottleneck with elastic demand. *Am. Econ. Rev.* 83, 161–179.
- Auld, J., Hope, M., Ley, H., Sokolov, V., Xu, B., Zhang, K., 2016. Polaris: Agent-based modeling framework development and implementation for integrated travel demand and network and operations simulations. *Transp. Res. C* 64, 101–116.
- Balmer, M., Meister, K., Rieser, M., Nagel, K., Axhausen, K.W., 2008. Agent-Based Simulation of Travel Demand: Structure and Computational Performance of Matsim-T. *Arbeitsberichte Verkehrs-und Raumplanung*, p. 504.
- Bastarionto, F., Hancock, T., Choudhury, C., et al., 2023. Agent-based models in urban transportation: Review, challenges, and opportunities. *Eur. Transp. Res. Rev.* 15 (19).
- Ben-Dor, G., Ben-Elia, E., Benenson, I., 2021. Population downscaling in multi-agent transportation simulations: A review and case study. *Simul. Model. Pract. Theory* 108, 102233.
- Bonabeau, E., 2002. Agent-based modeling: Methods and techniques for simulating human systems. *Proc. Natl. Acad. Sci.* 99 (suppl_3), 7280–7287.
- Colak, S., Lima, A., Gonzalez, M.C., 2016. Understanding congested travel in urban areas. *Nature Commun.* 7 (10793).

- Crooks, A., Malleon, N., Manley, E., Heppenstall, A., 2018. Agent-Based Modelling and Geographical Information Systems: A Practical Primer. Sage.
- Daganzo, C., 2007. Urban gridlock: Macroscopic modeling and mitigation approaches. *Transp. Res. B* 41 (1), 49–62.
- Daganzo, C.F., Lehe, L.J., 2015. Distance-dependent congestion pricing for downtown zones. *Transp. Res. B* 75, 89–99.
- Fosgerau, M., 2015. Congestion in the bathtub. *Econ. Transp.* 4 (4), 241–255.
- Geroliminis, N., Daganzo, C.F., 2008. Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings. *Transp. Res. B* 42 (9), 759–770.
- Godfrey, J., 1969. The mechanism of a road network. *Traffic Eng. Control* 11, 323–327.
- Hetland, M.L., 2010. Python Algorithms: Mastering Basic Algorithms in the Python Language. Apress.
- Hörl, S., Balac, M., Axhausen, K.W., 2019. Dynamic demand estimation for an amod system in Paris. In: 2019 IEEE Intelligent Vehicles Symposium. IV, pp. 260–266.
- Jin, W.-L., 2020. Generalized bathtub model of network trip flows. *Transp. Res. B* 136, 138–157.
- Jin, W.-L., Martínez, I., Menendez, M., 2021. Compartmental model and fleet-size management for shared mobility systems with for-hire vehicles. *Transp. Res. C* 129, 103236.
- Johari, M., Keyvan-Ekbatani, M., Leclercq, L., Ngoduy, D., Mahmassani, H.S., 2021. Macroscopic network-level traffic models: Bridging fifty years of development toward the next era. *Transp. Res. C* 131 (August).
- Johnson, D.B., 1975. Priority queues with update and finding minimum spanning trees. *Inform. Process. Lett.* 4 (3), 53–57.
- Kagho, G.O., Balac, M., Axhausen, K.W., 2020. Agent-based models in transport planning: Current state, issues, and expectations. *Procedia Comput. Sci.* 170, 726–732, The 11th International Conference on Ambient Systems, Networks and Technologies (ANT)/The 3rd International Conference on Emerging Data and Industry 4.0 (EDI40)/Affiliated Workshops.
- Keyvan-Ekbatani, M., Kouvelas, A., Papamichail, L., Papageorgiou, M., 2012. Exploiting the fundamental diagram of urban networks for feedback-based gating. *Transp. Res. B* 46 (10), 1393–1403.
- Kickhöfer, B., Kern, J., 2015. Pricing local emission exposure of road traffic: An agent-based approach. *Transp. Res. D* 37, 14–28.
- Lam, T.C., Small, K.A., 2001. The value of time and reliability: Measurement from a value pricing experiment. *Transp. Res. E* 37 (2), 231–251, Advances in the Valuation of Travel Time Savings.
- Lamotte, R., Geroliminis, N., 2016. The morning commute in urban areas: Insights from theory and simulation. In: Presented in 95th Annual Meeting of the Transportation Research Board. Washington, D. C..
- Lamotte, R., Geroliminis, N., 2018. The morning commute in urban areas with heterogeneous trip lengths. *Transp. Res. B* 117, 794–810.
- Lamotte, R., Murashkin, M., Kouvelas, A., Geroliminis, N., 2018. Dynamic modeling of trip completion rate in urban areas with mfd representations. In: Presented in 97th Annual Meeting of the Transportation Research Board. Washington, D. C..
- Laval, J.A., 2023. Effect of the trip-length distribution on network-level traffic dynamics: Exact and statistical results. *Transp. Res. C* 148, 104036.
- Llorca, C., Moeckel, R., 2019. Effects of scaling down the population for agent-based traffic simulations. *Procedia Comput. Sci.* 151, 782–787, The 10th International Conference on Ambient Systems, Networks and Technologies (ANT 2019)/The 2nd International Conference on Emerging Data and Industry 4.0 (EDI40 2019)/Affiliated Workshops.
- Mahmassani, H., Williams, J., Herman, R., 1987. Performance of urban traffic networks. In: 10th International Symposium on Transportation and Traffic Theory. Cambridge, Massachusetts.
- Mariotte, G., Leclercq, L., Laval, J.A., 2017. Macroscopic urban dynamics: Analytical and numerical comparisons of existing models. *Transp. Res. B* 101, 245–267.
- Martínez, I., Jin, W.-L., 2021. On time-dependent trip distance distribution with for-hire vehicle trips in chicago. *Transp. Res. Rec.* 2675 (11), 915–934.
- Mcardle, G., Furey, E., Lawlor, A., Pozdnoukhov, A., 2014. Using digital footprints for a city-scale traffic simulation. *ACM Trans. Intell. Syst. Technol.* 5 (3).
- Nicolai, T., 2012. Using Matsim as a Travel Model Plug-in to Urbansim (Chapter for Deliverable D7.2). Technical report, Berlin Institute of Technology (TU Berlin).
- Noland, R.B., Polak, J.W., 2002. Travel time variability: A review of theoretical and empirical issues. *Transp. Rev.* 22 (1), 39–54.
- Sirmatel, I.I., Tsitsokas, D., Kouvelas, A., Geroliminis, N., 2021. Modeling, estimation, and control in large-scale urban road networks with remaining travel distance dynamics. *Transp. Res. C* 128, 103157.
- Small, K.A., Chu, X., 2003. Hypercongestion. *J. Transp. Econ. Policy* 37 (3), 319–352.
- Vickrey, W., 1991. Congestion in Midtown Manhattan in Relation to Marginal Cost Pricing. Unpublished notes, Columbia University.
- Vickrey, W., 2020. Congestion in midtown manhattan in relation to marginal cost pricing. *Econ. Transp.* 21.
- Yang, H., Ke, J., Ye, J., 2018. A universal distribution law of network detour ratios. *Transp. Res. C*.
- Zheng, N., Geroliminis, N., 2016. Modeling and optimization of multimodal urban networks with limited parking and dynamic pricing. *Transp. Res. B* 83, 36–58.
- Zheng, N., Waraich, R.A., Axhausen, K.W., Geroliminis, N., 2012. A dynamic cordon pricing scheme combining the macroscopic fundamental diagram and an agent-based traffic model. *Transp. Res. A* 46 (8), 1291–1303.