Simulation Modelling Practice and Theory 108 (2021) 102233

Contents lists available at ScienceDirect

Simulation Modelling Practice and Theory

journal homepage: www.elsevier.com/locate/simpat

Population downscaling in multi-agent transportation simulations: A review and case study

Golan Ben-Dor [a,*], Eran Ben-Elia [b], Itzhak Benenson [a]

[a] Department of Geography and Human Environment, Porter School of Environmental and Earth Studies, Tel Aviv University, Israel
[b] Department of Geography and Environmental Development, Ben-Gurion University of the Negev, Israel

ARTICLE INFO

Keywords:
Multi-agent systems
Transportation simulations
MATSim
Downscaling
Sioux Falls

ABSTRACT

Simulating the dynamics and evolution of metropolitan transportation systems serving millions of travelers remains a difficult task, beyond existing standard software's abilities. MATSim (Multi-Agent Transportation Simulation) is the only high-resolution spatially-explicit framework that allows intrinsic population downscaling - simulating the entire system's dynamics based only on a fraction k of the traveling population. Till now, the choice of k was dictated by hardware performance, and a common rule was not to downscale below 10%. We investigate downscaling in MATSim by comparing the aggregate and disaggregate statistics that describe the dynamics of car traffic in full-scaled and downscaled simulations of the Sioux Falls test case road network. Simulations with 25% or higher shares of the traveler population preserve all major urban traffic statistics. Within the 10–25% interval, downscaling becomes unstable for some of the statistics. For scenarios that are downscaled 10% and below, statistics can substantially deviate from the full-scale model. We further discuss the problems related to a multimodal transportation simulation's downscaling that also includes public transportation.

1. Introduction

Multi-Agent Systems (MAS) possess an inherent ability to represent the heterogeneity of agents' behavior and decision-making [13]. However, to understand complex systems' collective dynamics, the agent population should be sufficiently large, especially when spatially-explicit transportation infrastructures are considered. Interactions of travelers in urban space could well have far-reaching consequences if traveled distances are large and long trips are frequent.

Agent-Based Modelling (ABM) lies at the core of MAS practice and refers to a category of complex computer simulation models. The goal of ABM is to quantitatively and qualitatively evaluate the dynamics of a society of agents with highly heterogeneous individual behavior. A single agent is a discrete, autonomous entity with its own goals and behavior that can adapt to varying conditions by modifying its behavior. Given this flexibility, numerous ABM models were applied in different scientific fields such as biology, economics, social sciences, business, system engineering, and more [31].

With their inherent traffic and transportation problems, cities are the objects of inquiry of spatially-explicit MAS modeling. However, the larger the city is, the higher is the demand for model performance. This relation is fundamentally super-linear, with every agent, be it a traveler or a vehicle, interacting with many other agents when moving in space. Thus, studying transportation dynamics in metropolitan areas populated by millions of travelers becomes a significant challenge, and the modeler has to balance between the

* Corresponding author.
E-mail address: golanben@mail.tau.ac.il (G. Ben-Dor).

https://doi.org/10.1016/j.simpat.2020.102233
Received 21 January 2020; Received in revised form 17 November 2020; Accepted 19 November 2020
Available online 29 November 2020
1569-190X/© 2020 Elsevier B.V. All rights reserved.

limitations of the hardware and the inherent demand of representing the entire variety of agents traveling and interacting in a highly heterogeneous urban space.

Parry and Bithell [55] analyzed the state of large-scale agent-based modeling in social science, ecology, military, telecommunications, and other applications. They listed four major approaches to cope with systems that exceed the abilities of the current hardware. The first two approaches do not account for the specificity of the models:

- *Hardware upgrading* – Advantages: No re-programming is needed. Disadvantages: It can be costly while yet insufficient.
- *Parallelization* – Disadvantages: requires advanced computing skills for coping with the overhead of communication between the cores. Solutions are problem-dependent and can be yet inefficient. [28]. Advantages: If successful can be employed on the cloud with a practically unlimited number of cores.

Specifically, for transportation modeling, Dobler et al. [23] demonstrated that the parallel version of the MATSim software's performance is faster than the sequential one for a lower number of cores (less than 12). However, a further increase in the number of cores does not decrease the simulation runtime. Parallel/GPU-based processing pushes this threshold somewhat further but is also limited [62]. Multi-core versions are claimed by two large scale transportation ABMs - POLARIS [22] and SimMobility [2]. SimMobility employs the boost threads library that synchronizes access to the data shared by multiple threads and the message passing interface library, which simplifies its development. POLARIS's central framework is the discrete-event engine (DEVE), which is an API for creating agents and their desires. The DEVE organizes the agents to execute in the order requested by the developer with multiple threads managing the agents without any interference [3]. Aimsun is a commercial software of microscopic agent-based simulations that simulate car-following and lane-changing. Aimsun applies a multi-core CPU parallelism to increase simulator performance decreases up to 8 cores simulations runtime; afterward, there is diminishing returns to the performance [29]. SUMO a popular microscopic agent-based tool that simulates car-following and lane-changing is not yet fully parallelized and can manage up to 200,000 vehicles with a single core. However, some of its routines, like routing are already parallelized [46].

The complexity of agents' behavior and the associated non-linear phenomena in the ABM can be exhibited differently by the system as a whole and its parts [55]. In order to perform the full-scaled research inquiry with the available hardware, large agent-based systems should be *scaled,* that is, simulated based on a population smaller than in reality of agents who operate in the environment that is also reduced [56]. Parry and Bithell [55] point to two approaches to such scaling:

- *Super-Agents* – the use of aggregate agents, each representing several agents of the initial model. Advantages: Reduction of the model agents' population. Disadvantages: Demands careful re-programming of agents' behavior and interactions [39]. Super-agents are employed in *SimMobility*, where agents of similar type can be grouped together into a super-agent [2]
- *One-Represents-Several (ORS) - Modeling the whole system with a fraction of the agent population*. Advantages: No need to re-program agents' behavioral rules. Disadvantages: Demands careful scaling of the model infrastructure.

One-Represents-Several downscaling demands adjusting the model parameters to guarantee that the dynamics of the downscaled modeling system remain the same as those of the full-scaled one [14]. Of the several spatially-explicit MAS traffic simulators, we are aware of, only the Multi-Agent Transportation Simulation (MATSim) software possesses an intrinsic downscaling procedure [36]. Downscaling in MATSim demands straightforward adjustments of the network parameters and is widely applied in MATSim application [[6,9,14,15,17,25,27,44,47,65,69]. However, studies of the *effects* of downscaling on model dynamics are yet limited in scope. Therefore, this study aims at closing this gap.

Recently, MATSim was compared to 83 different ABM in different fields and described as one of the most powerful and developed simulation environments where the number of agents is only constrained by hardware ability. MATSim was the only transportation ABM ranked at the top level ("extreme-scale") due to its ability to simulate millions of agents over vast urban networks and due to their adaptive behavior [1].

As the spatial extent of a large-scale ABM is always problematic, eventually, there is always some city that is too large to be fully simulated. This limitation is especially restricting when traffic and transportation simulations are considered since traffic cannot be limited to some parts of the city. In this case, downscaling can be considered i.e. simulating traffic of an entire spatial system with fewer agents. For example, downscaling can be implemented by assuming that one agent represents several agents in the full-scale model. To match the full-scale model, the downscaled one demands adjustment of the model parameters. In this paper, we investigate the solution implemented in MATSim – reduction of road capacity proportionally to the population sample size. The consistency between the emergent agent behavior and the resulting dynamics of the downscaled model compared to the full-scale model, are hardly ever investigated. As a top-end simulation model, MATSim is, therefore, an excellent candidate to investigate this research question, which also frames the motivation for our paper.

Accordingly, we aim to fill this gap between ABM practice and simulation downscaling and systematically investigate the consequences of downscaling in MATSim and, eventually, all other large scale ABMs. Our testbed was a well-documented dataset of the city of Sioux Falls, South Dakota, from 2014 [21]. In what follows, Section 2 presents an overview of Agent-Based Transportation Models. Section 3 presents the fundamentals of MATSim's downscaling procedure, including a review of relevant studies. Section 4 presents the methodology and investigated scenarios; Section 5 presents the obtained results; Section 6 shows a performance analysis of simulation runtime and Section 7, presents the conclusions and outlines pathways for future work. In transportation studies, ABM is used by MAS interchangeably, while the use of MAS often presumes that travel demand is represented by a detailed activity plan for each traveler [50]. We followed this practice in the paper.

2. Overview of agent-based transportation models

In an ABM, each agent is modeled individually, accounting for its unique properties, like age, residence, socioeconomic state, location. Transportation ABMs focus on the representation of travel decisions of individuals as they move about in space in and time [43]. Given their disaggregated nature, ABM emerged to capture the changes in individual travel behavior regarding local changes in land-use, travelers' activity schedules and mode choice, dynamic road pricing, as well as transportation innovations such as ride-sharing, mobility-as-a-service (MaaS), automated vehicles, and more [16]. Kagho et al. [43], specify three components of ABMs essential for transportation planning: (1) Physical environment, which includes a transportation network, land use, and other transportation supply elements, (2) Agents - vehicles and travelers who use these vehicles and (3) Agents' behavioral rules that represent the outcomes of their interactions with the physical environment and with other agents, e.g., choice of mode, activity (destination) or departure time. In addition, five major challenges for transportation ABMs are mentioned:

1 *Input data quality:* errors in synthetic/real population data, data collection process, data availability, etc., can threaten the validity of the simulation.
2 *Cost of Computation:* ABM usually demands a large number of agents and, consequently, computing resources. For example, a MATSim simulation run for the city of Paris, with a sample population of 10%, takes 5 h on a modern computing cluster [34].
3 *Transparency*: It is crucial to understand the "black box" aspects of an ABM to understand the resulting outputs as by nature, ABM are complex systems with a vast number of parameters.
4 *Validation* - It is important to show that the model can replicate reality in respect of real-world data on traffic counts, mode shares, trip distribution, and more.
5 *Reproducibility* - The complex nature of the ABM studies makes them harder to reproduce. We are not aware of any traffic multi-agent simulations that once investigated, were reproduced again by the same or a different research team.

The traffic dynamics are simulated in an ABM using the mobility simulator that formalizes the modeler's view of agents' interactions. The mobility simulator is at the core of the ABM representing traffic either as a dynamic queue e.g. MATSim [36], and more recently, POLARIS [22] or as a partially or fully implement high-resolution car-following model e.g. SUMO [46], Cube [68] and SimMobility ([8,2].

Unequivocally, the confirmed leader among the open-source systems is MATSim - an event-driven multi-agent MAS environment that aims at modeling the co-evolution of travelers and system-wide traffic conditions by imitation of individual travel choices and behavior [36].[1] MATSim agents adapt through reinforced learning and self-correction [49]. MATSim was developed for executing large-scale transportation simulations, and, basically, MATSim is capable of performing spatially-explicit simulations of an unlimited number of travelers [1]. Critically for our study, MATSim is oriented towards downscaling by design and practically, in all real-world MATSim applications downscaling, is implicitly employed when simulating urban traffic with several million agents.

The fundamental principle of downscaling in MATSim is "One Represents Several" (ORS) i.e., every agent is considered as representing several other agents, and model parameters are modified, accordingly to preserve the dynamics of the full-scale model. We define this as the ORS downscaling concept. Since MATSim dynamics are defined by agent adaptations to changes in the transportation network, it is not at all evident that the ORS concept, when translated into the rules governing the variation of MATSim's parameters, will guarantee, irrespective of k, that the downscaled simulations will replicate the dynamics of the full-scale model. Notably, the queue-based view of traffic flows and the ORS downscaling option were recently added to the POLARIS mode [22].

The inherent MATSim approach to downscale large-scale transportation simulations makes it an excellent candidate for investigating its effects on the integrity of this process. This could well be beneficial both to the ever increasing audience of MATSim's users as well as developers of other simulation environments. In our study we touch upon four of the MAS challenges mentioned by Kagho et al. [43]: the cost of computation in terms of running time; transparency, by describing the process, the data, and the parameters used for the study and reproducibility by using the open data of the well-known Sioux Falls scenarios. Our study offers general guidelines on how to conduct the ORS downscaling principle that is likely to be employed in every large mesoscopic scale transportation MAS model in the future.

3. Downscaling in MATSim

Population downscaling is brought about by the demands of large-scale transportation simulations. Regardless of the power of computers, simulating agents' adaptation to the varying traffic conditions requires re-estimating model's and agents' parameters at every time step. Consequently, computation time increases at least linearly with the increase in the number of agents, irrespectively of serial or parallel computing [23]. Transportation simulations demand many runs, with different sets of parameters and repeated runs for the same set of parameters, in order to account for the inherent stochastic variation of the traffic flows and agents' behavior. In practice, transportation modeling demands performing at least a "few" simulations per day i.e., every simulation run must be finished within a few hours. With standard up-to-date hardware, when the number of modeled agents is several hundred thousand, a MATSim run requires several hours to complete. As a result, almost all MATSim applications of large-scale metropolitan areas are downscaled to

[1] Google Scholar search on 21.4.2020 resulted in 19 hits of SimMobility, 214 of POLARIS, and 4,720 of MATSim.

200–500 K of agents at most that is, to 10–15% of the original population size.

3.1. The MATSim simulation scheme

In order to simulate millions of agents MATSim considers traffic flows at the mesoscale level and does not simulate car-following explicitly. The traffic simulation module of MATSim, QSim, employs a *First-In-First-Out* (FIFO) approach for simulating how vehicles traverse road links: A link is considered as a gate and a vehicle entering it at one end is added to the tail of the vehicle queue building up at the other end. The vehicle remains in this queue until it reaches the queue head (Fig. 1). Thus, no vehicle can enter a link if the *storage_capacity* - the maximal number of vehicles that can be physically allocated on a link, is exhausted.

The traversal time of a link is dependent on its current traffic conditions [36] that, in turn, is defined by the *flow_capacity* – the maximal number of cars that exit a link per time unit. Since both storage and flow capacity are limited, vehicles queueing on a link of a limited capacity generate traffic congestion.

Fig. 2 illustrates how *flow_capacity* and *storage_capacity* influence the traffic flow. Let us consider three connected links with the following capacities: link1: $storage\_capacity = 3\ car$, $flow\_capacity = 2\ car/min$ link2: $storage\_capacity = 6\ car$, $flow\_capacity = 2\ car/min$ link3: $storage\_capacity = 3\ car$, $flow\_capacity = 1\ car/min$

The following events will happen between $t = 0$ and $t = 1$ (looking from left to right): since *flow_capacity* of link1 is 2 cars/min, two cars will try to exit link1 and proceed to link2; since *flow_capacity* of link2 is also 2 cars/min, two cars will try to exit link2 and proceed to link 3; and, since *flow_capacity* of link3 is 1, one car will exit it. Looking from right to left: Possible transitions from link2 to link3 would not be fully executed since only one unit of the storage capacity on link3 will be emptied. At the same time, the transitions from link1 to link2 will be executed in full since, after one car, the unexploited *storage_capacity* of link2, that would transfer to link3, will be equal to 3 cars.

The queue-based approach substantially improves QSim's performance in comparison to other microsimulations that account for more explicit vehicle interactions [1]. In particular, the QSim approach prohibits the volume-to-capacity ratio (V/C) to exceed 1. In contrast, the traditional static traffic assignment models that use an impedance function, permit conditions where V/C is above 1 as applied in software environments such as EMME [27], TransCad [41], or Cube 68.

Agents in MATSim are characterized by their spatiotemporally-explicit activities, also referred to as plans. These plans are an input for MATSim, and this "Activity-Based" approach is exploited in several other models (e.g. ALBATROSS - [4]; FEATHERS - [10]; and CEDMAp, [63]) that aim at integration of agents' mobility at different levels of spatio-temporal aggregation. Plans are implemented with the MOBSim and Scoring simulation modules. MOBSim executes the trips made by the agents and enables adaptation of the agents to evolving traffic conditions. To simulate agents' adaptation, MOBSim selects some of them randomly for possible innovations: change of route, mode, or departure time. The performance of each agent is *scored* in the Scoring module using pre-established utility functions whereby agents for whom the innovation improves the score accept the change, while agents for whom the innovation decreases the score reject it. Changes in agents' behavior entail traffic changes, and the underlying assumption of MATSim is that, in a series of iterations of the same travel day, this process of system-travelers' co-adaptation will converge to a user equilibrium (UE), where on that day, no agent can improve its score by changing its behavior. This process, known as the MATSim *daily loop* (Fig. 3) is reiterated until UE is achieved.

3.2. MATSim downscaling rules

ORS downscaling in MATSim is defined, as customary in the MATSim literature, by the *downscaling ratio* k, the fraction of the traveler population that is considered in the downscaled simulation. We will interchangeably call the downscaled scenario "k-downscaled" and "k*100%-downscaled", e.g., "0.1-downscaled" or "10%-scenario". For example, in a 10%-downscaled model, 10% of the real population of the region is considered. In these terms, the full-scale model that operates with the entire population of the area is 100%-downscaled.

To implement downscaling in MATSim, the flow and storage capacities defined for each link are adjusted proportionally to k. In a full-scaled model, these capacities are established as:

$$flow\_capacity\_full = \frac{capacity\_value\_of\_link}{capacity\_period\_of\_network} \tag{1}$$

$$storage\_capacity\_full = \frac{length\_of\_road\_link * number\_of\_lanes}{effective\_cell\_size} \tag{2}$$

The *capacity_value_of_link* in Eq. (1) is established for all links of a given network applying standard traffic engineering methods of and calculated (per hour) according to the Highway Capacity Manual [75]. Since, traffic changes in MATSim, are considered per

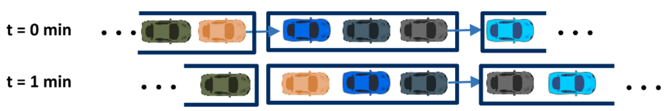

Fig. 1. Illustration of the QSim FIFO rule of vehicle advancement on one link and between two consecutive links.

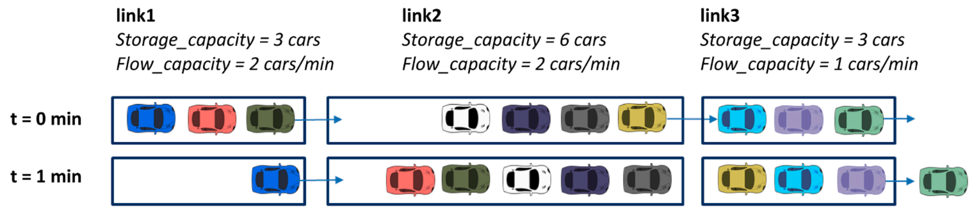

Fig. 2. Illustration of the effects of MATSim link storage capacity and flow capacity on the simulated traffic flow.

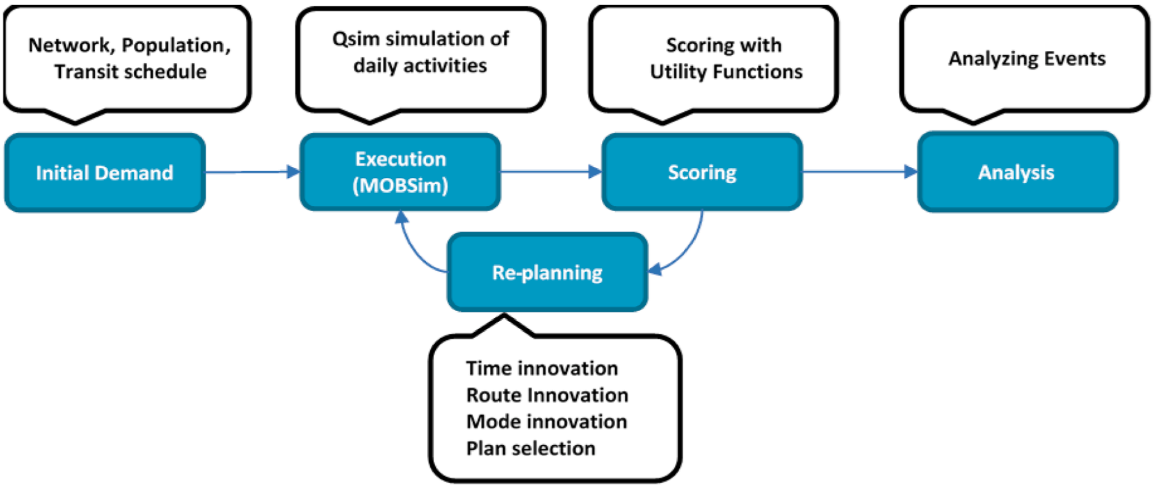

Fig. 3. Flowchart of the MATSim daily loop, adapted from [36].

second, the *capacity_value_of_link* is normalized using *capacity_period_of_network* = 3600. The *effective_cell_size* in Eq. (2) is the minimal length of a road lane a vehicle occupies calculated in MATSim as the length of the vehicle (in meters) multiplied by 1.5. The vast majority of MATSim applications set the *effective_cell_size* equal to 7.5 m, and we adopted this value as well.

When downscaling, the flow and storage capacities are adjusted relative to k. The adjustment of the flow capacity is linear, whereas that of the storage capacity is non-linear:

$$flow\_capacity\_downscaled = flow\_capacity\_full * k \quad (3)$$

$$storage\_capacity\_downscaled = full\_storage\_capacity * k^{0.75} \quad (4)$$

For example, in a 10%-downscaled model, according to (3), *full_flow_capacity* is multiplied by $k = 0.1$ and, according to (4), *full_storage_capacity* is multiplied by $k^{0.75} = (0.1)^{0.75} = 0.1778$.

The exponent of 0.75 in Eq. (4) was proposed by Nicolai [51], to avoid network breakdowns related to excessively congested links [57] and confirmed by Llorca & Moeckel [20].

In a full-scale simulation MATSim randomizes the order which agents are considered for performing their daily plans and which fraction of the agents are selected for innovations, in each iteration. This randomization is controlled by a *global random seed* set at the beginning of the full-scale model run. In a downscaled simulation, besides adjusting storage and flow capacities, MATSim randomly selects by activating a *downscale random seed,* k*100% agents that represent k agents in the full-scale model [36].

3.3. Review: downscaling in MATSim applications

MATSim studies extensively apply downscaling (see Table 1). To project the results obtained in the downscaled simulation to the full-scale model, one has to be sure that these are unbiased i.e. over several runs the average value of any output statistic is similar to that of the same statistic in the full-scale model. In addition, even for unbiased statistics, their variation should be known in order to estimate the number of repetitions necessary to estimate the statistic with the necessary precision. In this section, we shortly review the motivations for performing downscaling and the choice of the downscaling ratio k, and present the state-of-the-art regarding output statistics' variance in downscaled simulations.

3.3.1. Motivations for downscaling and the choice of the ratio k

As shown in Table 1, the values of the downscale ratio k vary between 1% and 25% and in 16 out of 31 studies the choice of k is

Table 1
MATSim downscaling applicatations and their attributes.

#	Authors	Year	Scenario	Research goal	Total Pop [100%]	Downscaling justification	values of *k* [%]	Network size [Nodes/ Links]	PT
1	Illenberger et al [40]	2007	Berlin, Germany	Perfection of re-planning procedure	1700,000	N/A	10%	N/A	N/A
2	Balmer et al [6]	2008	Greater Zürich, Switzerland	Model performance	1816,930	Common choice	10%	24,180/ 60,492	No
3	Horni et al [37]	2009	Switzerland	Perfection of travelers' choice model	2284,010	N/A	10%	N/A	N/A
4	Dobler [23]	2010	Berlin, Germany	Parallelization of MATSim	1600,000	N/A	1%	11,000/ 28,000	N/A
5	Dobler [23]	2010	Canton Zurich, Switzerland	Parallelization of MATSim	400,00	N/A	25%	73,000/ 163,000	N/A
6	Gao et al [27]	2010	Greater Toronto, Canada	Bridging between EMME/2 and MATSim	5570,000	Common choice	5%	16,337/ 40,549	N/A
7	Bekhor et al [9]	2011	Tel-Aviv, Israel	Traffic simulation	3200,000	Performance	10%	7879/ 17,118	N/A
8	Horni et al [35]	2011	Zurich city center, Switzerland	Study of output variance	680,000	Performance	10%	24,180 / 60,492	N/A
9	Erath et al [25]	2012	Singapore	National traffic modeling	4300,000	Performance	25%	12,420/ 26,972	Yes
10	Waraich et al [69]	2012	Zurich, Switzerland	Parking search	1800,000	Common choice	1%, 10%	1 million	No
11	Kickhöfer et al [45]	2013	Munich, Germany	Effect of travel costs on emissions	2093,530	Performance	10%	17,888/ 41,942	No
12	Röder et al [58]	2013	Brusseles, Belgium	Cordon pricing for a highway	860,214	N/A	1%	10,861/ 19,830	No
13	Dobler et al [24]	2014	Canton Zurich, Switzerland	Bicycle and pedestrian traffic	1800,000	N/A	25%	24,000/ 60,000	Yes
14	Hülsmann et al [38]	2014	Munich, Germany	Air pollution assessment	2093,530	Performance	1%	17,888/ 41,942	No
15	Nicolai et al [52]	2014	Zurich, Switzerland	Accessibility analysis in Zurich	336,290	Performance	10%	24,180/ 60,492	No
16	Bösch et al [18]	2015	Zurich, Switzerland	Extension of MATSim multimodal capabilities	N/A	Performance	0.1%,1%, 10%	N/A	Yes
17	Simoni et al [64]	2015	Zurich city center Switzerland	Road pricing	1800,000	N/A	10%	550/1224	No
18	Ziemke et al [74]	2015	Berlin, Germany	Population model calibration	1105,037	N/A	1%	11,345/ 24,335	No
19	Balać et al [5]	2016	Zurich, Switzerland	Activity re-planning	N/A	Performance	1%	N/A	N/A
20	Bischoff et al [14]	2016	Berlin, Germany	Introduction of automated taxi	N/A	Performance	10%	N/A	N/A
21	Bösch et al [19]	2016	Switzerland	National scenario	N/A	N/A	1%, 10%,,100%	N/A	Yes
22	Fourie [26]	2016	Zurich, Switzerland	Perfection of the QSim	2000,000	N/A	10%	N/A	No
23	McArdle et al [48]	2016	Dublin County, Ireland	Perfection of travelers' choice model	600,000	Performance	25%	N/A	No
24	Hülsmann קא 38]]	2016	Munich,, Germany	Detecting air pollution hotspots	2093,530	Performance	1%	17,888/ 41,942	No
25	Kickhofer et al [44]	2016	Santiago, Chile	Traffic simulation	6650,000	Performance	65%	N/A	No
26	Ordóñez [53]	2016	Singapore	Effects of iteration step exceeding one day	3742,500	N/A	1%	N/A	No
27	Saadi et al [61]	2016	Liège, Belgium	River flood impact on transportation	1094,791	Performance	1%	24,372/ 55,959	No
28	Hörl [33]	2017	Île-de-France, France	Introduction of automated vehicles	1000,000	Performance	10%	N/A	Yes
29	Maciejewski et al [47]	2017	Berlin, Germany	Congestion of automated taxis	N/A	Common choice	10%	N/A	N/A
30	Bassolas et al [7]	2018	Barcelona, Spain	Traffic simulation	2300,000	Performance	10%	9217/17,690	Yes
31	Tchervenkov et al. [65]	2018	Switzerland	Externalities of mobility (emissions, congestion, noise)	N/A	Common choice	10%	N/A	N/A

motivated by the desired computing time, while in 7 other papers by common agreement

3.3.2. Variance of output statistics in downscaled simulations

The stochastic variation in outputs of a full MATSim scenario, as determined by the *global random seed*, is low. Hemdan et al., [30] estimated the Coefficient of Variation (CV - the ratio between the standard deviation and the mean), of several global statistics, based on three repetitions of the full MATSim Sioux Falls simulation (84,110 agents). They report a CV of almost 0% for the modal shares (car, walk, transit); 0.08%−0.60% for the average travel time; 0.25% for the average morning peak traffic flow (veh/h). Only the average network morning peak density (veh/km) had the largest variability with CV = 2.13%.

Intuitively, the output statistic's variance in a downscaled simulation should be higher compared to the full-scale model. To the best of our knowledge, the studies that verify this fact are scarce. Zhuge & Shao [71], Zhuge et al. [72] and Zhuge et al. [73] investigated a unimodal car traffic simulation for Baoding, China and compared the daily averages of agent scores, travel distance, V/C, and travel speeds and compared the 20%-runs to 10%, 5%, and 1%-runs. They report that downscaling from 20% to 10% had no influence on these four statistics. However, downscaling of 5% and 1% resulted in an increase of the average travel distance by 30% and biased average scores. At the same time, the city-wide average V/C remained the same for all k. Llorca & Moeckel, [20] investigated the effect of downscaling on the distribution of agent travel times and daily scores in a traffic simulation of Munich. They performed one simulation run over the values of k between 0.01 to 1 and demonstrated that travel time distributions for k within the range of (0.05, 1] remained similar to that of $k = 1$ However, for $k < 0.05$ the deviation from the travel time of the full-scale model became substantial. The average daily score slightly increased with the decrease in k and for $k = 0.05$ was ~5% higher than that obtained for $k = 1$. They thus concluded that the lower threshold for k is 5%. Horni et al. [35] estimated the variance of the hourly traffic volume in 30 repetitions of a 10%-scenario for the City of Zurich and demonstrated that the CV of the hourly link volumes varied up to 10%−20%.

While the main motivation of performing downscaling is to reduce computing time, studies investigating its dependency on k, for a single-core or multi-core versions of MATSim, are very few. Llorca & Moeckel [20] showed in the traffic simulation of Munich that the runtime increased linearly with the increase in k. Dobler et al., [23] demonstrated that computation time was not reduced when parallelizing the traffic scenario of Zurich Canton, Switzerland for $k = 1$ and $k = 0.25$.

3.3.3. Downscaling in a mixed traffic model

Mixed traffic includes cars and public transportation (PT) and possibly other modes. In this case, adjusting the capacities according to the ORS downscaling principle and equations (3) – (4) is insufficient. The problem lies in the adjustment of the capacity, size and the number of PT vehicles. The capacity c_v that defines maximal vehicle occupation and the length l_v that defines the share of the storage and flow capacities of PT vehicles, can essentially be adjusted to $c_v{*}k$ and $l_v{*}k$. However, the overall number of PT vehicles must be preserved in order to preserve the waiting time of PT passengers in the downscaled model. While when converging to UE, car driving agents can change route to avoid congestion, public transport must always follow fixed routes and schedules. In downscaled simulations the interaction between unscaled private vehicles and scaled PT vehicles could well result in longer queues and congestion becoming worse along PT routes.

The simplest solution in this case is coding PT lines on a separate network while preserving PT vehicle length and reducing their passenger capacity proportionally to k [44]. However, this solution is unfit in cities where cars and PT vehicles drive on and compete over the same road lane space. For this case we recently suggested to modify the queueing model of QSim allowing PT vehicles to enter an exhausted link by linearly adjusting their capacity and length, thus allowing PT vehicles in mixed traffic to override the link flow capacity restrictions [11]. Unequivocally, downscaling with mixed traffic is under-investigated and demands more attention. However, in this paper we focus essentially on car traffic.

To summarize, while downscaling in MATSim is a common practice, to the best of our knowledge even for car traffic scenarios, the consequences have never been investigated in depth. We consider three open questions that demand systematic research necessary for establishing the lower threshold of the downscaling ratio k: (1) What is the bias of the statistics in downscaled simulations? (2) What is their variance? (3) What are the tradeoffs, in downscaled simulations, between the bias and variance of the output statistics as well as the computational performance?

In this paper we aim to investigate the impacts of downscaling on simulation output statistics based on the well-known Sioux Falls test case scenario. While MATSim can contend with multimodal transportation networks, given the need for consistent comparisons of different downscaled simulations, in this paper, we focused only on unimodal scenarios with private cars.

4. Methodology

4.1. The Sioux falls scenario

The road network of the City of Sioux Falls (SF), South Dakota, is popular test case long used in transportation planning investigations and was adapted to MATSim by Chakirov & Fourie [21]. The scenario has been widely applied since 1970 to the present day in different case studies from dedicated bus lanes [12], road pricing optimization [42], and automated vehicles [70] to unmanned airborne drones [60]. Two versions of the SF road network exist – the 2014 simplified version that contains only major roads and the 2017 version containing all links and junctions in the city [32]. In this study, we use the 2014 network. This network was selected as it is well studied, documented and its simplified topology allows easier understanding of its emergent traffic patterns and dynamics.

The 2014 network consists of 334 links and 282 junctions (Fig. 4). The road links are of two types: *Urban Road* – two-lane links with a flow capacity of 800–1000 cars per lane per hour, and *Highway* - three-lane links with a flow capacity of 1700–1900 cars per lane per hour. The storage capacity of the road links in Eq. (2) was estimated by [21] according to the attributes of the SF road network – link length and the number of lanes.

Uncommon to many MATSim studies, travel plans are available for 100% of the population in the SF scenario. In addition, this scenario is small enough (population and network) to simulate with the full- scale model with standard computer hardware. The daily plans were constructed based on a micro-census and land-use data and include three types of activities: staying at home, being at work, and secondary activities (ibid). The total number of travelers (agents) is 84,110 characterized by age, gender, and car ownership. However, with this number of agents, the level of traffic congestion observed on the road network, is relatively low. To generate higher congestion, we cloned each Sioux Falls agent, including their own travel plans and reconsidered 168,220 agents as the full i.e. $k = 1$ scale model. The initial modal split in SF is, roughly, 78% car use and 22% transit. However, as noted before, in our analysis, we assumed that all agents drive cars.

We investigated downscaling by varying k between 1 and 0.01, i.e. we consider scenarios with the number of agents varying between 168,220 and 1682. To downscale the simulation, we normalized *flow_capacity_downscaled* and *storage_capacity_downscaled* according to Eqs. (1) and (2) above. To estimate the variance of the model outputs, each scenario was repeated 10 times with a different random seed. Repetitions of the full-scale model were run each with a different global seed, while those of the downscaled models with different random seeds. In all scenarios, we applied the same scoring functions as proposed by Chakirov & Fourie [21]. All simulations were run with 3000 iterations, a sufficient number for convergence to UE. The results in this paper are based on the last iteration of each simulation (3000). Plan innovations - re-routing and change of departure time were applied in each run, for 1% of randomly selected agents from the 1st up to the 2700th iteration. During these iterations, the agents store in memory the five best plans and whenever a new plan was adopted, the worst one is forgotten. In the remaining 300 iterations, innovations were switched off.

4.2. Convergence to user equilibrium

Fig. 5 presents the dynamics of the executed scores of all agents averaged over 10 runs in the full-scale model. In the majority of the runs convergence to UE happened around the 400th iteration with some sporadic deviations of the scores in some runs up to around the 800th iteration, likely caused by the transfer from one UE to another. Further on, the system reaches a final UE where the average score remains stable (STD ≈ 0) in all 10 runs.

In the downscaled simulations, the number of iterations necessary to reach the final UE state was very similar to the full scale runs varying between 600 and 800. We thus assert that ~800 iterations are sufficient for convergence to the final UE in the SF scenario. Furthermore, we can conclude that the number of iterations necessary to reach the system UE does not depend on k and can be estimated in one "long" downscaled simulation.

4.3. Choice of the simulation output statistics

Conceptually, some statistics of MATSim outputs should automatically scale i.e. linearly change with the decrease of k, while others should remain constant. Examples of potentially scalable statistics are the number of departures or the traffic volumes per time unit that should change linearly with the decrease of k. Examples of non-scalable statistics are the average of agents' scores or the average travel distance that should remain the same regardless of k. When comparing the output statistics of downscaled and full-scaled models, we multiply the scalable statistics by $1/k$, whereas non-scalable ones remain as is.

In what follows, we investigated the effects of downscaling, at different spatiotemporal resolution levels, on four non-scalable

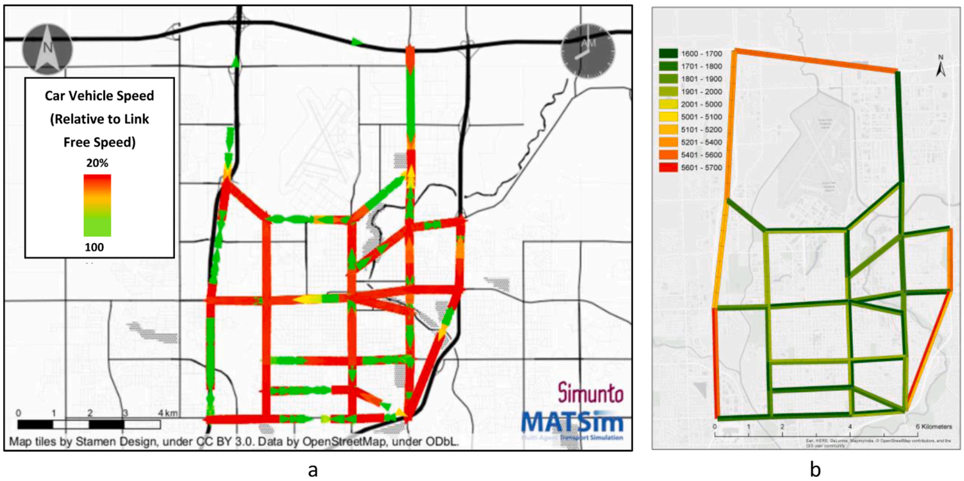

Fig. 4. Sioux Falls 2014 road network and (a) morning peak traffic flows at 08:00 for population of 168,220 agents; (b) flow capacities on links (vehicles/hour).

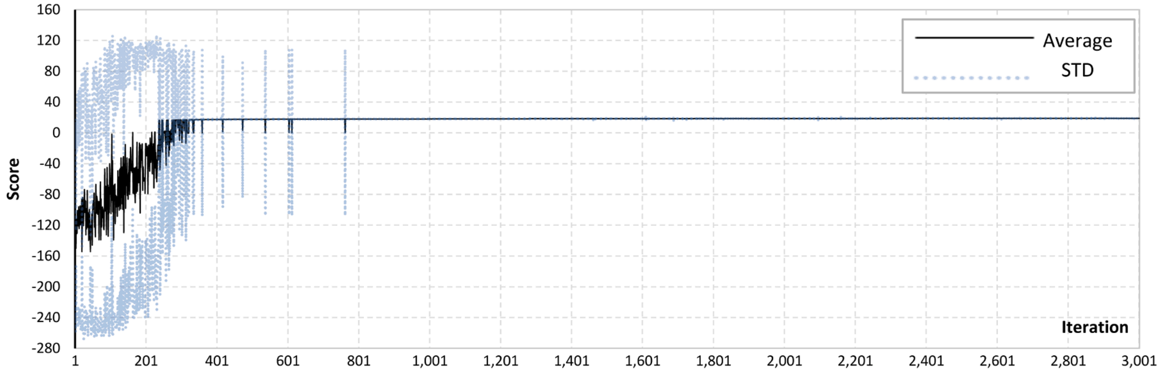

Fig. 5. Average (black) and average ± STD (gray) of the executed scores in the full-scale model (10 repetitions), by iterations.

statistics: (1) **executed scores** (reflecting the co-adaptation level of agents), (2) **travel distance**, (3) **trip duration** and (4) **volume-to-capacity (V/C) ratio**; and two scalable statistics: (1) the **hourly number of car departures** (number of agent departures) and (2) **traffic volumes** (number of traveling agents) (Table 2). The values of the last two statistics obtained in the downscaled runs are then multiplied by $1/k$. Each of the statistics is considered per day and per hour.

The downscaling effects are examined at the full aggregation level, i.e., spatially, over the entire network and temporally, over the entire day, as well as at higher spatiotemporal resolutions - hourly or aggregated by peak/off-peak periods of the day. The spatial distribution of traffic flows is investigated based on the hourly car volumes, aggregated over the entire network and disaggregated, by individual links by the peak/off-peak periods of the day.

The peak hour is defined as an hour of the maximal traffic volumes (M) and the *Peak Hour Interval* (PHI) is defined between the moment when the number of cars on the network exceeds $M/2$ and until it declines back again to $M/2$ [59]. Based on the full scenario, morning PHI was set to [06:30, 11:00], and evening PHI to [17:00, 19:45] (Fig. 6). The period between the morning and evening peaks (11:00–17:00) was considered as an off-peak period. We did not include night traffic, after 19:45 and before 5:00 into the analysis.

4.4. Measuring the similarity between downscaled and full-scale runs

In what follows we denote output statistic c obtained in a k-downscaled run i, as c_i^k, its average and standard deviation as m_c^k and s_c^k and coefficient of variation as $CV_c^k = \frac{s_c^k}{m_c^k}$. The base line for comparison between the full-scale and downscaled simulation runs was a set of output statistics averaged over 10 repetitions of the full-scale model: average m_c^1, standard deviations s_c^1 and coefficients of variation CV_c^1.

To compare between the downscaled and full-scaled runs we estimated the relative bias b_i^k between the statistic c_i^k of the downscaled run i and the full-scale run average m_c^1:

$$b_i^k = \frac{c_i^k - m_c^1}{m_c^1} \tag{5}$$

Below, we present the relative bias averaged over 10 repetitions of the downscaled scenario - m_b^k and its standard deviation -s_b^k. We consider the dependence of a certain statistic on k, varying k between $k = 1$ to $k = 0.01$ (in decreasing direction). Output statistic is therefore regarded as a robust to k-downscaling if its average relative bias m_b^k is close to 0 and, to account for the decrease in population size, its STD s_b^k is close to $CV_c^1/\sqrt{k}$.

4.5. Excluding outlier simulation runs

How certain can we be that, given k, the statistics of two downscaled simulation runs will be similar? Intuitively, in different runs, when $k < 1$, the population of agents is represented by different samples. These agents follow their specific plans, and thus, the model

Table 2
Investigated statistics of the simulation outputs.

Output statistic	Description	Category
Agent executed scores	Daily average	Non-scalable
Travel distance	Hourly average travel distance	Non-scalable
Trip duration	Hourly average travel time	Non-scalable
Volume-to-Capacity (V/C) ratio	volume (V) to flow capacity (C) ratio per hour	Non-scalable
Hourly number of car departures	Number of agent departures	Scalable
5-minute traffic volumes	Number of traveling agents	Scalable

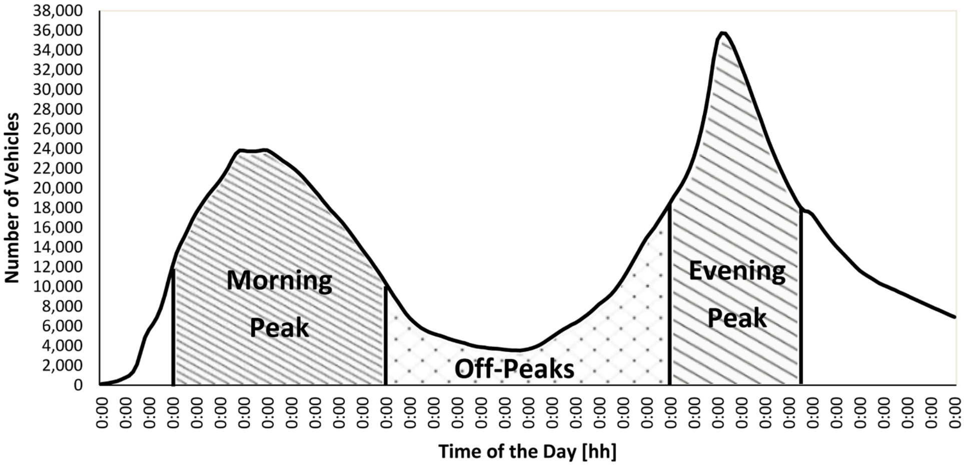

Fig. 6. Traffic volumes per hour, and morning and evening Peak Hour Intervals (PHI).

could possibly converge each time to a different UE. Even for $k = 1$, despite converging to the same UE (Fig. 5), similar average scores can represent different distributions of scores of the population and could entail differences in other statistics as well.

To understand whether different simulation runs converged to the same UE and that the average and STD that characterize the distribution of the output statistics were reliable we searched for possible outliers by analyzing the distribution of the average scores in a UE state, for each k separately. We classified a run as an outlier if the average score in the UE state was beyond Tukey's fences [67] constructed based on all 10 runs for a given k:

$$[Q_1 - w(Q_3 - Q_1),\ Q_3 + w(Q_3 - Q_1)] \tag{6}$$

where $w = 2.5$, and Q_1 and Q_3 are the 25th and 75th percentiles, respectively.

Note that for the normal distribution of scores, the probability to obtain a score outside Tukey's fences with $w = 2.5$ is less than 5×10^{-5}.

There were no outliers within the runs for $k = 1$, while 3 out of 100 simulation runs with $k < 1$ had converged to UE with average scores at the border or beyond Tukey's fences with $w = 2.5$: One for $k = 0.5$ ($w = -2.48$); one for $k = 0.4$ ($w = -3.21$); and one for $k = 0.05$ ($w = 2.84$). Some of the statistics for these three runs were also beyond Tukey's fence with $w = 2.5$. These runs were naturally excluded from the analysis, and were substituted by an additional run for the same value of k. After excluding the outliers, the STD of the relative bias, for all statistics, was $CV_c^1/\sqrt{k}$ or lower.

5. Results

5.1. Full-scale model output statistics

The values of m_c^1, s_c^1 and CV_c^1 for the selected statistics in the full-scale runs are presented in Table 3 One can observe that the values of CV_c^1 never exceed 5% and remain below 2.5% for most of the statistics.

The output statistics of the full-scale runs by hours of the day and by links are presented in Figs. 7 and 8. Fig. 7 presents the mean and STD of the hourly number of departures and traffic volumes and the CV of these statistics, which remained, always, below 10%.

The spatial patterns of traffic congestion was evaluated using the volume-to-capacity ratio (V/C) where $V/C \geq 1$ indicates extreme

Table 3
The average, STD, and CV of output statistics for the full-scaled ($k = 1$) SF simulation estimated based on 10 repetitions with different global seeds.

Output Statistic	Average	STD	CV
Executed Scores	18.82	0.04	0.19%
Travel Distance (km)	5.22	0.0037	0.07%
Trip duration (mm:ss), Morning Peak (06:30 – 11:00)	52:33	02:23	4.54%
Trip duration (mm:ss), Evening Peak (17:00 – 19:45)	57:58	00:46	1.31%
Trip duration (mm:ss), Daily Off-Peak (11:00 – 17:00)	35:39	00:47	2.20%
Hourly Departures, Morning Peak (06:30 – 11:00)	20,680	114	0.55%
Hourly Departures, Evening Peak (17:00 – 19:45)	30,055	91	0.30%
Hourly Departures, Daily Off-Peak (11:00 – 17:00)	20,480	60	0.29%
5-minute traffic volumes, Morning Peak (06:30 – 11:00)	18,591	891	4.79%
5-minute traffic volumes, Evening Peak (17:00 – 19:45)	26,122	332	1.27%
5-minute traffic volumes, Day Off-Peak (11:00 – 17:00)	7145	154	2.15%

levels of congestion and the existence of bottlenecks [[66]] . Fig. 8 presents a map of the average hourly value of V/C during the morning peak, 6:30 – 11:00 and relative bias b_i^1 that remains below 2% for all most all levels of V/C. This variation of V/C is similar to that reported by [54] for Santiago, Chile.

5.2. Results of downscaled models

The effects of downscaling on the statistics of different outputs vary. Some of the statistics for $k = 1$ are closely repeated in every downscaled runs down even at very low values of k, whereas some needed to be averaged to be closer to the statistics of the full-scaled models, while some deviated even for relatively high values of k. We present the results first for aggregate statistics followed by disaggregate ones.

5.2.1. Aggregate statistics

We first present the results obtained for three non-scalable statistics (executed scores, travel distance and trip duration) followed by two scalable statistics (car departures and traffic volumes).

5.2.1.1. Non-scalable statistics. Fig. 9 shows two non-scalable statistics for which the relative bias remained below 1% even for low values of k. One is the average daily executed scores and the other the daily traveled distance. The relative bias of both statistics was systematically positive, but very low, and remained below 1% up to $k = 0.05$. The STD of the bias remained below $CV_c^1\ /\sqrt{k}$. We thus can assert that with a few downscaled runs (aimed at ruling out existence of outliers), are sufficient for obtaining reliable estimates of these statistics possibly with a slightly positive bias.

In contrast, the hourly average trip duration was more biased compared to the previous two statistics and its dependence on k changed by periods of the day (Fig. 10). To simplify, we averaged the results over the morning peak, evening peak and off-peak hours. For each of these three periods, the relative bias and its STD were low up to $k \sim 0.30$–0.25, but below these values, the relative bias steadily decreased for the evening peak; steadily increased for the off-peak period, and kept fluctuating for the morning peak. Applying a t-test, for each $k < 0.4$, the bias for the evening peak and the off-peak periods significantly and systematically deviated from zero ($p <$

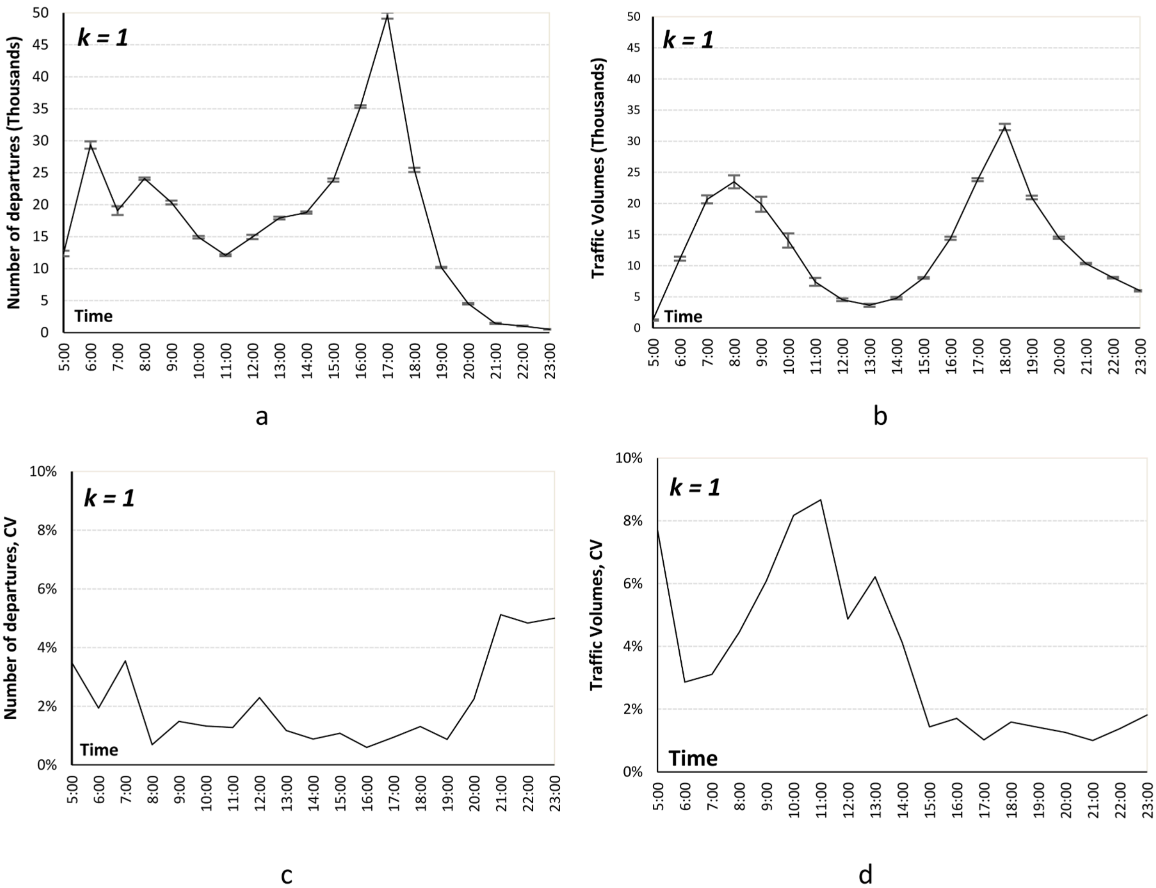

Fig. 7. Hourly dynamics of the full-scale Sioux Falls runs: (a) mean and STD (error bars) number of departures; (b) mean and STD (error bars) of 5-minute traffic volumes; (c) CV of the number of departures; (d) CV of traffic volumes.

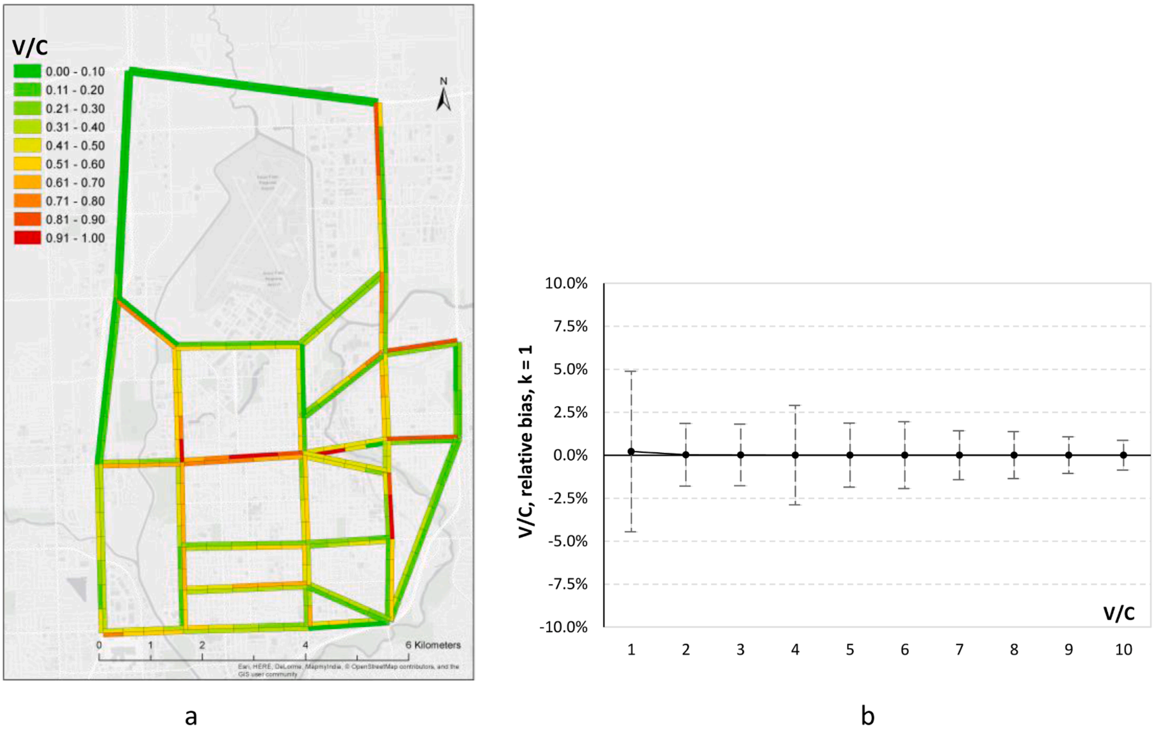

Fig. 8. V/C patterns by road links for $k = 1$: (a) Average V/C; (b) Average and STD (error bars) of relative bias of the V/C by links.

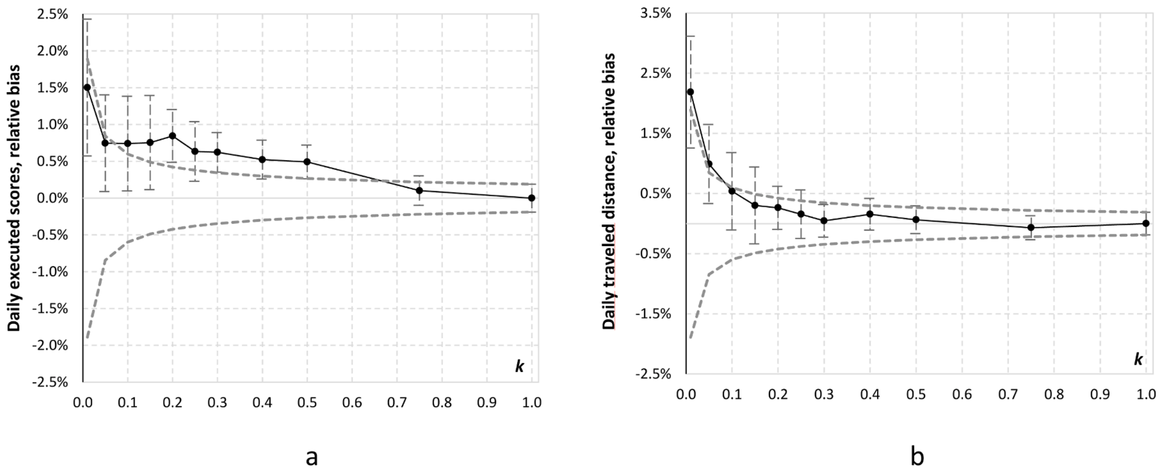

Fig. 9. The dependency of the average relative bias m_b^k and its standard deviation s_b^k (error bars) on the downscaling fraction k for (a) the average daily executed scores and (b) traveled distance. Dashed continuous lines denote $\pm CV_c^1/\sqrt{k}$ boundaries.

0.01). In order for the bias of the average trip duration to remain within the safe boundary for all hours of the day, the value of k must exceed 0.20. Note that for $k \leq 0.20$, the STD of the average trip duration for the morning peak and off-peak periods was at least 5% and the statistics of one downscaled run, even if not identified as an outlier, can deviate from the full-scale statistics by 20% or more. Averaging over 10 repetitions reduced this bias to ~10% (Fig. 10).

5.2.1.2. Scalable statistics. Fig. 11 presents the average relative bias and its STD for the number of daily car departures. As observed, the average bias remained below 5% for all tested k. The STD proportionally increased with the decrease of k for the morning peak period but remained below 5% up to $k = 0.1$. For the evening and off-peak periods STD remained low for all k. We thus conclude that, as above, a few downscaled runs, necessary to rule out existence of outliers, are sufficient for obtaining reliable estimates. The estimates' bias is close to 0 for the morning peak period; negative but low for the evening peak and positive but low for the off-peak hours.

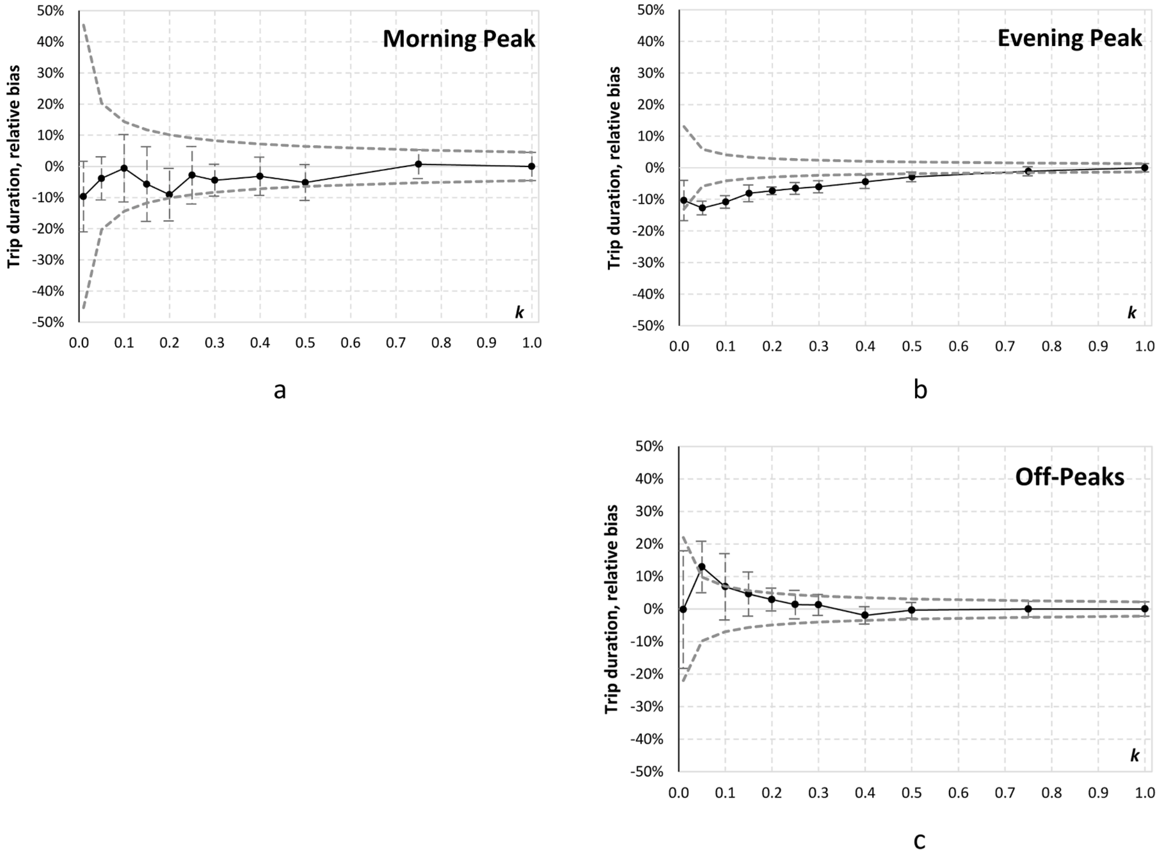

Fig. 10. The average relative bias and its STD (error bars) of the average trip duration - (a) Morning peak; (b) Evening peak; (c) Off-peak. Dashed continuous lines denote $\pm CV_c^1/\sqrt{k}$ boundaries.

Fig. 12 shows the average relative bias and its STD for –the average traffic volumes. The average bias for this statistics was higher than for the number of daily departures, varied for the morning peak hours and systematically negatively increased with the decrease in k for the evening and off-peak hours. The bias STD was very low for the evening peak hours, while starting from $k = 0.20$, reaching 10% for the morning peak hours. Overall, we conclude that both scalable statistics are properly represented for $k > 0.2$.

5.3. Disaggregate statistics

The only disaggregate (and also non-scalable) statistic that we consider is the volume to capacity ratio (V/C). Fig. 13 presents maps - for three specific values of k - 0.25, 0.1 and 0.05 - of the average hourly V/C during the morning peak (06:30–11:00). All appear visually similar to the map of $k = 1$ (Fig. 8) including the average relative bias (by links) as dependent on the V/C value for each link.

A quantitative comparison (Fig. 13d-f) suggests that the downscaled scenarios properly reflect the link values of V/C for $V/C > 0.1$. The STD of the relative bias of V/C remained for these links around 5% and below while reaching ~15% for links with $V/C \leq 0.1$. The latter represents links with very low volumes. As can be seen, the STD of the relative bias decreased with the increase of V/C i.e., proper recognition of the congested links demands less repetitions of the downscaled simulations compared to the estimates of flows where link is very low.

Fig. 14 presents traffic flows as dependent on V/C for each of 10 simulation runs performed for $k = 0.05$ with different downscaling random seeds. As can be seen, the value of the traffic flow for the road links with $V/C \geq 0.15$ is very similar in all runs. This synchronization terminates for $V/C = 0.05$.

6. Simulation runtime

We estimated the model runtime as dependent on k, with the parallel version of QSim [23] and one simulation run for each value of k. All simulations were performed on the same workstation with 7820x Intel® Core™ X-series Processor, with 8 physical and 16 logical cores, with the CPU manually overclocked to 4.5 ghz. The workstation had 32GB of Corsair® RAM, overclocked to 3000mhz.

The parameters of the simulation (besides the value of k) remained the same in all simulations. In our preliminary experiments we tested that the runs' time, which differ in the random seeds only, was the same and, thus, each simulation was run only once for each

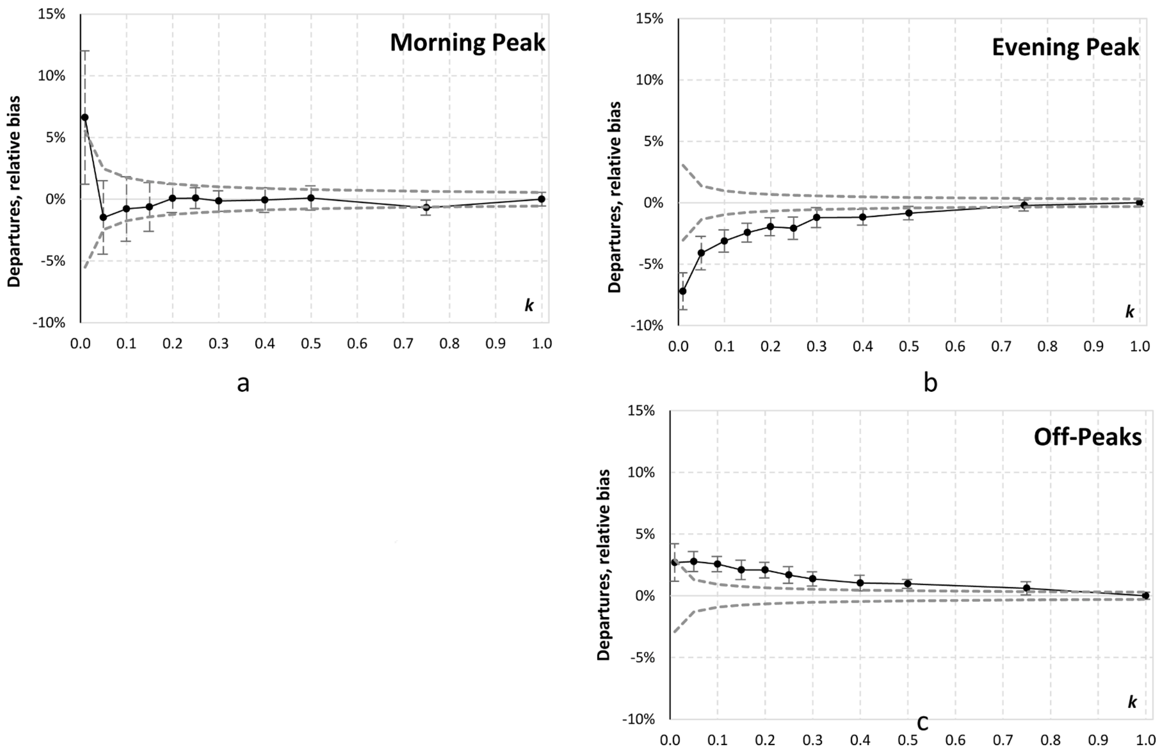

Fig. 11. Average relative bias and STD (error bars) of the number of daily car departures - (a) Morning peak; (b) Evening peak; (c) Off-peak. Dashed continuous lines denote $\pm CV_c^1/\sqrt{k}$ boundaries.

number of cores and k.

The lion share - 76% - 92% - of the runtime in each iteration was taken by the QSIM mobility module. The next processes were scoring, iteration startup, plan innovations, and disk writing. The dependency of the runtime per iteration on k, for different number of cores, is presented in Fig. 15. The performance was similar for the number of cores between 2 and 11 and, in all these cases, the performance was higher than in the case of 1 and 12 cores. The estimates fluctuated for $k < 10\%$, while for higher k fractions, the performance with 2–11 cores was 15–20% higher than with 1 or 12 cores. This phenomenon was initially mentioned by [23] who assumed that the lack of improvements might be related to the growing synchronization time.

It is worth noting that the decline in the performance with the increase in k is quite low. For example, in the case of 2–11 cores, the computation time improvements, per iteration time for $k = 1$ is, roughly, three times longer than for $k = 0.1$. That is, even for an output statistic with a low variance, e.g., averaged daily scores or traveled distance, one could achieve better estimates by running simulations with the highest possible k compared to repeating runs with lower k several times.

Evidently, if the full-scale simulation is too heavy for the available hardware and running the same downscaled simulation several times is the only possible solution to contend with this limitation, the overall run time may well exceed that of the full-scale one. Naturally, a stronger machine could well be a better solution though.

7. Discussion and conclusions

With standard hardware of the year 2020, the number of agent-travelers that can be practically managed in a largescale MATSim scenario lies between 200 K and 500 K. With this number of agents, depending on the computer memory and other hardware components, a single model run will take between a few hours and a full day. The number of travelers in a typical metropolitan region is often higher, and practically MATSim cannot manage full-scale metropolitan simulations. How can modelers then be certain that randomly selected fractions - say 10% of the travelers - will generate the dynamics that correctly replicate a full-scale simulation?

The One-Represents-Several (ORS) downscaling concept, as implemented in MATSim, is essentially simple and effective. Namely, a fraction k of travelers is randomly sampled to represent the entire population and the capacities of each road link are decreased proportionally (linearly or non-linearly) to the reduction in the number of travelers. If schematically, the number of travelers is reduced to 10% and the number of cars located on the links and those that can traverse them will also be reduced by a factor of 10 – nothing will change. However, this simple statement may be incorrect since (1) road links are connected as part of the network, (2) we are interested in dynamics of congested network traffic and (3) agents in MATSim are adaptive and their choice of routes, departure times and daily plans are influenced by the traffic state. Intuitively, decreasing the downscaling fraction k, results in fewer cars representing traffic on a link and in the decrease in the system's robustness to variations in agent behavior. Consequently, the level of

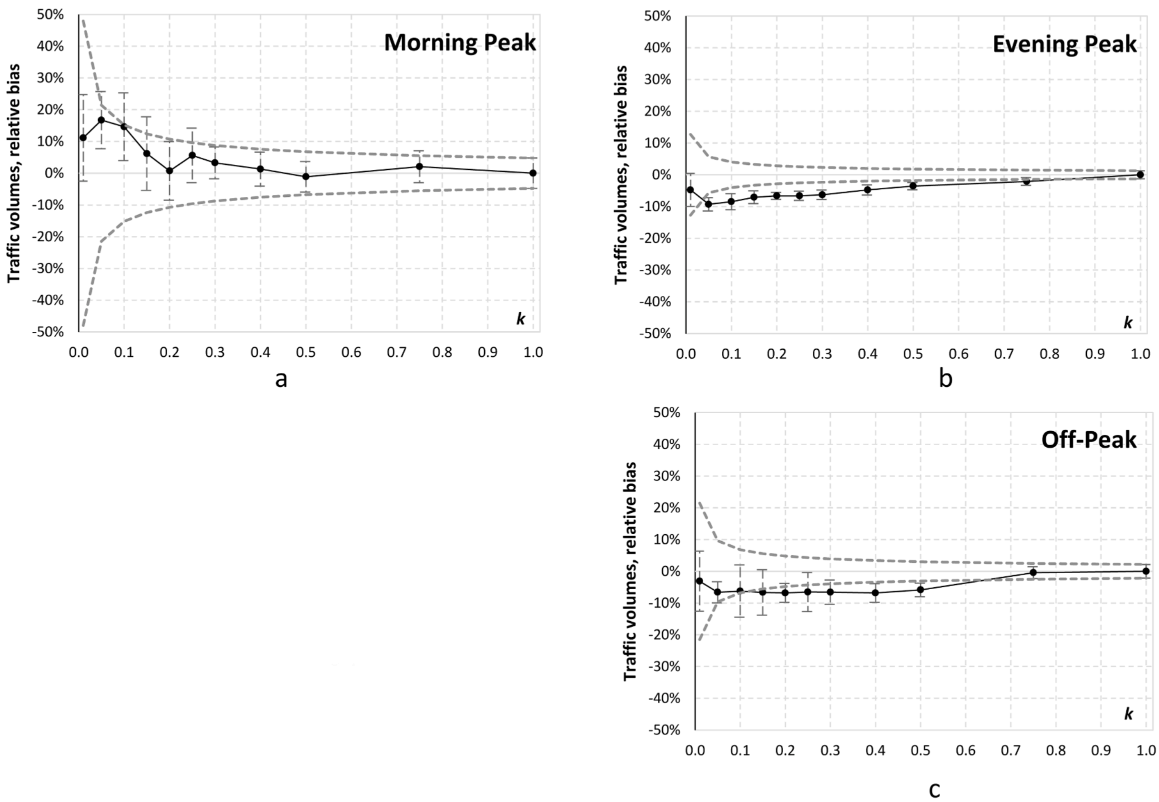

Fig. 12. Average relative bias and STD (error bars) of the traffic volume - (a) Morning peak, (b) Evening Peak and (c) off-peak periods. Dashed continuous lines denote $\pm CV_c^1/\sqrt{k}$ boundaries.

traffic on the congested and, especially, non-congested links may well become non-robust to the decrease in *k,* and traffic dynamics for low *k* will not replicate sufficiently well the dynamics of the full-scale model.

To our best knowledge, these potential effects of ORS downscaling have been somewhat narrow in scope ([20,30,73] and [23]). The aforementioned studies did not fully address any spatial/hourly temporal aspects of ORS downscaling. Rather, they principally investigated only a small number of *k* values, compared only a limited number of simulation outputs, and did not investigate any bias that might arise when replicating simulation runs. Other MATSim studies that applied downscaling, used coarse thumb (e.g., 10% downscaled) regarding the acceptable downscaling fractions. To narrow this knowledge gap, we considered several key output statistics of road traffic dynamics and studied whether they replicate the values obtained in a full-scale simulation for the well-known Sioux Falls test case road network. For each fraction *k* of agents selected for downscaling, we investigated two types of statistics: scalable ones, that are expected to change proportionally to the fraction of the population considered in the downscaled model and non-scalable ones that are expected to be independent of that fraction. For each statistic, we repeated the same scenario, with a different random seed, 10 times, for all investigated values of *k* between $k = 1$ and $k = 0.01$, and then investigated the average relative bias of the statistic, its STD and their dependency on k. Of course, this is a non-exhaustive list of possible network performance measures and others can be considered as well in the future.

Usually the best approach would be to simulate once a 100% scenario given that the variance is likely very low for all output parameters - around 2%. Furthermore, analyzing one run is simpler and more intuitive to communicate than ten repetitions of 10%. However, the inherent hardware limitations make the use of the full population practically impossible. Therefore, the sensible solution should be to use the highest possible fraction, which the modeler's computer hardware can cope with while minimizing the runtime for optimal performance. Never is it wise to decrease the fraction beyond 10%, and it is important to remember that the lower the *k* values are, the more repetitions will be needed.

Our major conclusion is that downscaling up to the level of 20% is qualitatively safe. This result is based on the quantitative robustness of the congestion patterns generated in the downscaled version of the model i.e. the level congestion over the links as reflected by the V/C ratio remained constant on all links with $V/C > 0.1$ up to the values of $k = 0.05$. This local robustness entails sufficient stability of the other output statistics that we investigated. Their average bias for the values of *k* between from 1 down to 0.20–0.25, always remained below 10%. Downscaling to 10–15%, commonly applied in MATSim can result in a larger but still acceptable bias. Downscaling below 10% can be risky. These results are in good agreement with the literature. We thus assert that downscaled runs should be repeated, as many times as necessary to recognize possible outliers and reduce the variance of the relevant

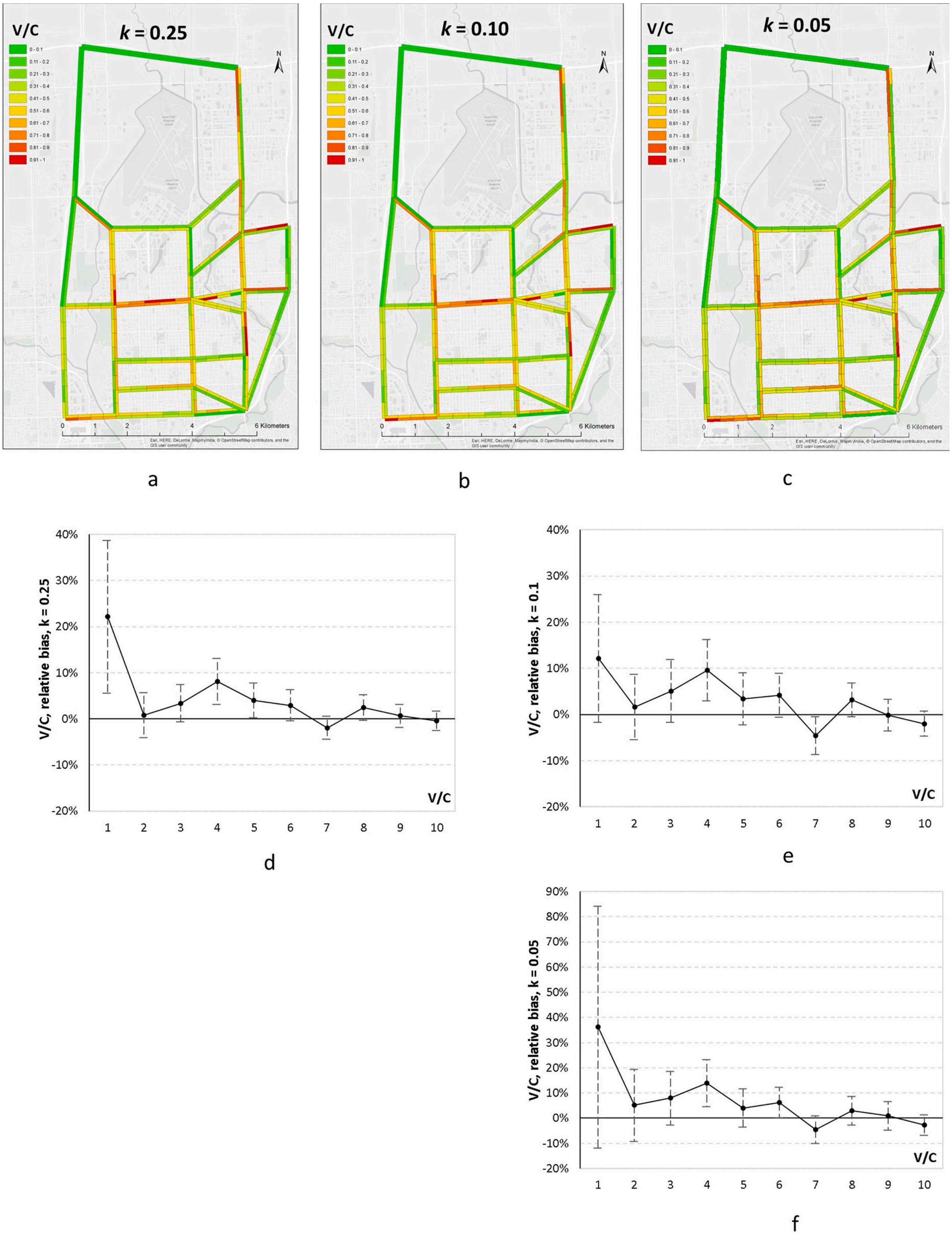

Fig. 13. Average hourly V/C during the morning peak for (a) $k = 0.25$; (b) $k = 0.1$; (c) $k = 0.05$ and the average relative bias of V/C (by link) and STD (error bars) in downscaled scenarios, for $k = 0.25$ (d) $k = 0.1$ (e), $k = 0.05$ (f).

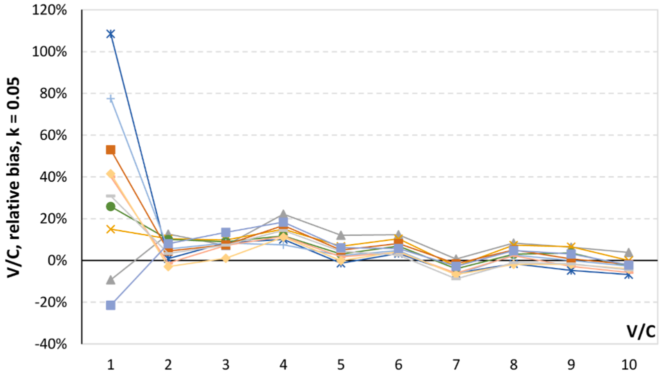

Fig. 14. The relative bias of V/C for each of 10 repetitions of the downscaled scenario $k = 0.05$.

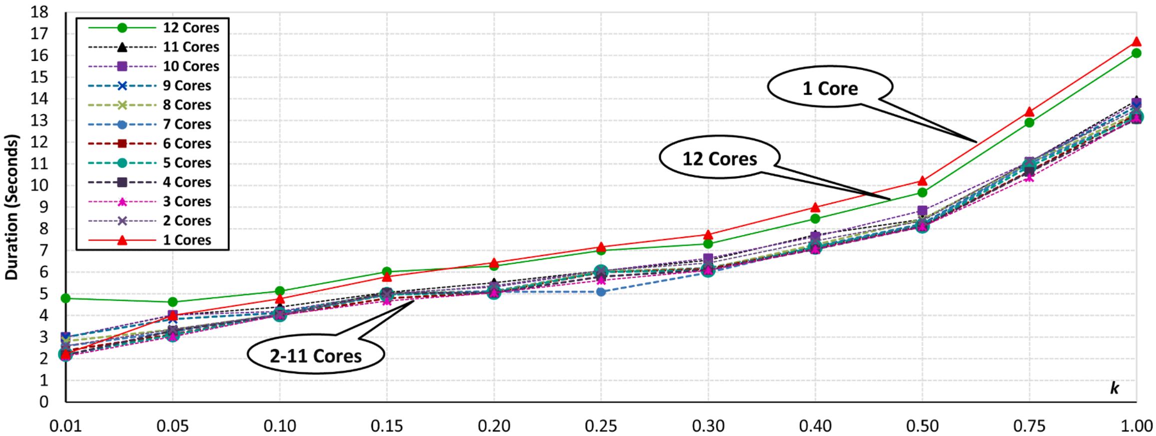

Fig. 15. Average daily iteration runtime, as dependent on k, for different number of cores.

statistic averaged over the downscaled runs to the level expected for the full-scale model. The number of repetitions depends, evidently, on the performance of each statistic. Our estimates could well provide a benchmark for modelers in planning the number of such experiments.

In addition, we compared MATSim's parallel version performance, repeating the same simulation with the number of cores varying between 1 and 12, similar to [23]. Several processors accelerate calculations, but only by 15–20%. Moreover, the performance of the parallel version of MATSim with 12 cores was, again, lower than with 2–11 cores and similar to the model performance with 1 core.

Some studies suggested it is possible to downscale as low as 5%, which would be very beneficial from a practical and operational viewpoint. Nevertheless, it is important to remember these works base their conclusions on the daily averages of the outputs. Investigating hourly and spatially-explicit characteristics of the model outputs, we recommend to use a safe 20% downscaling factor that still economizes on simulation runtime while maintaining relatively low estimation bias of the model's output parameters compared to the full-scale scenario. The more repetitions of a downscaled scenario are performed, the lower is the variance of the parameters' estimates, and, practically, we recommend 10 or more repetitions.

While our study is among the very few that positively confirm the application of downscaling in MATSim, several issues of importance remain that demand further investigation including the trustworthiness of downscaled results transferred from one network to another. No clear answer exists. A downscaled study that is based on one network is evidently insufficient for far-reaching conclusions. The topology of the street network and demand patterns may likely also affect the performance of downscaling. In this respect it is important to note that in the full-scale SF simulation, the average number of agents that traverse a network link per day is about 500. This value is quite similar to the reviewed simulation studies presented in Table 1 - varying between 100 and 1000. We thus assert that in this respect the SF car traffic patterns are roughly similar to the rest of the scenarios simulated with MATSim.

Important limitations of the current study worth noting are that all calculations were considered with only one transportation mode, i.e., cars. As explained in Section 3.3.3, MATSim capacity adjustments (3) – (4), are insufficient in mixed traffic simulations, where cars and PT interact on the same road space. Future studies should also stress downscaling with other Agent-Based transportation models. Especially with the inclusion of Shared-Automated-Vehicle (SAV), where even more computational power will be needed for the SAV dispatching algorithms, downscaled simulations will likely become necessary. Out-of-the box downscaling with SAV traffic may well not work as the number of SAV vehicles must be reduced proportionally to equations (3) – (4) alongside the population. This could raise new questions such as how the SAV passenger occupancy in an ORS-downscaled scenario could be reflected in the full scale one.

Notwithstanding, our study demonstrates, to our best knowledge, a comprehensive investigation of major spatiotemporal downscaling impacts on simulation model robustness. Our results can well provide modelers sensible benchmarks how to cautiously downscale their simulations and in particular, for transportation agent-based models that employ queue-based traffic simulators.

Acknowledgments

Funding for this research was provided by The Israeli Ministry of Transport, Division of the Chief Scientist via the Research Grant "Adoption of the shared autonomous vehicle: A Game-based model of urban travelers and transportation system co-adaptation" and the of Ministry of Science & Technology, Israel and The Ministry of Science & Technology of the People's Republic of China, Grant No. 3-15741; Mobility as a Service: From Rigid to Smart Evolving Public Transport. Special thanks to Dr. M. Maciejewski (TU Berlin). Ph.D. candidate Golan Ben-Dor was partially supported by a scholarship from the Shlomo Shmeltzer Institute for Smart Transportation in Tel-Aviv University and by the Dr. Etel Friedman Memorial Scholarship, Center for Economic and Social Research, Tel Aviv Municipality.

References

[1] S. Abar, G.K. Theodoropoulos, P. Lemarinier, G.M. O'Hare, Agent based modelling and simulation tools: a review of the state-of-art software, Comput. Sci. Rev., 24 (2017) 13–33.

[2] M. Adnan, F.C. Pereira, C.M.L. Azevedo, K. Basak, M. Lovric, S. Raveau, M. Ben-Akiva, SimMobility: a multi-scale integrated agent-based simulation platform, in: Paper presented at the *95th Annual Meeting of the Transportation Research Board Forthcoming in Transportation Research Record*, 2016.

[3] J. Auld, M. Hope, H. Ley, V. Sokolov, B. Xu, K. Zhang, POLARIS: agent-based modeling framework development and implementation for integrated travel demand and network and operations simulations, Transp. Res. Part C: Emerg. Technol., 64 (2016) 101–116.

[4] T. Arentze, F. Hofman, H. van Mourik, H. Timmermans, ALBATROSS: multi-agent, rule-based model of activity pattern decisions, Transp Res Rec 1706 (1) (2000) 136–144.

[5] M. Balać, K.W. Axhausen, Activity rescheduling within a multi-agent transport simulation framework (MATSim), Arbeitsberichte Verkehrs-Und Raumplanung 1180 (2016), https://doi.org/10.3929/ethz-b-000249228.

[6] M. Balmer, K. Meister, M. Rieser, K. Nagel, K.W. Axhausen, Agent-based simulation of travel demand: structure and computational performance of MATSim-T, in: Paper presented at the *Proceedings of the Second TRB Conference on Innovations in Travel Modeling, Portland*, 2008, p. 504, https://doi.org/10.3929/ethz-a-005626451. *June 2008*.

[7] A. Bassolas, J.J. Ramasco, R. Herranz, O.G. Cantu-Ros, Mobile phone records to feed activity-based travel demand models: mATSim for studying a cordon toll policy in Barcelona, arXiv Preprint (2018).

[8] B.H. Nahmias-Biran, J.B. Oke, N. Kumar, C. Lima Azevedo, M. Ben-Akiva, Evaluating the impacts of shared automated mobility on-demand services: an activity-based accessibility approach, Transportation (Amst) 1-26 (2020).

[9] S. Bekhor, C. Dobler, K. Axhausen, Integration of activity-based and agent-based models: case of Tel Aviv, israel, Transp. Res. Rec.: J. Transp. Res. Board 2255 (2011) 38–47.

[10] T. Bellemans, B. Kochan, D. Janssens, G. Wets, T. Arentze, H. Timmermans, Implementation framework and development trajectory of FEATHERS activity-based simulation platform, Transp. Res. Rec. 2175 (1) (2010) 111–119.

[11] G. Ben-Dor, B. Dmitrieva, M. Maciejewski, J. Bischoff, E. Ben-Elia, I. Benenson, MATSim simulations in the tel aviv metropolitan area: direct competition between public transport and cars on the same roadway, in: hEART 2017: 6th Symposium of the European Association for Research in Transportation. Technion, Haifa, 2017. https://transp-or.epfl.ch/heart/2017/abstracts/hEART2017_paper_110.pdf (accessed 8 January 2020).

[12] G. Ben-Dor, E. Ben-Elia, I. Benenson, Assessing the impacts of dedicated bus lanes on urban traffic congestion and modal split with an agent-based model, Procedia Comput. Sci. 130 (2018) 824–829.

[13] I. Benenson, P.M. Torrens, Geosimulation: object-based modeling of urban phenomena, Comput. Environ. Urban Syst. 28 (1–2) (2004) 1–8.

[14] J. Bischoff, M. Maciejewski, Autonomous taxicabs in berlin–a spatiotemporal analysis of service performance, Transp. Res. Procedia 19 (2016) 176–186.

[15] J. Bischoff, M. Maciejewski, Simulation of city-wide replacement of private cars with autonomous taxis in berlin, Procedia Comput. Sci. 83 (2016) 237–244. Bonabeau, E. (2002). Agent-based modeling: Methods and techniques for simulating human systems.

[16] E. Bonabeau, Agent-based modeling: methods and techniques for simulating human systems, Proceed. Natl. Acad. Sci. 99 (2002) 7280–7287.

[17] P.M. Boesch, F. Ciari, K.W. Axhausen, Autonomous vehicle fleet sizes required to serve different levels of demand, Transp. Res. Rec.: J. Transp. Res. Board 2542 (2016) 111–119.

[18] P.M. Bösch, F. Ciari, A multimodal network for MATSim, in: Paper presented at the *Proceedings of the 15th Swiss Transport Research Conference, Ascona*, Switzerland, 2015, pp. 15–17, https://doi.org/10.3929/ethz-b-000100802.

[19] P.M. Bösch, K. Müller, F. Ciari, The IVT 2015 baseline scenario, in: Paper presented at the *16th Swiss Transport Research Conference (STRC 2016)*, 2016. http://www.strc.ch/2016/Boesch_EtAl.pdf (accessed 8 January 2020).

[20] C. Llorca, R Moeckel, Effects of scaling down the population for agent-based traffic simulations, Procedia Comput. Sci. 151 (2019).

[21] A. Chakirov, P. Fourie, Enriched sioux falls scenario with dynamic and disaggregate demand, Arbeitsberichte Verkehrs-Und Raumplanung (2014) 978, https://doi.org/10.3929/ethz-b-000080996.

[22] F. De Souza, O. Verbas, J. Auld, Mesoscopic traffic flow model for agent-based simulation, Procedia Comput. Sci. 151 (2019) 858–863.

[23] C. Dobler, Implementation of a time step based parallel queue simulation in MATSim, in: Paper presented at the *10th Swiss Transport Research Conference. Monte Verita*, Ascona, 2010. http://www.strc.ch/2010/Dobler.pdf (accessed 8 January 2020).

[24] C. Dobler, G. Lämmel, Integration of a multimodal simulation module into a framework for large-scale transport systems simulation, Pedestrian Evacuat. Dynam. 2012 (2014) 739–754.

[25] A. Erath, P. Fourie, Eggermond Van, S. Ordóñez, A. Chakirov, K. Axhausen, Large-scale agent-based transport demand model for singapore, in: Paper presented at the *13th International Conference on Travel Behaviour Research (IATBR). Toronto: International Association for Travel Behaviour Research*, 2012, https://doi.org/10.3929/ethz-b-000306926.

[26] P.J. Fourie, Multi-modeling in matsim: psim, in: Nagel Horni, Axhausen (Eds.), The Multi-Agent Transport Simulation MATSim, Ubiquity Press, London, 2016, pp. 263–266.

[27] W. Gao, M. Balmer, E. Miller, Comparison of MATSim and EMME/2 on greater Toronto and Hamilton area network, Transp. Res. Rec.: J. Transp. Res. Board 2197 (2010) 118–128.

[28] L. Gasser, K. Kakugawa, B. Chee, M. Esteva, Smooth scaling ahead: progressive MAS simulation from single PCs to Grids, in: Multi-agent and multiagent-based simulation. Joint Workshop MABS 2004, New York, NY, USA, 2005. July 19, 2004.

[29] P. Heywood, S. Maddock, J. Casas, D. Garcia, M. Brackstone, P. Richmond, Data-parallel agent-based microscopic road network simulation using graphics processing units, Simul. Modell. Pract. Theo. 83 (2018) 188–200.

[30] S. Hemdan, A.M. Wahaballa, F. Kurauchi, Travel time variability and macroscopic fundamental diagram relationships in multimodal networks, Int. J. Intell. Transp. Syst. Res. 17 (2) (2018) 114–124.
[31] D. Helbing, S. Balietti, How to do agent-based simulations in the future: from modeling social mechanisms to emergent phenomena and interactive systems design, in: D. Helbing (Ed.), Social Self-Organization: Agent-Based Simulations and Experiments to Study Emergent Social Behavior, Understanding Complex Systems, Springer, Heidelberg, 2013, pp. 25–70.
[32] S. Hörl, Agent-based simulation of autonomous taxi services with dynamic demand responses, Procedia Comput. Sci. 109 (2017) 899–904.
[33] S. Hörl, A MATSim scenario for autonomous vehicles in la défense and île-de-france, Arbeitsberichte Verkehrs-Und Raumplanung 1239 (2017), https://doi.org/10.3929/ethz-b-000128143.
[34] S. Hörl, M. Balac, K.W. Axhausen, Dynamic demand estimation for an AMoD system in Paris, in: 2019 IEEE Intelligent Vehicles Symposium (IV), 2019, pp. 260–266.
[35] A. Horni, D. Charypar, K.W. Axhausen, Variability in transport microsimulations investigated for MATSim: preliminary results, [Arbeitsberichte Verkehrs-Und Raumplanung], (2011) 676, https://doi.org/10.3929/ethz-a-006482318.
[36] A. Horni, K. Nagel, K.W. Axhausen, The Multi-Agent Transport Simulation MATSim, Ubiquity Press, London, 2016.
[37] A. Horni, D. Scott, M. Balmer, K. Axhausen, Location choice modeling for shopping and leisure activities with MATSim: combining microsimulation and time geography, Transp. Res. Rec.: J. Transp. Res. Board 2135 (2009) 87–95.
[38] F. Hülsmann, R. Gerike, M. Ketzel, Modelling traffic and air pollution in an integrated approach–the case of munich, Urban Clim., 10 (2014) 732–744.
[39] M. Lozano, P. Morillo, D. Lewis, D. Reiners, C. Cruz-Neira, A distributed framework for scalable large-scale crowd simulation, in: International Conference on Virtual Reality, Berlin, Heidelberg 111-121, Springer, 2007.
[40] J. Illenberger, G. Flotterod, K. Nagel, Enhancing MATSim with capabilities of within-day re-planning, in: Paper presented at the *Intelligent Transportation Systems Conference, 2007. ITSC 2007*, IEEE, 2007, pp. 94–99.
[41] M.A. Ji, P.E.I. Yu-long, Development and application of TransCAD for urban traffic planning [J], J. Harbin Univ. Civ. Eng. Architec. 5 (2002).
[42] I. Kaddoura, B. Kickhöfer, Optimal road pricing: towards an agent-based marginal social cost approach, in: VSP working paper 14-01, TU Berlin, transport systems planning and transport telematics, 2014. https://www.researchgate.net/profile/Benjamin_Kickhoefer/publication/275637691_Optimal_Road_Pricing_Towards_an_Agent-Based_Marginal_Social_Cost_Approach/links/554737130cf24107d39820e1/Optimal-Road-Pricing-Towards-an-Agent-based-Marginal-Social-Cost-Approach.pdf (accessed 31 May 2020).
[43] G.O. Kagho, M. Balac, K.W. Axhausen, Agent-based models in transport planning: current state, issues, and expectations, Procedia Comput. Sci. 170 (2020) 726–732.
[44] B. Kickhofer, D. Hosse, K. Turnera, A. Tirachinic, Creating an open MATSim scenario from open data: the case of santiago de chile, Tech. Rep., VSP Work. Pap. 16-02 (2016) (accessed 8 January 2020).
[45] B. Kickhöfer, F. Hülsmann, R. Gerike, K. Nagel, Rising car user costs: comparing aggregated and geo-spatial impacts on travel demand and air pollutant emissions, in: Thomas Vanoutrive, Ann Verhetsel (Eds.), Smart Transport Networks: Decision Making, 2013, pp. 180–207.
[46] D. Krajzewicz, M. Bonert, P. Wagner, The open source traffic simulation package SUMO, RoboCup 2006 (2006).
[47] M. Maciejewski, J. Bischoff, Congestion effects of autonomous taxi fleets, Transport (2017) 1–10.
[48] G. McArdle, E. Furey, A. Lawlor, A. Pozdnoukhov, Dublin, in: Nagel Horni, Axhausen (Eds.), The Multi-Agent Transport Simulation MATSim, Ubiquity Press, London, 2016, pp. 413–418.
[49] J. McCarthy, Generality in artificial intelligence, Commun ACM 30 (12) (1987) 1030–1035.
[50] K. Nagel, F. Marchal, Computational methods for multi-agent simulations of travel behavior, in: Proceedings of International Association for Travel Behavior Research (IATBR), Lucerne, Switzerland, 2003.
[51] T.W. Nicolai, Using MATSim as a travel model plug-in to UrbanSim, in: VSP Working Paper, 2012, pp. 12–29. *TU Berlin, Transport Systems Planning*, http://svn.vsp.tu-berlin.de/repos/public-svn/publications/vspwp/2012/12-29/chapter_for_D72_final_version_submitted20121002.pdf (accessed 8 January 2020).
[52] T.W. Nicolai, K. Nagel, High resolution accessibility computations. In Ana Condeço-Melhorado, Aura Reggiani & Javier Gutiérrez (Eds.). Accessibility and Spatial Interaction, 2014, pp. 62–91.
[53] S. Ordóñez, Multi-day activity models: an extension of the multi-agent transport simulation (MATSim), Arbeitsberichte Verkehrs-Und Raumplanung 1211 (2016), https://doi.org/10.3929/ethz-b-000120636.
[54] M. Paulsen, T.K. Rasmussen, O.A. Nielsen, Output variability caused by random seeds in a multi-agent transport simulation model, Procedia Comput. Sci. 130 (C) (2018) 850–857.
[55] H.R. Parry, M. Bithell, Large scale agent-based modelling: a review and guidelines for model scaling. Agent-based Models of Geographical Systems, Springer, New York, 2012, pp. 271–308.
[56] H.R. Parry, Agent based modeling, large scale simulations, editor., in: R.A. Meyers (Ed.), Encyclopedia of Complexity and Systems Science, Springer, New York, 2009, pp. 148–160.
[57] M. Rieser, K. Nagel, Network breakdown "at the edge of chaos" in multi-agent traffic simulations, Eur. Phys. J. B 63 (3) (2008) 321–327.
[58] D. Röder, I. Cabrita, K. Nagel, Simulation-based sketch planning, part III: calibration of a MATSim-model for the greater brussels area and investigation of a cordon pricing for the highway ring, in: VSP Working Paper 13-16, TU Berlin, Berlin, Germany, 2013. http://svn.vsp.tu-berlin.de/repos/public-svn/publications/vspwp/2013/13-16/brussels-2013-06-23.pdf (accessed 8 January 2020).
[59] F.D. Rollo, A.G. Schulz, A contrast efficiency function for quantitatively measuring the spatial-resolution characteristics of scanning systems, J. Nucl. Med. 11 (2) (1970) 53–60.
[60] R. Rothfeld, M. Balac, K.O. Ploetner, C. Antoniou, Initial analysis of urban air mobility's transport performance in Sioux falls, in: 2018 Aviation Technology, Integration, and Operations Conference, 2018. https://arc.aiaa.org/doi/pdf/10.2514/6.2018-2886/ (accessed 30 May 2020).
[61] I. Saadi, A. Mustafa, J. Teller, M. Cools, Calibration of MATSim in the context of natural hazards in belgium, in: Paper presented at the *XII Congreso De Ingeniería Del Transporte. 7, 8 Y 9 De Junio, Valencia (España)*, 2016, pp. 859–869.
[62] A. Saprykin, N. Chokani, R.S. Abhari, GEMSim: a GPU-accelerated multimodal mobility simulator for large-scale scenarios, Simul. Modell. Pract. Theo. 94 (2019) 199–214.
[63] I.N. Sener, C.R. Bhat, R. Copperman, S. Srinivasan, J.Y. Guo, A. Pinjari, N. Eluru, in: Activity-based travel-demand analysis for metropolitan areas in Texas: CEMDAP Models, Framework, Software Architecture and Application Results, Midwest Regional University Transportation Center, 2006.
[64] M.D. Simoni, A.J. Pel, R.A. Waraich, S.P. Hoogendoorn, Marginal cost congestion pricing based on the network fundamental diagram, Transp. Res. Part C: Emerg. Technol. 56 (2015) 221–238.
[65] C. Tchervenkov, J. Molloy, K.W. Axhausen, Estimating externalities from GPS traces using MATSim, in: Paper presented at the *18th Swiss Transport Research Conference (STRC 2018)*, 2018, https://doi.org/10.3929/ethz-b-000264806.
[66] The Highway Capacity Manual, 6th Edition A Guide for Multimodal Mobility Analysis (2016), Transportation research board of the national academies, ISBN 0309369975, Washhington DC.
[67] John W Tukey, Exploratory Data Analysis, Addison-Wesley, 1977, p. 720.
[68] Vorraa, T. (2009). Transport modelling supported by GIS–an overview of GIS features now within cube. urban transport XV: urban transport and the environment, 15, 235.
[69] R. Waraich, K. Axhausen, Agent-based parking choice model, Transp. Res. Rec.: J. Transp. Res. Board 2319 (2012) 39–46.
[70] B. Wang, S.A.O. Medina, P. Fourie, Simulation of autonomous transit on demand for fleet size and deployment strategy optimization, Procedia Comput Sci 130 (2018) 797–802.
[71] C. Zhuge, C. Shao, Baoding: a case study for testing a new household utility function in MATSim, Multi-Agent Transp. Simul. MATSim.Ubiquity (2016) 409–412.
[72] C. Zhuge, C. Shao, J. Gao, C. Dong, H. Zhang, Agent-based joint model of residential location choice and real estate price for land use and transport model, Comput. Environ. Urban. Syst. 57 (2016) 93–105.

[73] C. Zhuge, C. Shao, S. Wang, Y. Hu, Sensitivity analysis of integrated activity-based model: using MATSim as an example, Transp. Lett. 11 (2) (2017) 93–103.
[74] D. Ziemke, K. Nagel, C. Bhat, Integrating CEMDAP and MATSim to increase the transferability of transport demand models, Transp. Res. Rec.: J. Transp. Res. Board (2493) (2015) 117–125.
[75] Manual, H.C.. Highway Capacity Manual, ourth edition Edition, Transportation Research Board, Washington, DC, 2000, ISBN 0-309-06681-6.