Uncertainties of Sub-Scaled Supply and Demand in Agent-Based Mobility Simulations with Queuing Traffic Model

Aleksandr Saprykin^{1,2} · Ndaona Chokani^{1,2} · Reza S. Abhari^{1,2}

Accepted: 13 January 2021 / Published online: 27 May 2021
© The Author(s) 2021

Abstract

Agent-based models for dynamic traffic assignment simulate the behaviour of individual, or group of, agents, and then the simulation outcomes are observed on the scale of the system. As large-scale simulations require substantial computational power and have long run times, most often a sample of the full population and downscaled road capacities are used as simulation inputs, and then the simulation outcomes are scaled up. Using a massively parallelized mobility model on a large-scale test case of the whole of Switzerland, which includes 3.5 million private vehicles and 1.7 million users of public transit, we have systematically quantified, from 6 105 simulations of a weekday, the impacts of scaled input data on simulation outputs. We show, from simulations with population samples ranging from 1% to 100% of the full population and corresponding scaling of the traffic network, that the simulated traffic dynamics are driven primarily by the flow capacity, rather than the spatial properties, of the traffic network. Using a new measure of traffic similarity, that is based on the chi-squared test statistic, it is shown that the dynamics of the vehicular traffic and the occupancy of the public transit are adversely impacted when population samples less than 30% of the full population are used. Moreover, we present evidence that the adverse impact of population sampling is determined mostly by the patterns of the agents' behaviour rather than by the traffic model.

Keywords Population sampling · Demand scaling · Supply scaling · Agent-based simulation · Activity-based model · Traffic simulation

✉ Aleksandr Saprykin
saprykin@lec.mavt.ethz.ch

¹ Laboratory for Energy Conversion, Sonneggstrasse 3, 8092 Zurich, Switzerland

² Swiss Mobility Lab, Institute of Science, Technology and Policy, Universitätstrasse 41, 8092 Zurich, Switzerland

1 Introduction

The complexity of modern transportation systems continues to increase in order to efficiently serve the demand from the increasing population in urban areas and to ensure compliance with policies that support the energy transition (Litman 2013; Speranza 2018). Emerging technologies such as electric vehicles (EVs) and automated vehicles (AVs) are among the main drivers of the ongoing transformations in the transportation sector (Burns 2013; Hars 2015). EVs and AVs require new infrastructure (such as, chargers for EVs and intelligent traffic management systems for AVs) to be built and maintained (Huang et al. 2015). The increased penetration of EVs and AVs is also paralleled by changes in the behaviour of people, for example due to the fact that EVs require additional time for charging, or that EVs have limited range compared to internal combustion engine vehicles (Dijk et al. 2013), or that cheaper AVs result in a different perception of the value of time (Childress et al. 2015). As further reductions in mobility-related CO₂ emissions are targeted by many countries across the globe (Biresselioglu et al. 2018), it is anticipated that the transition to EVs will intensify in the near future. Furthermore, traditional and free-floating car-sharing services are expected to reduce the cost of car ownership (Prettenhaler and Steininger 1999) and to have a positive impact on sustainable travel (Cervero et al. 2007) by changing patterns of behaviour. Simultaneously, there is a trend towards a closer integration of different transportation modes into multi-modal networks with the idea that mobility-as-a-service (Jittrapirom et al. 2017) may potentially improve the overall inter-modal interaction of service providers, and especially make public transit more attractive to existing car owners.

The above mentioned changes in mobility require substantial investments in the upgrade of infrastructure in the transportation, electricity and telecommunication sectors (Schroeder and Traber 2012). Moreover, new policies are being formulated and implemented in order to facilitate the upcoming changes and to accelerate the adoption of new technologies (Bakker and Trip 2013). However, uncertainties in such complex systems, which involve both people behaviour and policy regulation, increase investment risks (De Palma et al. 2012; Pagani et al. 2019) and slow down the required improvements to infrastructure. Thus, there is need for a holistic planning tool, especially for large-scale metropolitan areas, whereby for plausible "what-if" scenarios the anticipated changes in demand and supply of mobility can be simulated, and in advance, the required adaptations in infrastructure and needed financing be determined.

Agent-based modelling (ABM) is one approach to model the behaviour and interaction of people in complex urban environments (Zhang and Levinson 2004; Eppstein et al. 2011; Fagnant and Kockelman 2014; Zhuge and Shao 2018). In ABM, agents are considered as individuals with their own logic of behaviour and rules of interaction; thus, ABM simulations are more realistic, and provide a framework to understand causal effects in the complex system as a whole. ABM is not limited to humans, but cover a broad range of disciplines from particles in physics to living species in biology (Abar et al. 2017). Typically, in mobility ABM, each agent represents a human being or a vehicle within a simulated environment that is comprised of roads, public transit system, and other mobility services as well as the policies and

regulations that influence the decision-making of agents. During simulation, agents can react to events and adapt their behaviour. For example, if a road is blocked because of an accident, then agents can re-route to reach their intended destinations.

On the one hand, ABM yields transport simulations with a higher level of detail and complexity in which behaviour and decision-making are driven by the same factors as in the real world (for example, congestion, road accidents, public transport delays, social networks, locations of places of activity, etc.). Thus, very realistic scenarios can be formulated and the impacts of potential changes in policies, infrastructure and transport services can be accurately assessed, at the spatial and temporal resolution of each simulated agent. On the other hand, large-scale and detailed agent-based models – that integrate complex behaviour and rules of interaction – impose a substantial computational burden on the simulation process. ABM applied to large metropolitan areas may encounter bottlenecks in computational performance with correspondingly long simulation run times; these challenges reduce the attractiveness of such models for transportation planners and policy makers. Moreover, the cost of required hardware is also a burden when high performance is required.

One approach to get feasible run times is to improve the computational performance of the simulation itself using either multi-threading and distributed computing (Cameron and Duncan 1996; Rickert and Nagel 2001; Nökel and Schmidt 2002; Cetin et al. 2003; Klefstad et al. 2005), efficient algorithms (Goldberg and Harrelson 2005; Sanders and Schultes 2005) or dedicated acceleration hardware (Strippgen and Nagel 2009). Despite the improvements made with such an approach in recent years, it seems that the complexity of simulated transportation systems is increasing at a rate faster than that at which performance improvements come into place.

Therefore, another approach adopted by researchers is, first to downscale the demand and supply of mobility in the simulated scenario: a fraction of agents is randomly sampled from the full population, while the capacity of the road network (that is, the rate at which vehicles pass through a transportation link) and other infrastructure are also downscaled. And the simulation outcomes are then scaled up, and these results considered to be a good approximation of the scenario with the full population and the actual network. For example, if a 10% sub-sample of the whole population is used, then the capacity of the roads is downscaled also approximately to 10% (some variation is possible due to calibration). Then, the simulation outcomes are scaled up by a factor of 10. Basically, each agent in the downscaled scenario represents an aggregate group, not each individual, of the real world.

It is evident that such an aggregated behaviour in a simulated scenario can lead to errors. For example, a scaled road network introduces rounding errors (such as, when the physical length of a street is too short when large downscaling factors are used), or there may be an over- and under-supply of travel demand on links and for public transit due to the skew of the sampled population on certain routes and at certain locations. Depending on the context of the scenario, distortions in the simulation outcomes may lead to a biased interpretation of results. In light of the ongoing shift to e-mobility one can also consider the impact on electricity distribution grids: in a simulation with a sub-sampled population, the demand for charging would be concentrated at certain aggregated points and may lead to the overloading of power lines, while in reality the demand may actually be distributed more uniformly in space

and time without causing line overloading issues. Another open question is how one should properly scale the charging infrastructure of EVs in such simulation with a sub-sampled population.

It should be noted that in general ABM introduces uncertainties into simulations (Ševčíková et al. 2007; Beykai and Miller 2017) because input data, such as synthetic population and travel demand, are typically derived from surveys and aggregated statistics, and furthermore, the modelled patterns of behaviour are also mostly based on surveys. In contrast to the almost unavoidable uncertainties that result from input data, the uncertainties that result from the use of downscaled input data are introduced artificially because of the lack of required computational power, and these uncertainties can be avoided either by employing a more efficient simulation approach or by running the simulations on more powerful and expensive hardware. Until now, there has been relatively little effort made to understand and quantify the uncertainties that are introduced by sub-sampling the demand and supply in mobility simulations. Moreover, there is a strong evidence (see Section 2 below) that, in comparison to a simulation with the full-scale input data, sub-sampling distorts the spatio-temporal characteristics of the simulated travel demand and of the traffic externalities.

In this paper, the impacts of using sampled populations and scaled input in agent-based scenarios, with a mesoscopic queueing model for traffic propagation, are quantified. The contributions of the paper are the following:

- A systematic and quantitative assessment of errors that arise in agent-based transport simulations with downscaled input data is presented; specifically, the sensitivities of predicted traffic flow dynamics and public transit occupancy rates are quantified. While the public transit system does not share network links with cars and itself was not scaled in this study, the impacts of scaled travel demand on the occupancies of vehicles in a public transit system are captured.
- A GPU-accelerated mesoscopic queueing traffic model is described, and its sensitivity to variations in traffic dynamics under a range of scale factors is evaluated.
- A large-scale scenario for Switzerland that includes 3.5 million cars and 1.7 million users of public transit is used in the assessment.
- A novel similarity measure, based on the chi-squared statistic, is proposed to assess the quality of simulated spatio-temporal traffic dynamics.

The rest of the paper is organized as follows. In Section 2 an overview of previous related work is given. In Section 3, a large-scale scenario for Switzerland and a mobility model are presented, and the approach that is used to assess the impacts of sub-scaled inputs on the simulated traffic flows and public transit utilization is described. The results of the study are then presented in Section 4; the conclusions and a discussion are presented in Section 5.

2 Related Work

There are only a few multi-agent mobility simulators capable of running truly large-scale and multi-modal scenarios with millions of agents and millions of links and nodes in the network: TRANSIMS (Smith et al. 1995), MATSim (Horni et al. 2016),

GEMSim (Saprykin et al. 2019a). MATSim, as one of the most widely used, provides the main evidence for issues with sub-sampled populations.

It is common among researchers to use small population samples (1% to 10%) to run large-scale scenarios with scaled input data. Hülsmann et al. (2014) used 1% of population to run a Munich scenario while studying traffic-related air pollution. The speed of computations was the major reason for downscaling, therefore simulation output results were scaled up back to 100% of population. Zhang et al. (2013) used 1% of population to run a large-scale scenario for the city of Shanghai. The memory constraints of the available hardware were stated as the main reason for downscaling. Bekhor et al. (2011) integrated activity-based and agent-based models for the Tel Aviv metropolitan area using 10% of population to avoid long running times. Kickhofer et al. (2016) used a population sample of 0.65% to run a large-scale scenario for the city of Santiago de Chile, where the authors conclude that the congestion patterns of the simulated scenario do not match the real patterns well and recommend expanding the population sample in the range of 10% to 100%.

More scenarios and case studies with population samples are available in the MATSim book (Horni et al. 2016). Furthermore, MATSim developers present in the above mentioned book the use of a 1% to 10% sample of the population to obtain reliable results with the acceptable run times. Therefore, the most popular sample sizes for large-scale scenarios run with MATSim are within this range.

In recent years, ABM simulations with the use of public transit and other emerging modes of transportation have increased. As a consequence, the required computational power has also increased and researchers have continued to use small population samples so that the run times of their simulations are kept within a reasonable timeframe. However, there is evidence that small population samples are the source of some discrepancies in traffic simulations.

Ben-Dor et al. (2017) faced issues using network links shared by private cars and public transit vehicles in MATSim simulations of the Tel Aviv metropolitan area with a 10% population sample. In the simulations buses tended to get stuck in long waiting queues because the link flow capacity was over-used by cars, resulting in disruptions of public transit services. The study showed that when a coarse population sample is used, then the traffic flows on links are not scaled by the same ratio. Furthermore, while for private cars it is possible to calibrate the predicted road traffic flows in order to reduce the adverse effects of scaling, on the other hand public transit has to run according to schedule and thus, public transit is more sensitive to changes in network capacity.

Bischoff and Maciejewski (2016) in a study of autonomous taxis in the Berlin area showed that use of a 10% population sample leads to 11% of the demand for autonomous taxis compared to a simulation with the full population; that is a 10% relative error. On the other hand, changes in fleet occupancy statistics were considered acceptable, and the deviations in the durations of both pickup trips and trips with customers were no more than 3% of the length.

Simoni et al. (2015) compared the accumulation-production relationship for the links in a MATSim simulation of central Zurich using different sized population samples (10%, 20% and 50%). The simulation results were seen to vary with size of the population sample. It was also noted that flows on links decreased faster with the

increased flow density of the larger population samples. However, only qualitative graphical comparison were presented, without any quantitative results.

Erath et al. (2012) had artefacts of overcrowded buses in a MATSim simulation of the public transit in Singapore using a 10% population sample. In order to improve the simulation, a 25% population sample was used. No further details on the issues with public transit were presented.

Bösch et al. (2016) in the Switzerland baseline scenario for MATSim put an attention on the usage of population sub-samples by providing an example from car-sharing simulations: using a smaller population sample means that a reduction in shared vehicles is required to prevent over-supply, and at the same time reduction of the number of shared vehicles leads to reduced availability in the area or under-supply. Thus, it is not recommended to use population samples with less than 5%–10% for scenarios with shared cars. Also issues with the scaling of public transit were mentioned: because of the inability to scale a public transit schedule itself (that is, a fleet size, the frequency of operation), it is suggested to scale down the size (and, as consequence, link flow capacity consumed by a vehicle) of public transit vehicles proportionally to the size of the population sample. But as it was shown by Ben-Dor et al. (2017), this approach does not solve the issues with public transit completely when using population sub-samples.

Kwak et al. (2012) conducted probably the first attempt to systematically study the errors that arise in traffic simulations due to the use of samples of the full population. In the work, a macroscale static traffic assignment model was used for different periods of the day. This traffic model was not an agent-based model, but rather used origin-destination (OD) matrices with zones. The OD matrices were scaled up to match the flows of the full population. While not an ABM simulation, nevertheless, the work showed that the use of samples of the full population affects the predicted traffic flows even at the macro-scale.

Recently, Llorca and Moeckel (2019) studied the effects of scaled down populations in agent-based traffic simulations of the Munich metropolitan area. The study focused mostly on the impacts on average travel time and the distribution of travel times. The results showed that the average travel time depends on the size of the population sample and is minimum with a sample that is 10%–20% of the full population. Travel time distributions for 5% and 100% of population samples were observed to be very similar. Another interesting finding was that the scale factor for spatial length of the links (streets) did not have a strong influence on the average travel time of agents. The authors concluded that a scale factor of 5% seems reasonable for simulations where only analysis of highly aggregated results are required. However, the authors emphasized that for traffic flow analysis of a single corridor a full population sample is likely required.

3 Methodology

To evaluate the impact of scaled input data on simulation outputs, our bottom-up in-house agent-based simulation framework, EnerPol (Eser et al. 2016; Marini et al. 2018), is used. EnerPol provides integrated scenario-based assessments of energy

and transportation infrastructure, land use, and urban development. EnerPol's GPU-Enhanced Mobility Simulator, GEMSim (Saprykin et al. 2019a), is used in this study. GEMSim utilizes a massively parallel mobility model and allows one to assess, in a reasonable amount of time, "what-if" scenarios from city to country scales.

3.1 Travel Demand

The travel demand part of a scenario consists of a set of individual agents derived from the EnerPol's agent-based population model. The population model generates a synthetic population for the whole of Switzerland (approximately 8.3 million agents) using data from the Interactive Statistical Atlas of Switzerland (Federal Statistical Office 2019). The agents are then linked to geo-referenced households based on the Swiss Federal Register of Buildings and Dwellings (Federal Statistical Office 2018b), and the Swiss Business and Enterprise Register (Federal Statistical Office 2018a) is used to identify the locations of jobs of employed agents. On this basis, each agent is assigned an individual daily plan that is based on the Swiss Transportation Survey 2010 (Federal Statistical Office / Federal Office for Spatial Development 2012). The calibration was performed in the population model to match census data, and the validation with real world traffic counts is presented in prior work (Saprykin et al. 2019a).

An individual daily plan has a set of activities that an agent must perform, and the legs connecting the activities. Each activity has a spatial location and a type (that is, work, shopping, leisure, etc.), and each agent performs an activity by spending a certain amount of time at the corresponding location. A leg describes the path that an agent should take to move from the location of one activity to the location of another using either a car, public transit, or walking. The Switzerland test case that is used in this paper consists of approximately 3.5 million agents with private cars and approximately 1.7 million agents using public transit (inhabitants below the age of 6, people staying home, and car passengers are excluded from the scenario). The walk mode is a sub-mode of public transit, and is employed when it is faster to walk than to wait for a public transit vehicle.

Travel demand is executed repeatedly in a loop, where agents can learn from the previous experience and adapt their plans between the iterations. In this scenario, only re-routing strategy is allowed, and agents neither switch their mode of transport between simulated iterations nor shift departure times.

3.2 Travel Supply

The supply part of the scenario consists of a road network and a public transit schedule. The network is represented as a directed graph, where each edge represents a road segment that spans from an upstream to a downstream node (vertex), and each node may have multiple upstream or downstream links, for example at an intersection. The Switzerland test case has a network with 513 770 nodes and 1 127 775 road links, and is shown in Fig. 1a. The road network was constructed using data from OpenStreetMap (2018).

Fig. 1 Supply part of the agent-based simulation: road **a** and public transit **b** networks of Switzerland

The public transit schedule is a set of transit routes, where a transit route is defined as a set of transit stops with scheduled arrival and departure times, and departure frequencies from the terminal transit stop. Public transit vehicles can either use separate network links or share routes with private cars. The Switzerland test case has a transit schedule with 27 873 stops and 21 847 routes as shown in Fig. 1b. Public transit vehicles do not mix on the links with private cars in the present simulations. The data for the transit schedule of Switzerland is for the year of 2018, taken from Swiss Federal Railways (Swiss Open Data Platform 2018), and includes the transit routes for trains, buses, tramways and other means of public transportation.

3.3 Traffic Model

GEMSim utilizes a spatial queue model, based on Gawron (1998) and Cetin et al. (2003), and adapted for massively parallel processing on GPUs with additional improvements and optimizations. In this section, a brief description of the traffic queueing model is provided, while a more detailed description is available in Saprykin et al. (2019a). The principle of this queue model is presented in Fig. 2. A link in the network is represented by two buffers: one buffer corresponds to the physical length of a link (spatial buffer), and another buffer corresponds to the designed traffic flow (capacity buffer). The spatial buffer limits the number of vehicles that can simultaneously queue on a road segment, while the capacity buffer limits the number of vehicles that can leave the link in a time period (that is, the number of vehicles per hour).

Fig. 2 Queue model: (1) moves a vehicle from the spatial buffer to the capacity buffer, (2) moves the vehicle from the upstream link to the downstream

The size of a spatial buffer N_l is calculated as

$$N_l = \max \left\{ \left\lceil k_l \cdot N_{lanes} \cdot \frac{L_{link}}{L_{veh}} \right\rceil, 1 \right\} \quad (1)$$

where k_l is a spatial scaling coefficient, L_{link} is the physical length of the link, L_{veh} is the length of the space occupied by a vehicle (gross space), N_{lanes} is the number of lanes on the link, and $\lceil \cdot \rceil$ means integer. The size of a capacity buffer N_f is calculated as

$$N_f = \max \left\{ \left\lceil k_f \cdot q \cdot \frac{t_{cycle}}{t_{period}} \right\rceil, 1 \right\} \quad (2)$$

where k_f is a capacity scaling coefficient, q is the maximum allowed traffic flow through the link in terms of the number of vehicles per time period, t_{cycle} is the duration of a single simulation time step, and t_{period} is the duration of the time period used to define allowed traffic flow.

Both coefficients k_l and k_f are typically in the range (0; 1] and are used to scale the network for the size of the population sample. When full scale input data are used, the coefficients are equal or close to 1, while for downscaled input data the coefficients are typically equal or close to the fraction of the population sample that is used for sampling. The reason why these coefficients do not necessarily scale in direct proportion to the size of sample is that the simulation process is highly non-linear and stochastic. The dependence of these coefficients on the scale that is used is discussed later in the paper.

A vehicle moves between links in a two-stage procedure. First, the vehicle moves from the spatial buffer to the capacity buffer if the following criteria are met:

- A vehicle stayed the minimum required time on the link defined as $t_{link} = L_{link}/v_{link}$, where v_{link} is the allowed free speed.
- A vehicle is at the front of the queue in the spatial buffer.
- Enough flow capacity has been accumulated: during the simulation step t_{cycle} a link accumulates $q \cdot t_{cycle}/t_{period}$ for the flow capacity to keep the physical constraints for the designed traffic flow.

Second, when a vehicle is in the capacity buffer, it moves into the spatial buffer of the downstream link if there is enough free space. When a downstream link has multiple upstream links connected through the same node, the order in which upstream links are processed is defined to be uniformly at random and proportional to the flow capacity of each link:

$$p_i = q_i / \sum_{k=1}^n q_k \quad (3)$$

where q_i and q_k are the flow capacities of the i -th and k -th upstream links, respectively, and p_i is the probability of an upstream link being selected. This selection process introduces additional stochastic perturbation into the simulation, and multiple runs of the same scenario can produce slightly different results. The order of upstream links is updated in each simulation time step.

The same queueing model is used to simulate the movement of public transit vehicles. For each public transit vehicle a driver agent is assigned, and he/she follows

a route from the schedule, while passenger agents embark and disembark at transit stops.

3.4 Similarity Measure

In order to evaluate the impacts of using scaled input data, a similarity measure is required. This similarity measure must not only be able to compare the outputs from two simulations in relation to each other, but must also provide a quantitative measure of how close the predictions from a simulation with scaled input data are to a scenario with the full-scale population. As we also want to compare simulation outputs both spatially and temporally, the notion of a spatio-temporal point is utilized for the generalization of a similarity measure. A spatio-temporal point represents a measure taken at a certain location over a certain period of time. For traffic flow, the measure may be a set of counts from inductive loops throughout a day, and for public transit the measure may be vehicle occupancy at each stop along a route.

As the problem of uncertainty quantification of sub-sampled populations has received little attention in prior works, there are no specific measures suggested in the literature. However, one can see that a comparison of the outputs of two simulations is very close to the procedures that are used in the calibration and validation of traffic models. While these procedures qualitatively compare outputs from simulations using a common set of measures of goodness-of-fit, most of these measures have a notable limitation: that is, one has to specify a threshold that indicates whether the goodness-of-fit is considered to be acceptable or not.

Below we briefly describe the most commonly used measures of goodness-of-fit provided, and discuss their limitations and applicability to the present study. Some of the measures are subsequently evaluated if the measure qualitatively captures the same trends in the goodness-of-fit as identified in the new proposed measure which we later describe.

The mean absolute error (*MAE*), mean absolute normalized error (*MANE*), root mean squared error (*RMSE*) and root mean squared normalized error (*RMSNE*) are used by many researchers for calibration purposes (Hourdakakis et al. 2003; Balakrishna et al. 2007; Hollander and Liu 2008). *MAE* and *MANE* are insensitive to large errors, and the non-normalized *RMSE* may give a biased assessment as different roads have different traffic volumes throughout a day; notwithstanding, *RMSNE* is a good candidate for further evaluation:

$$RMSNE = \sqrt{\frac{1}{N} \cdot \sum_{i=1}^N \left(\frac{x_i - y_i}{y_i} \right)^2} \quad (4)$$

where x_i is the prediction at the i -th spatio-temporal point, y_i is an empirical or observed value at the same point, and N is the total number of evaluated spatio-temporal points.

While the *RMSNE* can show system-wide relative differences in the outputs of scenarios, the *RMSNE* does not show the nature of the differences. In that regard, Theil's inequality coefficient (Theil 1958) is a preferred measure and has been used

in numerous studies (Hourdakakis et al. 2003; Brockfeld et al. 2004):

$$U = \frac{\sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - y_i)^2}}{\sqrt{\frac{1}{N} \sum_{i=1}^N (y_i)^2} + \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i)^2}} \quad (5)$$

An inequality coefficient of 0 indicates a perfect fit, while an inequality coefficient of 1 indicates the worst possible fit. Theil's inequality coefficient can be decomposed into three parts:

$$U_m = \frac{N(\bar{y} - \bar{x})^2}{\sum_{i=1}^N (y_i - x_i)^2} \quad (6)$$

$$U_s = \frac{N(\sigma_y - \sigma_x)^2}{\sum_{i=1}^N (y_i - x_i)^2} \quad (7)$$

$$U_c = \frac{2 \cdot (1 - r) \cdot N \cdot \sigma_x \cdot \sigma_y}{\sum_{i=1}^N (y_i - x_i)^2} \quad (8)$$

where $\bar{x}$ and $\bar{y}$ are the averages of the predicted and the observed quantities X and Y , σ_x and σ_y are standard deviations of the quantities, and r is a Pearson's correlation coefficient of the quantities. The three parts, the bias proportion U_m , the variance proportion U_s , and the covariance proportion U_c indicate the sources of the differences, respectively the systematic bias, the distribution mismatch between the predicted and the observed quantities, and the non-systematic bias. A value of 0 for U_m and U_s and value of 1 for U_c indicate a perfect match between the predicted and the observed quantities. Additionally, the sum of the three parts of Theil's inequality coefficient is unity:

$$U_m + U_s + U_c = 1 \quad (9)$$

While the *RMSNE* and Theil's coefficient are generally applicable to any type of simulation model, the *GEH* measure of goodness-of-fit has been adopted by many highway agencies around the world (in USA (Dowling et al. 2004), UK (Great Britain Highways Agency 1996), Australia (Roads and Maritime Services 2013), New Zealand (NZ Transport Agency 2014), etc.) for the validation of traffic models:

$$GEH(x_i, y_i) = \sqrt{\frac{2 \cdot (x_i - y_i)^2}{x_i + y_i}} \quad (10)$$

$GEH(x_i, y_i)$ values less than 5 are considered to be indicative of a good fit; values in the range of 5 to 10 indicate that the model's outcomes are still acceptable but some inconsistency in the model's predictions is present; and, values greater than 10 require that the inconsistencies be explained in order that the model's outcomes be accepted. Typically, at least 85% of spatio-temporal points must yield $GEH(x_i, y_i)$ values less than 5 in order that a model is considered to be properly validated. Although $GEH(x_i, y_i)$ is not a real statistic, it is very similar to the chi-squared two-sample test statistic introduced by Pearson, which is for a pair of measurements x_i and y_i defined as:

$$\chi^2(x_i, y_i) = \frac{(x_i - y_i)^2}{x_i + y_i} \quad (11)$$

where $\chi^2(x_i, y_i)$ approaches a χ^2 distribution with one degree of freedom. The tested null hypothesis H_0 is that x_i and y_i come from the same distribution, and H_0 is accepted if the value of the $\chi^2(x_i, y_i)$ test statistic is less than the critical value of $\chi^2_{1,\alpha}$ distribution for a chosen significance level α (typically 0.05) and one degree of freedom, otherwise the null hypothesis H_0 is rejected. As predicted outputs that are inherently variable must be compared, it seems reasonable to use the GEH measure of goodness-of-fit to assess similarity with a selected significance level. However, even though GEH has been adopted by many highway agencies for the validation of predictions compared to field data, as discussed below, we do not consider GEH to be a good similarity measure for two predicted outputs.

From (10) and (11) one can obtain:

$$\chi^2_{GEH}(x_i, y_i) = \frac{GEH(x_i, y_i)^2}{2} \quad (12)$$

where $\chi^2_{GEH}(x_i, y_i)$ approaches a χ^2 distribution with one degree of freedom. It can be shown that a given critical value GEH_{th} used to test $GEH(x_i, y_i)$, with its corresponding value of $\chi^2_{GEH,th}$, acts as a non-linear scaling factor for the critical value of $\chi^2_{1,\alpha}$ that is used to test the chi-squared statistic for the same pair of quantities x_i and y_i :

$$K_{GEH} = \frac{\chi^2_{GEH,th}}{\chi^2_{1,\alpha}} = \frac{GEH_{th}^2}{2 \cdot \chi^2_{1,\alpha}} \quad (13)$$

In other words, the use of the GEH formula is similar to using the chi-squared test statistic, but with the critical value of $\chi^2_{1,\alpha}$ corresponding to the different (and shifted) significance level α , and K_{GEH} establishes the relation between these two different critical values of the χ^2_1 distribution. Essentially, the value of $K_{GEH} > 1$ decreases (and shifts) the initial significance level α (that is rate of type I errors) in the chi-squared test, therefore increasing the rate of type II errors significantly. Moreover, the same initial (and unshifted) value of α is used to make the decision on the goodness-of-fit for the model that is being validated: the number of spatio-temporal points that needs to satisfy the scaled critical value of $\chi^2_{1,\alpha}$ is not increased due to the decreased α . It is for this reason that we consider that GEH is not a real statistic.

For example, the recommended value of $GEH = 5$ with an initial critical value of $\chi^2_{1,\alpha} = 2.07$ for a significance level $\alpha = 0.15$ (that is, the requirement that at least 85% of the spatio-temporal points pass the GEH test) gives approximately $K_{GEH} = 6$, thus shifting α to 0.00041 while ensuring that 85% of points pass the GEH test. Similarly, a threshold value of $GEH = 4$ with at least 95% of the points passing the GEH test (that is, initial $\alpha = 0.05$), gives $K_{GEH} = 2$ and a shifted $\alpha = 0.00468$.

Another interesting interpretation of the GEH formula can be obtained from the following:

$$\chi^2(x_i, y_i) = \frac{(x_i - y_i)^2}{2 \cdot E_{xy,i}} = \frac{(x_i - E_{xy,i})^2}{E_{xy,i}} + \frac{(y_i - E_{xy,i})^2}{E_{xy,i}} = \frac{2 \cdot \sigma_{xy,i}^2}{E_{xy,i}} \quad (14)$$

where $E_{xy,i}$ and $\sigma_{xy,i}^2$ are average and variance of x_i and y_i . The variance $\sigma_{xy,i}^2$ is multiplied by a factor of 2 because both x_i and y_i are random variables. Combining (13) and (14) gives:

$$K_{GEH} = \frac{\chi_{GEH,th}^2}{\chi_{1,\alpha}^2} = \frac{\sigma_{GEH,th}^2 \cdot E_{1,\alpha}}{E_{GEH,th} \cdot \sigma_{1,\alpha}^2} \Big|_{E_{1,\alpha}=E_{GEH,th}} = \frac{\sigma_{GEH,th}^2}{\sigma_{1,\alpha}^2} \quad (15)$$

where $\sigma_{GEH,th}^2$ and $\sigma_{1,\alpha}^2$ are variances for the corresponding critical values of $\chi_{GEH,th}^2$ and $\chi_{1,\alpha}^2$ conditional on the equality of corresponding means $E_{GEH,th}$ and $E_{1,\alpha}$. Equation (15) shows that GEH with $K_{GEH} > 1$ yields model predictions with a higher variance than allowed with the chi-squared statistic, and over-dispersion is exactly defined by K_{GEH} .

The lack of clarity in how the GEH_{th} values that are suggested as critical for traffic model validation are determined casts doubt on the general applicability of the GEH formula as a good similarity measure for the outcomes from simulations. Neither of these GEH_{th} values can be considered to be a reliable statistical measure. Instead, the chi-squared statistic test is considered to be more suitable for assessing similarity, and this statistic is perhaps the most widely used statistic with well-known properties. Moreover, the chi-squared test allows one to choose a significance level α , making this measure of goodness-of-fit more widely applicable. Nevertheless, it worthwhile to see if the GEH measure is able to capture the trends in similarity, as well.

Considering the above, we propose the following similarity measure for a given significance level α :

$$S_\alpha = \min \left\{ \frac{\sum_{i=1}^N \delta_i}{N \cdot (1 - \alpha)}, 1 \right\} \quad (16)$$

where

$$\delta_i = \begin{cases} 1, & \text{if } \chi^2(x_i, y_i) < \chi_{1,\alpha}^2 \\ 0, & \text{otherwise.} \end{cases} \quad (17)$$

S_α has a minimum of 0 when none of spatio-temporal points pass the chi-squared test, and a maximum of 1 when at least $(1 - \alpha) * 100\%$ of spatio-temporal points pass the test. Thus, S_α represents a straightforward statistical and quantitative approach to measure the similarity of outputs from mobility simulations. This similarity measure is generic enough and can be applied to both traffic flows and occupancy of public transit vehicles.

Thus below we further evaluate the following four measures: $RMSNE$, Theil's inequality coefficient, GEH and the proposed S_α .

3.5 Scaling Scenario Input and Output

To investigate the impact of scaling one's input data, the following set of population samples is used: 1%, 2%, 5%, and 10% to 90% with 10% intervals. Uniformly random sampling of the full population was applied, and the simulation outputs with the full population (that is, 100% population sample) are used as the reference (that is, quantity Y). Hence, all simulated population samples are compared with respect to

the full population sample simulated as a reference case. A 30 hour period, starting from midnight, is simulated. As the dependence of the network parameters k_l and k_f on the size of population sample is unknown, a calibration (impedance matching) was performed for each of the samples to match the simulation outputs of the reference case as closely as possible. Two sets of spatio-temporal points were chosen for calibration of car traffic: counts on all car-related network links during the morning (07:00–08:00) and the evening (17:00–18:00) peak hours. During simulation, the traffic counts are aggregated over 15-minute time intervals, and therefore for each peak hour each network link has four distinct spatio-temporal points which are used in the assessment of similarity. The use of multiple spatio-temporal points on each link allows the temporal dynamics of traffic flows to be compared rather than limiting the comparison to only time-averaged states. In addition to the peak hours, two other sets of spatio-temporal points were used to evaluate similarity measures with calibrated coefficients: noon hour (12:00–13:00) when the traffic is relaxed and daily (00:00–24:00) aggregated values. Network links where no cars pass in the reference case are excluded from the analysis. As public transit is not scaled (neither the network links nor the schedule), it need not be calibrated. The occupancy of each public transit vehicle after its departure from each stop are used as the set of spatio-temporal points to assess the similarity of simulated public transit. The traffic and occupancy counts from simulations with population samples are scaled up by the inverse of the sampling fraction size to match the counts from the reference case.

The determination of the optimal values k_l^* and k_f^* for the calibrated parameters k_l and k_f is the subject of the following optimization problem:

$$\begin{aligned} \max \quad & S_\alpha^m(k_l, k_f) + S_\alpha^e(k_l, k_f) \\ \text{s.t.} \quad & \{k_l, k_f\} \in (0; 1] \end{aligned} \quad (18)$$

where a significance level α of 0.05 was chosen, and S_α^m and S_α^e are the similarity measures for the morning and evening peak hours, respectively. To reduce the number of required simulations for the calibration, the search space for values k_l^* and k_f^* was adjusted depending on the sampling fraction size as follows:

- $\{k_l, k_f\} \in [0.1; 1]$ with a step of 0.1 for population samples from 10% to 90%;
- $\{k_l, k_f\} \in [0.005; 0.015]$ with a step of 0.001 for the 1% population sample;
- $\{k_l, k_f\} \in [0.01; 0.1]$ with a step of 0.01 for the population samples of 2% and 5%.

Additionally, the small population samples (less than 10%) were examined with a broader range of k_l and k_f , but the variation in the optimization objective was in the range of 1%. Similarly, larger population samples did not show significantly different behaviour for intermediate scaling ratios. In total, 1 221 simulations were performed for the whole set of population samples. After calibration the k_l^* and k_f^* that were determined for each population sample were used to calculate similarity measures of car traffic for the population sample. This procedure was repeated for five times with different sets of the random population samples and the results of the similarity measures were averaged to smooth out fluctuations. Thus, a total of 6 105 simulation runs were performed.

To determine the number of iterations to run for each of the population samples, the average score of the agents between iterations was examined for different sample sizes. Figure 3a shows that larger population samples require more iterations to converge the average score, while samples smaller than 10% require only a few iterations. The same behaviour, when the number of iterations required for a simulation to converge depends on the population sample size, was previously reported by Llorca and Moeckel (2019). Moreover, use of larger population samples may lead to oscillations of the score between iterations. Figure 3b shows a zoomed-in part of the previous plot, after 80 iterations, when agents stopped innovating their daily plans and only choose one of the previously memorized daily plans with the best scores. After 100 iterations, the difference of the average score between 1% and 100% samples is less than 1%. However, after 20 iterations, there is only marginal improvement for the average score even for large sub-samples. Hence, sub-scaled scenarios were run for 20 iterations with re-routing strategy followed by 5 iterations with memorized plans only.

4 Results

The mean, standard deviation (shaded area) and 95% confidence intervals using t-distribution (vertical bars) for the optimal values of the scaling coefficients k_l^* and k_f^* are shown in Fig. 4. The standard deviation and confidence intervals are obtained based on five runs for each of the combinations of k_l and k_f during the calibration process as described above. It is interesting to note that the capacity coefficient k_f^* matches exactly the scale of the input data for almost all population samples; thus, it is evident that the flow capacity is the main driver of the traffic dynamics in the traffic model. For almost all population samples, the standard deviation of k_f^* is either zero or very close to zero, therefore the flow capacity coefficient is rather stable and is not affected by traffic fluctuations between simulation runs.

Fig. 3 Average score of agents during the simulation: **a** overview of the whole simulation, **b** zoomed-in part (dashed) after 80 iterations

Fig. 4 Optimized scaling coefficients for spatial **a** and capacity **b** buffers of the network

The spatial coefficient k_l^* is close to 1 for population samples in the range of 20%–50%, and decreases for larger sized population samples. However, for very small population samples, less than 10%, k_l^* decreases substantially, and does not match the overall input scale factor. This can be explained as follows: the more the network is scaled down, the more the performance of spatial buffers degrades due to rounding errors when smaller values of k_l are applied; on the other hand, larger k_l better matches the impedance of the reference case. The same behaviour is reported by Llorca and Moeckel (2019) where spatial buffers are scaled with factors larger than the corresponding size of the population sample. The standard deviation of k_l^* is larger than the standard deviation of k_f^* , therefore indicating that spatial buffers are more sensitive to stochastic fluctuations in the simulated traffic.

To show the impact of rounding errors on the simulated traffic flow, let's consider the following example with a network link of $N_l = 8$ (about 60 meters long) and $N_f = 1$ (flow capacity of 900 cars/hour) which is very typical for the city of Zurich. For the sake of simplicity, we assume that the flow of this link is fully matched with downstream links such that there is always free space available in downstream spatial buffers. Further, we consider a stable flow of 900 cars/hour from upstream links, or 1 car in each 4 seconds of simulated time. In a non-scaled scenario, this link should never be overflown, and no spill-over should occur under the assumed conditions. In the first scenario, a 40% sample is used, and, according to (1)–(2), link buffers are scaled to $N_{l,0.4} = 3$ and $N_{f,0.4} = 1$, whereas the link accumulates 0.4 cars of flow capacity in each 4 seconds. In the second scenario, a 70% sample is used, and link buffers are scaled to $N_{l,0.7} = 5$ and $N_{f,0.7} = 1$, whereas the link accumulates 0.7 cars of flow capacity in each 4 seconds. As soon as the link has accumulated flow capacity ≥ 1.0 , it can release 1 car into the capacity buffer and downstream consequentially. Table 1 shows the state of the spatial buffer of the link during a simulation when the upstream flow is kept constant (1 car in each 4 seconds). Since a flow of a single car cannot be downscaled, there exists a probability that for a certain time period a flow

Table 1 State of the capacity buffer of a scaled down link

Step	40% sample		70% sample	
	Flow capacity (0.4/step)	Queue state ($N_{l,0.4} = 3$)	Flow capacity (0.7/step)	Queue state ($N_{l,0.7} = 5$)
1	0.4	1 (0)	0.7	1 (0)
2	$0.4 + 0.4 = 0.8$	$1 + 1 = 2(0)$	$0.7 + 0.7 - 1.0 = 0.4$	$1 - 1 + 1 = 1(0)$
3	$0.8 + 0.4 - 1.0 = 0.2$	$2 - 1 + 1 = 2(0)$	$0.4 + 0.7 - 1.0 = 0.1$	$1 - 1 + 1 = 1(0)$
4	$0.2 + 0.4 = 0.6$	$2 + 1 = 3(0)$	$0.1 + 0.7 = 0.8$	$1 + 1 = 2(0)$
5	$0.6 + 0.4 - 1.0 = 0.0$	$3 - 1 + 1 = 3(0)$	$0.8 + 0.7 - 1.0 = 0.5$	$2 - 1 + 1 = 2(0)$
6	$0.0 + 0.4 = 0.4$	$3 + 1 = \mathbf{3(1)}$	$0.5 + 0.7 - 1.0 = 0.2$	$2 - 1 + 1 = 2(0)$
...	...	...	...	...
15	$0.6 + 0.4 - 1.0 = 0.0$	$9 - 1 + 1 = 3(6)$	$0.8 + 0.7 - 1.0 = 0.5$	$5 - 1 + 1 = 5(0)$
16	$0.0 + 0.4 = 0.4$	$9 + 1 = 3(7)$	$0.5 + 0.7 - 1.0 = 0.2$	$5 - 1 + 1 = 5(0)$
17	$0.4 + 0.4 = 0.8$	$10 + 1 = 3(8)$	$0.2 + 0.7 = 0.9$	$5 + 1 = \mathbf{5(1)}$

of 900 cars/hour is kept. In Table 1, a simulation step equals to 4 seconds to keep the calculations simple.

Table 1 shows that the spatial buffer of the link is overflowed in only 6 steps (24 simulated seconds) when using a 40% population sample with the network scaled down accordingly. However, when using a 70% population sample, it takes 17 steps (68 simulated seconds) to overflow the link. In total, after 17 steps, the link with a smaller scaling factor of 0.4 spills over 8 cars causing congestion in upstream links. Hence, rounding errors in downscaled network links can severely affect congestion patterns and traffic dynamics during the simulation.

Additional insight into the nature of the impacts of k_l and k_f is given in Fig. 5, where the optimization objective for one of the sets of samples (not averaged across five sets) is shown. The characteristics of the optimization objective, given in (18), in the $k_l - k_f$ plane are sensitive to the size of the population sample that is used. For very small population samples of 1% the objective is both far in similarity compared to the reference case and the surface of the objective is flat. With this flat surface, there is a high likelihood that many different combinations of k_l and k_f may be optimal. Thus, with the very small population sample, the scaled flow capacity of the network does not drive the traffic dynamics, and the model generates mostly noise rather than proper traffic dynamics. When the size of the population sample is increased to 10%, the model is more similar to the reference case in terms of the traffic dynamics, and the flow capacity of the buffers generate more realistic traffic flows. With a 50% population sample the model is almost fully driven by the flow capacity and generates traffic that is quite similar to the reference case. It is also clear that the k_l coefficient does not significantly affect the simulation outputs. With a 90% population sample, the model is sensitive to both the flow capacity and the spatial length of the links; in this regard k_l effectively fine-tunes the traffic dynamics.

As the size of the population sample increases, the objective (traffic similarity) improves, and the surface of the objective in the $k_l - k_f$ plane becomes more

Fig. 5 Normalized surface of the optimization objective during the calibration procedure: **a** 1% of population, **b** 10% of population, **c** 50% of population, **d** 90% of population

concave. The concave surface indicates that the objective is more sensitive to changes of the scaling coefficients, therefore the optimal values of k_l and k_f are closer to the scaling factor when the surface is concave. Nevertheless, as the simulations are highly non-linear, the optimum coefficient k_l^* for spatial buffers is smaller than the corresponding scale factor for the flow capacity. Optimum coefficients for spatial and capacity buffers, evaluated with different sized population samples, are presented in Table 2 (Appendix A).

The measures of goodness-of-fit for the different population samples are presented in Figs. 6–7. In the legend, GEH_5 and GEH_{10} indicate that the links satisfy, respectively, the conditions $5 \leq GEH(x_i, y_i) < 10$, and $GEH(x_i, y_i) \geq 10$, and HW indicates that the similarity measure is only applied to a subset of network links consisting from motorway and expressway links with a minimum speed of 80 km/h. The GEH measure is given as a percentage of the total number of links that are analyzed.

Fig. 6 Measures of goodness-of-fit (hourly time periods): **a** Theil's inequality coefficient U , **b** RMSNE, **c** GEH as % of total links, **d** proposed S_α measure

A total of 1 033 642 links, including 67 483 motorway and expressway links, were analyzed.

As expected, the GEH measure satisfies the validity condition $GEH < 5$ for at least 85% of the links even with the 5% population sample during hourly evaluated periods (morning, noon and evening), while the traffic similarity S_α based on the chi-squared statistic yields statistically significant results only with a population sample of at least 30% in the morning, 60% in the noon and 40% in the evening. The lower similarity in the noon and evening peak hours is only observed for motorways and expressways, while for the whole set of links a statistically significant result is already achieved with a population sample of at least 30%. Motorways and expressways are also affected by discrepancies in traffic dynamics when small population samples are used. However, one may not intuitively expect this result considering that links with smaller volumes are affected more by traffic fluctuations, but

simulations show that both road types are affected by the size of the population samples. Nevertheless, lower traffic volumes on highways in the noon could be the reason for higher fluctuations and lower similarity. The same trend is also observed for daily aggregated counts in Fig. 7, where neither GEH nor S_α reaches a statistically significant result, as the error is accumulated throughout a day for each of the links. It is also noteworthy that as the size of the population sample is increased from 1% to 30%, the GEH measure decreases sharply from more than 30% to less than 1% (except daily evaluated period where errors are accumulated), and then decreases only slightly up to a population sample of 90%. Therefore, GEH captures qualitatively a critical point with a population sample of 30%, while S_α captures this point quantitatively. The shape of curves for the GEH measure with the GEH_{10} condition decreases more rapidly than the curves for the GEH measure with the GEH_5 condition as the more heavily penalized links fit better to the reference distribution as the size of the population sample is increased.

Fig. 7 Measures of goodness-of-fit (daily time periods): **a** Theil's inequality coefficient U , **b** RMSNE, **c** GEH as % of total links, **d** proposed S_α measure

All other measures of goodness-of-fit qualitatively capture the impacts of the size of the population sample on traffic dynamics. However, Theil's inequality coefficient highlights a critical point when the size of the population sample is in the range of 10%–20%, and further increases in the size of the population sample lead to only minor improvements. One can also note that for the noon hour the errors are higher, as was previously mentioned. More detailed data of the three component parts of Theil's inequality coefficient are presented in Tables 3–6 (Appendix B) for cars and in Table 7 (Appendix B) for public transit. These data show that there is a larger systematic bias and a larger mismatch in the distribution for small population samples. Such errors can be expected when small traffic volumes are scaled up by large multiples. For large sized population samples, non-systematic errors contribute more than the systematic bias and mismatch of distribution, as the overall match of traffic dynamics and transit occupancy is better.

The *RMSNE* also shows that there is a tendency for the error to reduce with larger population samples, but it is very difficult to interpret these results. For example, the *RMSNE* is below 2 for cars when a 5% population sample is used, as the simulation results are averaged over a large number of links when the *RMSNE* is evaluated.

Although the public transit infrastructure is not scaled using k_l and k_f coefficients and public transit vehicles do not share the roads with other cars, in general, all similarity measures indicate that public transit has higher uncertainties compared to the peak hours, however the similarity stays almost unaffected in a daily evaluated period. Public transit vehicles always run according to the schedule, and therefore are unaffected by the queueing model (that is, no congestion forms, and vehicles can always move forward). Thus, we can infer that discrepancies in the predicted traffic dynamics that result from the use of small population samples do not depend mainly on the traffic model that is used, but rather are a consequence of the sub-scale population itself which distorts the demand.

Figure 8 and Table 8 (Appendix C) show the performance of the simulated car transport in terms of the total vehicle-hours travelled (*VHT*) and total

Fig. 8 Performance of the simulated car transport: **a** vehicle-hours travelled (VHT), **b** vehicle-kilometres travelled (VKT)

vehicle-kilometres travelled (*VKT*) during hourly and daily evaluated periods; the performance is normalized relative to the reference case. The simulations with a 5% population sample give errors in the range of 5%–6%, and the errors are 2%–3% for simulations with 10%–20% population samples.

It is worthwhile to note that *VKT* varies much less than *VHT* and has an error no more than 6%, even with a 1% population sample, while *VHT* has errors in the range of 7%–30% for small sized population samples. A possible explanation is that the distribution of trip distances performed by the agents is more narrow and concentrated around the mean value, therefore it is easier to approximate the distribution with a few agents. On the other hand, the distribution of trip durations is broader, and the use of a small population sample gives a larger error. However, it is also evident that discrepancies in the measures of goodness-of-fit depend on the patterns of behaviour of the simulated population, and these patterns can be different in other scenarios.

The runtime performance of simulations with different population sub-samples was evaluated on a cluster node with two Intel Xeon E5-2620 v4 CPUs clocked at 2.60 GHz, 256 GB of RAM and four NVIDIA P100 (16 GB RAM onboard) GPUs. In total, 20 CPU threads have been used to run a scenario. Runtime of one simulated daily iteration for each of the population sub-samples is shown in Fig. 9. The runtime performance scales non-linearly, that is, when the size of a sub-sample increases from 20% to 40%, the runtime increases from 220 to 350 seconds only. The reason for this behaviour is the structure of a GPU-accelerated simulation loop. As it was shown in prior work (Saprykin et al. 2019b), there are two main performance bottlenecks: (i) data transfers between GPU and the host side, and (ii) re-routing of the agents between iterations. Both bottlenecks almost linearly depend on the size of a simulated population, but the traffic propagation part which is accelerated with a GPU has a non-linear runtime dependency (Saprykin et al. 2019a). Thus, the overall runtime of a simulated iteration has a non-linear dependency on the size of a population sample.

Fig. 9 Runtime performance of one simulated iteration

5 Conclusions and Discussion

The GPU-enhanced spatial queuing model is seen to be more sensitive to the traffic flow coefficient k_f rather than the spatial coefficient k_l . Thus, in the calibration of a simulation that uses a sampled population the optimal flow and spatial coefficients can be more efficiently determined: specifically, the search space for the optimum k_f can be reduced by using a narrower search range for k_f . As the optimum flow coefficient is within a known narrow band, only a few steps are required in the k_f dimension. On the other hand, as the spatial coefficient does not have a strong impact on traffic dynamics, coarser step-sizes can be used in the k_l dimension.

All commonly used measures of goodness-of-fit, as well as our proposed measure, capture the same trends in regards to the impacts of using sampled populations: namely, smaller population samples yield traffic dynamics that are more dissimilar than the traffic dynamics in a simulation that uses the full population. Our proposed measure of goodness-of-fit S_α shows that there is a critical size of approximately 30% of the full population, below which the simulated traffic dynamics are markedly different. Therefore, as a trade-off between the quality of the simulated traffic dynamics and the required computational power, a 30% population sample will give predictions that are insignificantly different from a simulation that uses the full population. The dynamics of the occupancy of public transit are also accurately captured in simulations with a 30% population sample. However, if high accuracy in the predicted traffic dynamics and in the occupancy of public transit are not desired, and the main interest is only aggregated parameters such as VHT and/or VKT , with errors in the range of 2%–6% being acceptable, then smaller population samples, in the range of 5%–10% of the full population, can be used. Nevertheless, as emerging technologies continue to increase the complexity of modern transportation systems, the impact of the size of the population sample needs to be always considered. One example is the transition to electric mobility, where simulations with a 30% population sample will yield accurate predictions of the traffic dynamics, but may not necessarily accurately capture the spatio-temporal characteristics of the electricity demand for the charging of EVs. Only simulations with the full population will unambiguously yield accurate predictions related to both the transportation and electricity infrastructures.

In the simulations with sampled populations, whereas for private traffic both the supply and demand are scaled, in the case of public transit only the demand is scaled. The measures of goodness-of-fit show the same trends for both private traffic and public transit in regards to the impacts of using sampled populations. A public transit system can be seen here as a case of an ideally scaled road network which does not impact travel times of the agents. It is thus evident that it is the scaling of demand which is the main reason for distortions in the simulated traffic dynamics. While a scaled traffic queueing model may also impact traffic dynamics, the impacts can be reduced during the calibration process. We expect that scenarios with other types of agent-based traffic models will also be adversely impacted by the scaling of input data.

In the study, five samples from the full population for each of the scaling factors were used to smooth out fluctuations caused by the probabilistic traffic model (that is, the order in which upstream links are processed is randomized in each simulation step). However, each sample may itself introduce uncertainties in the simulation results. Hence, in future research, it would be interesting to quantify the impact of a single population sample on such uncertainties. To do that, one can simulate a scenario for each population sample multiple times with different random seeds and then average the simulation results. While this would require even more computing power, one would be able to separate per-sample uncertainties from the other uncertainties due to the use of scaling factors.

The results presented in the paper are obtained using the large-scale scenario of Switzerland. It would be interesting to see similar studies for other areas to better understand the impacts of different network topologies, for example how a grid network structure affects the results. Second, as the public transit vehicles were not mixed with the rest of the traffic and were not delayed in this study, one can do further analysis of scalability issues in the scenarios where public transit vehicles do share roads with other cars and trucks. Finally, other re-planning strategies than re-routing may also impact the scaling process differently. For example, departure time mutation or switch of a transport mode would be interesting to study in the context of the scaling issues.

Appendix A: Calibration Coefficients

Table 2 Scaling coefficients for spatial and capacity buffers of the network

Sample, %	$\overline{k_l^*}$	$\overline{k_f^*}$	$\sigma[k_l^*]$	$\sigma[k_f^*]$
1	0.0086	0.0110	0.0025	0.0012
2	0.0680	0.0200	0.0192	0.0000
5	0.0920	0.0500	0.0130	0.0000
10	0.6200	0.1000	0.2588	0.0000
20	0.9000	0.2200	0.1414	0.0447
30	0.9400	0.3000	0.0548	0.0000
40	0.9600	0.4000	0.0548	0.0000
50	0.9800	0.5000	0.0447	0.0000
60	0.7800	0.6000	0.2049	0.0000
70	0.5600	0.7000	0.0548	0.0000
80	0.7000	0.8000	0.0707	0.0000
90	0.8000	0.9000	0.0707	0.0000

Appendix B: Measures of Goodness-of-Fit

Table 3 Measures of goodness-of-fit for the morning peak hour (07:00–08:00)

Sample, %	U	U_s	U_m	U_c	$RMSNE$	$GEH_5, \%$	$GEH_{10}, \%$	S_α^m	$S_\alpha^{m,hw}$
1	0.1970	0.0022	0.0113	0.9864	2.7663	19.77	13.58	0.51	0.23
2	0.1441	0.0021	0.0035	0.9944	1.9397	21.62	5.85	0.55	0.33
5	0.0932	0.0003	0.0078	0.9918	1.2105	13.61	0.79	0.65	0.50
10	0.0694	0.0064	0.0083	0.9852	0.8493	5.42	0.11	0.77	0.66
20	0.0543	0.0036	0.0027	0.9936	0.5756	1.19	0.05	0.92	0.84
30	0.0370	0.0001	0.0038	0.9958	0.4278	0.11	0.02	1.00	0.97
40	0.0310	0.0003	0.0058	0.9935	0.3442	0.04	0.01	1.00	1.00
50	0.0258	0.0002	0.0042	0.9950	0.2819	0.02	0.01	1.00	1.00
60	0.0203	0.0002	0.0034	0.9954	0.2321	0.01	0.00	1.00	1.00
70	0.0169	0.0001	0.0025	0.9959	0.1879	0.01	0.00	1.00	1.00
80	0.0143	0.0000	0.0025	0.9954	0.1464	0.01	0.00	1.00	1.00
90	0.0111	0.0002	0.0042	0.9920	0.1033	0.00	0.00	1.00	1.00

Table 4 Measures of goodness-of-fit for the noon hour (12:00–13:00)

Sample, %	U	U_s	U_m	U_c	$RMSNE$	$GEH_5, \%$	$GEH_{10}, \%$	S_α^m	$S_\alpha^{m,hw}$
1	0.2610	0.0013	0.0920	0.9067	3.3671	14.24	10.58	0.67	0.21
2	0.1954	0.0026	0.0625	0.9348	2.3822	15.95	4.58	0.70	0.30
5	0.1321	0.0067	0.0506	0.9427	1.4629	10.89	0.77	0.76	0.47
10	0.1041	0.0096	0.0419	0.9485	1.0027	4.77	0.15	0.84	0.63
20	0.0810	0.0168	0.0573	0.9259	0.7208	1.38	0.08	0.93	0.81
30	0.0748	0.0264	0.0702	0.9034	0.6504	0.83	0.08	0.97	0.87
40	0.0719	0.0305	0.0845	0.8849	0.5754	0.64	0.07	0.99	0.91
50	0.0698	0.0324	0.0851	0.8824	0.5310	0.53	0.08	1.00	0.94
60	0.0710	0.0334	0.0887	0.8778	0.4971	0.51	0.09	1.00	0.95
70	0.0653	0.0394	0.0963	0.8642	0.4704	0.47	0.08	1.00	0.97
80	0.0560	0.0495	0.1056	0.8448	0.4494	0.41	0.05	1.00	0.98
90	0.0550	0.0531	0.1130	0.8338	0.4383	0.39	0.05	1.00	0.99

Table 5 Measures of goodness-of-fit for the evening peak hour (17:00–18:00)

Sample, %	U	U_s	U_m	U_c	$RMSNE$	GEH_5 , %	GEH_{10} , %	S_α^e	$S_\alpha^{e,hw}$
1	0.1856	0.0052	0.0124	0.9824	2.7455	19.23	14.49	0.52	0.24
2	0.1357	0.0061	0.0020	0.9919	1.8854	20.81	6.71	0.57	0.33
5	0.0910	0.0002	0.0204	0.9794	1.1995	13.74	1.09	0.66	0.50
10	0.0723	0.0023	0.0021	0.9955	0.8389	6.06	0.33	0.77	0.65
20	0.0622	0.0013	0.0255	0.9731	0.5827	1.76	0.29	0.91	0.82
30	0.0477	0.0042	0.0372	0.9585	0.4414	0.46	0.17	0.99	0.94
40	0.0445	0.0074	0.0539	0.9385	0.3632	0.37	0.17	1.00	0.98
50	0.0405	0.0088	0.0610	0.9300	0.3039	0.32	0.17	1.00	1.00
60	0.0356	0.0093	0.0574	0.9330	0.2603	0.29	0.15	1.00	1.00
70	0.0293	0.0114	0.0590	0.9291	0.2212	0.28	0.07	1.00	1.00
80	0.0221	0.0097	0.0457	0.9438	0.1808	0.14	0.04	1.00	1.00
90	0.0135	0.0034	0.0153	0.9791	0.1261	0.04	0.01	1.00	1.00

Table 6 Measures of goodness-of-fit for a whole day (00:00–24:00)

Sample, %	U	U_s	U_m	U_c	$RMSNE$	GEH_5 , %	GEH_{10} , %	S_α^m	$S_\alpha^{m,hw}$
1	0.0782	0.0296	0.0746	0.8957	2.3386	31.14	27.59	0.22	0.19
2	0.0637	0.0439	0.0857	0.8703	1.6581	32.96	16.27	0.29	0.24
5	0.0547	0.0716	0.1398	0.7884	1.0607	25.54	8.23	0.40	0.28
10	0.0473	0.0651	0.1309	0.8038	0.7403	16.46	4.74	0.55	0.36
20	0.0451	0.0732	0.1488	0.7778	0.5306	11.84	3.90	0.68	0.40
30	0.0480	0.0935	0.1718	0.7346	0.4659	13.55	4.61	0.69	0.28
40	0.0476	0.0971	0.1810	0.7217	0.4146	13.40	4.47	0.71	0.28
50	0.0471	0.0980	0.1781	0.7237	0.3802	13.26	4.33	0.73	0.27
60	0.0468	0.0987	0.1782	0.7230	0.3539	13.14	4.26	0.73	0.28
70	0.0461	0.1026	0.1806	0.7166	0.3359	13.14	4.21	0.74	0.27
80	0.0456	0.1047	0.1776	0.7175	0.3205	13.16	4.14	0.74	0.27
90	0.0452	0.1042	0.1730	0.7227	0.3082	13.03	4.07	0.74	0.27

Table 7 Measures of goodness-of-fit for public transit occupancy

Sample, %	U	U_s	U_m	U_c	$RMSNE$	$GEH_5, \%$	$GEH_{10}, \%$	S_{α}^{pt}
1	0.4083	0.0187	0.0048	0.9765	4.5617	23.93	9.99	0.49
2	0.2536	0.0000	0.0708	0.9291	2.0408	24.21	3.76	0.53
5	0.1650	0.0001	0.0263	0.9737	1.3663	13.54	0.47	0.65
10	0.1150	0.0001	0.0165	0.9834	1.0205	4.74	0.03	0.79
20	0.0779	0.0001	0.0066	0.9933	0.8017	0.68	0.02	0.94
30	0.0591	0.0000	0.0025	0.9974	0.6759	0.09	0.01	1.00
40	0.0477	0.0000	0.0025	0.9974	0.6185	0.03	0.01	1.00
50	0.0391	0.0000	0.0021	0.9978	0.5245	0.02	0.00	1.00
60	0.0323	0.0000	0.0007	0.9991	0.5216	0.02	0.00	1.00
70	0.0263	0.0000	0.0005	0.9992	0.5454	0.02	0.01	1.00
80	0.0205	0.0000	0.0005	0.9990	0.5027	0.02	0.00	1.00
90	0.0153	0.0000	0.0002	0.9990	0.5968	0.02	0.01	1.00

Appendix C: Performance of Cars

Table 8 Performance of the simulated car transport

Sample, %	Morning (07:00–08:00)		Noon (12:00–13:00)		Evening (17:00–18:00)		Daily (00:00–24:00)	
	$VHT, \%$	$VKT, \%$	$VHT, \%$	$VKT, \%$	$VHT, \%$	$VKT, \%$	$VHT, \%$	$VKT, \%$
1	123.90	95.22	131.85	99.16	132.59	94.68	133.05	100.42
2	119.19	97.38	119.43	99.08	123.52	95.56	121.47	100.11
5	106.32	99.64	102.41	99.78	106.89	101.03	101.50	100.21
10	102.58	99.61	99.31	99.80	104.92	100.60	100.20	100.02
20	98.32	100.35	95.53	100.22	97.33	103.07	93.93	100.20
30	99.51	100.37	95.32	100.21	98.32	102.41	94.68	100.21
40	99.27	100.37	95.63	100.28	97.30	102.67	94.17	100.23
50	99.14	100.28	95.54	100.22	96.81	102.64	94.00	100.17
60	99.30	100.24	95.80	100.40	96.54	102.28	94.22	100.17
70	99.24	100.08	96.72	100.29	96.71	101.77	94.78	100.06
80	99.57	100.05	97.65	100.18	98.02	101.21	96.31	100.08
90	99.80	99.99	98.95	100.19	99.09	100.52	97.98	100.00

Reference case for *VHT* (vehicle-hours): 219 469 in the morning, 107 355 in the noon, 359 633 in the evening and 3 217 881 in a day. Reference case for *VKT* (vehicle-km): 12 387 429 in the morning, 6 068 409 in the noon, 16 899 257 in the evening and 168 031 450 in a day.

Acknowledgements We thank the anonymous reviewers for their valuable comments which helped to improve the manuscript. This work was supported by an ETH Zurich ISTP Research Incubator Grant.

Funding Open Access funding provided by ETH Zurich

Declarations

Conflict of Interests The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abar S, Theodoropoulos GK, Lemarinier P, O'Hare GM (2017) Agent based modelling and simulation tools: a review of the state-of-art software. *Comput Sci Rev* 24:13–33
- Bakker S, Trip JJ (2013) Policy options to support the adoption of electric vehicles in the urban environment. *Transp Res Part D: Transp Environ* 25:18–23
- Balakrishna R, Antoniou C, Ben-Akiva M, Koutsopoulos H, Wen Y (2007) Calibration of microscopic traffic simulation models: Methods and application. *Transp Res Record: J Transpo Res Board* 1999:198–207. <https://doi.org/10.3141/1999-21>
- Bekhor S, Dobler C, Axhausen K (2011) Integration of activity-based and agent-based models: case of Tel Aviv, Israel. *Transp Res Record: J Transp Res Board* 2255:38–47. <https://doi.org/10.3141/2255-05>
- Ben-Dor G, Dmitrieva B, Maciejewski M, Bischoff J, Ben-Elia E, Benenson I (2017) 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*
- Beykaei SA, Miller E (2017) Testing uncertainty in ILUTE—an integrated land use-transportation micro-simulation model of demographic updating. *J Civil Environ Eng* 7:1
- Biresselioglu ME, Kaplan MD, Yilmaz BK (2018) Electric mobility in Europe: a comprehensive review of motivators and barriers in decision making processes. *Transp Res Part A: Polic Pract* 109:1–13
- Bischoff J, Maciejewski M (2016) Autonomous taxicabs in Berlin - a spatiotemporal analysis of service performance. In: *Transportation research procedia*, vol 19. Elsevier, pp 176–186, <https://doi.org/10.1016/j.trpro.2016.12.078>
- Bösch P, Ciarl F, Axhausen K (2016) The IVT 2015 baseline scenario. In: *16th Swiss transport research conference*
- Brockfeld E, Kühne R, Wagner P (2004) Calibration and validation of microscopic traffic flow models. *Transp Res Record: J Transp Res Board* 1876:62–70. <https://doi.org/10.3141/1876-07>
- Burns LD (2013) Sustainable mobility: a vision of our transport future. *Nature* 497(7448):181
- Cameron GD, Duncan GI (1996) PARAMICS - parallel microscopic simulation of road traffic. *J Supercomput* 10(1):25–53. <https://doi.org/10.1007/BF00128098>

- Cervero R, Golub A, Nee B (2007) City CarShare: longer-term travel demand and car ownership impacts. *Transp Res Rec* 1992(1):70–80
- Cetin N, Burri A, Nagel K (2003) A large-scale agent-based traffic microsimulation based on queue model. In: *Proceedings of Swiss transport research conference (STRC)*. Monte Verita, CH
- Childress S, Nichols B, Charlton B, Coe S (2015) Using an activity-based model to explore the potential impacts of automated vehicles. *Transp Res Rec* 2493(1):99–106
- De Palma A, Picard N, Andrieu L (2012) Risk in transport investments. *Netw Spat Econ* 12(2):187–204
- Dijk M, Orsato RJ, Kemp R (2013) The emergence of an electric mobility trajectory. *Energy Polic* 52:135–145
- Dowling R, Skabardonis A, Alexiadis V (2004) Traffic analysis toolbox volume III: guidelines for applying traffic microsimulation modeling software. Tech. Rep. FHWA-HRT-04-040
- Eppstein MJ, Grover DK, Marshall JS, Rizzo DM (2011) An agent-based model to study market penetration of plug-in hybrid electric vehicles. *Energy Polic* 39(6):3789–3802. <https://doi.org/10.1016/j.enpol.2011.04.007>
- Erath A, Fourie P, van Eggermond M, Ordoñez S, Chakirov A, Axhausen KW (2012) Large-scale agent-based transport demand model for Singapore. In: *13th International conference on travel behaviour research, future cities laboratory (FCL)*, vol 790. <https://doi.org/10.3929/ethz-b-000306926>
- Eser P, Singh A, Chokani N, Abhari RS (2016) Effect of increased renewables generation on operation of thermal power plants. *Appl Energy* 164:723–732. <https://doi.org/10.1016/j.apenergy.2015.12.017>
- Fagnant DJ, Kockelman KM (2014) The travel and environmental implications of shared autonomous vehicles, using agent-based model scenarios. *Transp Res Part C: Emerg Technol* 40:1–13. <https://doi.org/10.1016/j.trc.2013.12.001>
- Federal Statistical Office (2018a) Business and enterprise register. <https://www.bfs.admin.ch/bfs/en/home/registers/enterprise-register/business-enterprise-register.html> [Accessed: 19-Sep-2019]
- Federal Statistical Office (2018b) Federal register of buildings and dwellings. <https://www.bfs.admin.ch/bfs/en/home/registers/federal-register-buildings-dwellings.html> [Accessed: 19-Sep-2019]
- Federal Statistical Office (2019) Interactive Statistical Atlas of Switzerland. <https://www.bfs.admin.ch/bfs/en/home/statistics/regional-statistics/atlas/interactive-statistical-atlas-switzerland.html> [Accessed: 19-Sep-2019]
- Federal Statistical Office / Federal Office for Spatial Development (2012) Mikrozensus Mobilität und Verkehr (MZMV) 2010. <https://www.bfs.admin.ch/bfs/en/home/news/whats-new.gnpdetail.2012-0796.html> [Accessed: 19-Sep-2019]
- Gawron C (1998) Simulation-based traffic assignment: computing user equilibria in large street networks. PhD thesis University of Cologne, Cologne, Germany
- Goldberg AV, Harrelson C (2005) Computing the shortest path: A* search meets graph theory. In: *SODA '05 Proceedings of the sixteenth annual ACM-SIAM symposium on discrete algorithms*, pp 156–165
- Great Britain Highways Agency (1996) Design manual for roads and bridges. Volume 12. Section 2. Traffic appraisal advice. Part 1. Traffic appraisal in urban areas. HMSO
- Hars A (2015) Self-driving cars: the digital transformation of mobility. In: *Marktplätze im Umbruch*. Springer, pp 539–549
- Hollander Y, Liu R (2008) The principles of calibrating traffic microsimulation models. *Transportation* 35(3):347–362. <https://doi.org/10.1007/s11116-007-9156-2>
- Horni A, Nagel K, Axhausen KW (2016) The multi-agent transport simulation MATSim. Ubiquity Press, London
- Hourdakis J, Michalopoulos P, Kottommannil J (2003) Practical procedure for calibrating microscopic traffic simulation models. *Transp Res Record: J Transp Res Board* 1852:130–139. <https://doi.org/10.3141/1852-17>
- Huang Y, Li S, Qian ZS (2015) Optimal deployment of alternative fueling stations on transportation networks considering deviation paths. *Netw Spat Econ* 15(1):183–204
- Hülsmann F, Gerike R, Ketzl M (2014) Modelling traffic and air pollution in an integrated approach—the case of Munich. *Urban Clim* 10:732–744. <https://doi.org/10.1016/j.uclim.2014.01.001>
- Jitrapitrom P, Caiati V, Feneri AM, Ebrahimigharehbaghi S, González MJA, Narayan J et al (2017) Mobility as a service: a critical review of definitions, assessments of schemes, and key challenges. *Urban Plan* 2(2):13–25
- Kickhofer B, Hosse D, Turnera K, Tirachinic A (2016) Creating an open MATSim scenario from open data: the case of Santiago de Chile. Tech. rep., Technical report. VSP Working Paper 16–02

- Klefstad R, Zhang Y, Lai M, Jayakrishnan R, Lavanya R (2005) A distributed, scalable, and synchronized framework for large-scale microscopic traffic simulation. In: IEEE Conference on intelligent transportation systems proceedings (ITSC), pp 813–818. <https://doi.org/10.1109/ITSC.2005.1520154>
- Kwak M, Arentze T, de Romph E, Rasouli S (2012) Activity-based dynamic traffic modeling: influence of population sampling fraction size on simulation error. In: Proceedings international association of travel behavior research conference
- Litman T (2013) The new transportation planning paradigm. *Instit Transp Eng ITE J* 83(6):20–28
- Llorca C, Moeckel R (2019) Effects of scaling down the population for agent-based traffic simulations. *Procedia Comput Sci* 151:782–787
- Marini M, Gawlikowska AP, Rossi A, Chokani N, Klumpner H, Abhari RS (2018) The impact of future cities on commuting patterns: An agent-based approach. *Environ Plan B: Urban Anal City Sci* 0(0):1–18. <https://doi.org/10.1177/2399808317751145>
- Nökel K, Schmidt M (2002) Parallel DYNEMO: Meso-scopic traffic flow simulation on large networks. *Netw Spat Econ* 2(4):387–403
- NZ Transport Agency (2014) Transport model development guidelines
- OpenStreetMap (2018). <https://www.openstreetmap.org> [Accessed: 15-Jan-2019]
- Pagani M, Korosec W, Chokani N, Abhari RS (2019) User behaviour and electric vehicle charging infrastructure: an agent-based model assessment. *Appl Energy*, 254
- Prettenhaler FE, Steininger KW (1999) From ownership to service use lifestyle: the potential of car sharing. *Ecolog Econ* 28(3):443–453
- Rickert M, Nagel K (2001) Dynamic traffic assignment on parallel computers in TRANSIMS. *Futur Gener Comput Syst* 17(5):637–648. [https://doi.org/10.1016/S0167-739X\(00\)00032-7](https://doi.org/10.1016/S0167-739X(00)00032-7)
- Roads and Maritime Services, New South Wales Government (2013) Traffic modelling guidelines
- Sanders P, Schultes D (2005) Highway hierarchies hasten exact shortest path queries. Springer, Berlin, pp 568–579. https://doi.org/10.1007/11561071_51
- Saprykin A, Chokani N, Abhari RS (2019a) GEMSim: a GPU-accelerated multi-modal mobility simulator for large-scale scenarios. *Simul Model Pract Theory* 94:199–214
- Saprykin A, Chokani N, Abhari RS (2019b) Large-scale multi-agent mobility simulations on a GPU: towards high performance and scalability. *Procedia Comput Sci* 151:733–738
- Schroeder A, Traber T (2012) The economics of fast charging infrastructure for electric vehicles. *Energy Polic* 43:136–144
- Ševčíková H, Raftery AE, Waddell PA (2007) Assessing uncertainty in urban simulations using Bayesian melding. *Transp Res B* 41:652–669
- Simoni MD, Pel AJ, Waraich RA, Hoogendoorn SP (2015) Marginal cost congestion pricing based on the network fundamental diagram. *Transp Res Part C: Emerg Technol* 56:221–238. <https://doi.org/10.1016/j.trc.2015.03.034>
- Smith L, Beckman R, Anson D, Nagel K, Williams ME (1995) TRANSIMS: transportation analysis and simulation system. Tech. rep. <https://doi.org/10.2172/88648>
- Speranza MG (2018) Trends in transportation and logistics. *Eur J Oper Res* 264(3):830–836
- Strippgen D, Nagel K (2009) Multi-agent traffic simulation with CUDA. In: Proceedings of the 2009 international conference on high performance computing and simulation, HPCS 2009, pp 106–114. <https://doi.org/10.1109/HPCSIM.2009.5192895>
- Swiss Open Data Platform (2018) <https://opentransportdata.swiss/en/dataset/timetable-2018-gtfs> [Accessed: 15-Jan-2019]
- Theil H (1958) Economic forecasts and policy / by H. Theil, assisted by J.S. Cramer. H. Moerman and A Russchen. North-Holland Publishing Company, Amsterdam
- Zhang L, Levinson D (2004) Agent-based approach to travel demand modeling: exploratory analysis. *Transp Res Record: J Transp Res Board* 1898:28–36. <https://doi.org/10.3141/1898-04>
- Zhang L, Yang W, Wang J, Rao Q (2013) Large-scale agent-based transport simulation in Shanghai, China. *Transp Res Rec* 2399(1):34–43. <https://doi.org/10.3141/2399-04>
- Zhuge C, Shao C (2018) Agent-based modelling of locating public transport facilities for conventional and electric vehicles. *Netw Spat Econ* 18(4):875–908