Effects of population sampling on agent-based transport simulation of on-demand services

Conference Paper

Author(s):
Kagho, Grace Orowo; Meli, Jonas; Walser, Dominique; Balac, Milos

Publication date:
2022

Permanent link:
https://doi.org/https://doi.org/10.3929/ethz-b-000662210

Originally published in:
Procedia Computer Science 201, https://doi.org/10.1016/j.procs.2022.03.041

Available online at www.sciencedirect.com
ScienceDirect
Procedia Computer Science 201 (2022) 305–312
Procedia Computer Science
www.elsevier.com/locate/procedia

The 13th International Conference on Ambient Systems, Networks and Technologies (ANT)
March 22 - 25, 2022, Porto, Portugal

Effects of population sampling on agent-based transport simulation of on-demand services

Grace O. Kagho[a,*], Jonas Meli[a], Dominique Walser[a], Milos Balac[a]

[a] *ETH Zurich, 8093 Zurich, Switzerland*

Abstract

Large-scale agent-based simulations require higher computing resources than are usually available. Consequently, many applications rely on downscaling, i.e., simulating with smaller population samples. Existing studies have shown a need to investigate the impact of downscaling on the output statistics of such simulations. In this study, we investigate the impacts of population downscaling on a ride-sharing service with a focus on vehicle occupancy and wait time, travel time and detour time. Our findings reveal that if transport modelers want to model on-demand services with ride sharing, it is strongly recommended to use a 100% population, or when using a smaller population sample, to estimate the relative biases of their desired metrics compared to the results of a 100% population in order for their results to be applicable for real-world situations.

Keywords: Agent-based simulations; population sampling; MATSim; downscaling; shared mobility

1. Introduction

Agent-based modelling has gained traction as a suitable method for solving today's dynamic and complex transport problems. This is because present transport problems in our cities require models that can properly capture the complex interactions that exist at a microscopic level among travellers and the transport system. However, large-scale applications of agent-based simulations pose challenges to today's computing capabilities especially as these transport models introduce more complexity to be able to produce models that are as close to the real world scenario as possible. For example, a MATSim transport model of a demand responsive transit system in Wayne County, Michigan with a travel demand of 9 million trips takes approximately 43 hours to simulate 30 iterations on a high performance computing cluster with 12 cores and 144G memory [6].

The computational tractability issues encountered with agent-based simulation are currently being approached through a variety of ways such as the use of approximation algorithms, parallelization, integration robust programming

* Corresponding author. Tel.: +41-44-633-33-17
E-mail address: grace.kagho@ivt.baug.ethz.ch

10.1016/j.procs.2022.03.041

techniques, and increase in number of hardware introducing parallelization [9]. As these methods still have their limits, one method that has gained the most popularity is downscaling — an approach of simulating these large-scale agent-based systems with smaller population samples in which their results are then scaled.

While downscaling has become a common practice, not many studies have looked at its consequences on various transport simulation results, with a few studies looking at the effect on travel times, link volumes, boardings or demand.[1, 7, 11]. As transport studies look at emerging on-demand services and the potential of ridepooling to reduce congestion and environmental impact of car drivers [10], these services introduce new dynamics to transport simulations, and hence, it becomes difficult to interpret their downscaled results. For example, a pertinent question that arises is, how should one compare a single passenger in a 10% sample versus a 100% population? To answer such a question, it is the aim of this study to investigate the effects of downscaling on vehicle occupancy and other simulation results of on-demand services. The study focuses on a common vehicle occupancy scenario; that of ride-sharing services in an urban area. Selecting the city of Zurich as a case study, and using MATSim, a multi agent-based simulation framework [4], we look at the variance and bias of the results of the downscaled population and examine the tradeoffs between the sample sizes and the computation time.

This paper is structured as follows, section 2 gives an overview of the state of the art in regards to the impact of downscaling, section 3 describes the methodology applied to this study and in section 4 the results are presented. Conclusions are then highlighted in the final section.

2. Background

Large-scale agent-based systems that are designed for transport simulation of various cities require higher computing resources than is usually available. This is because these large-scale transport simulations demand many simulation runs with different varying parameter sets, which leads to higher computation times with increasing number of simulated agents and the complexity of agent interactions. As such, to perform a robust study with available computing power, many large-scale agent-based transport simulation studies use a smaller sample of the population and a corresponding downscaled transport environment. Therefore, it is essential to ensure that the dynamics and evolution of the sampled model are the same as those of the total population.

MATSim, one of the most widely used agent-based simulation frameworks for large-scale scenarios has been used in studies covering large urban areas, with the modelled scenario (the population and transport system) downscaled to smaller sample sizes ranging between 1% and 25% in order to reduce computing time or just to follow general trends [1]. Only a few studies exist that have investigated the downscaling effects. These studies explore the impacts on various output statistics with varied results.

Ben-Dor et al. [1], in a recent study, explore downscaling effects in MATSim, and they provide an in-depth investigation of these effects across various output statistics for a Sioux Falls car traffic simulation. They find that, with sampled populations equal to or greater than 25%, one can preserve the majority of the urban traffic statistics. 10% or lower population samples lead to significant deviations from the full-scale model, while population sizes between 10 and 25% may render some spatial/temporal statistics, such as trip duration and traffic volume, unstable. Their results are different from previous works. For example, Llorca and Moeckel [7] examined the effects of downscaling on the travel times and daily scores of agents and found that the travel time distribution of sample populations that are 5% or above are similar to that of the full population. Zhuge et al. [11], in their simulation of the Baoding city in China, carried out sensitivity analysis for various output statistics (travel speeds, travel distance, Volume-to-Capacity ratio and agent scores) and found different results for the different output statistics for different population samples. They suggest that 50% is the minimum sample size that would keep all four analysed outputs the same as those of the 100% population.

The above studies show that the sample size matters, depending on the output statistic in which one is interested in investigating. This means that, as various agent-based traffic simulations are expanded to capture more complex and dynamic transport activities, the ensuing output statistics need to be investigated for downscaling effects. On-demand services and related shared mobility services are one of the emerging transport alternatives that allow for flexibility in the use of different mobility tools, i.e., vehicles adjust their service to the needs of a traveler. These modes allow for ride-sharing or pooling of passengers. There are two recent studies [12, 5] where the the effects of downsampling are highlighted, although this is not investigated in detail. Zwick et al. [12] investigate the potential of ride-sharing in six regions in Munich, Germany with varying population ranging between 16 000 and 1.5 million. Their results show that, an increase in trip density brings about a logarithmic increase in the ride-sharing service parameters. Kaddoura and Schlenther [5] found similar nonlinear results when they investigated the relationship between trip density and fleet sizing using population samples ranging between 5% and 100% of two German regions. Both studies do not investigate in detail the effects of population sampling on their results, especially for vehicle occupancy. They work with dynamic fleets and compare across populations with different characteristics, making it difficult to understand the downscaling effects holistically. Thus, in this study, we look at the effects of downscaling on vehicle occupancy and other simulation results of on-demand services, working with one region and a fixed fleet size.

3. Methodology

3.1. Scenario and Simulation setup

This study is set up using an existing baseline scenario of the city of Zurich which has been extracted from the MATSim Switzerland model [3]. The scenario consists of about 1.1 million synthetic persons who generate a travel demand of about 4.7 million trips. This travel demand has been generated by running the "MATSim Loop", which consists of running a mobility simulation iteratively with modules that recalculate alternatives for each agent and allow agents to perform mode and route choice before the next run of the mobility simulation. Eventually, an equilibrium state is reached typically after 40 iterations. The resulting mode shares of the output population plans are roughly 32% for car trips and 23% transit trips with the remaining shares for walk, bike, and car passenger trips.

In this study, which aims to quantify the effect of downscaling on modelling a demand responsive transit (DRT), we extract all car trips that start and end within the study area and replace the mode of these trips with ride-sharing.This leaves about 700 000 trips that are represented as individual agents in the simulation. The ride-sharing service is enabled by MATSim's DRT extension [2], which is designed to allow MATSim to simulate a dynamic ride-sharing service where vehicles can pick up and drop off agents in response to their requests. The MATSim version 13, with its corresponding DRT version, has been used. The DRT service in this study, operates a door-to-door scheme for 24 hours. It is possible to set constraints for a maximum wait time and maximum detour time (the time an in-vehicle passenger is willing to lose while the DRT vehicle picks up another passenger), whereby agents are rejected when these constraints are not met. To avoid such rejections in this study, the on-demand service operator is required to serve all requests, i.e., a no rejection criteria, unless the trip cannot be carried out within the simulation time. In a simulation run with the full population, a DRT service with 12 000 4-passenger vehicles is offered to serve the demand. This fleet size has been chosen to ensure an average wait time within 10 minutes. For sample populations, we downscale the DRT fleet proportionally. The fleets are placed in a starting location based on population densities. In this study, we run four iterations to achieve a stable state in the number of DRT trips served, average travel times and wait times. This is because the DRT system operates a dynamic vehicle routing problem for its dispatching algorithm, in which, the network freespeeds are used to compute the travel times in the first iteration, and then later updated until the times are stable.

3.2. Population Sampling

Downscaling in MATSim is usually done using a downscaling ratio, which represents a fraction of the total population. This value, which we will refer to as k, is not only used to sample from the synthetic population but also used to adjust the scenario's network's link flow and storage capacity factors in proportion to the value of k. For example, for a 10% sample scenario ($k = 0.1$), 10% of the agents from the synthetic population are randomly selected for simulation and the link flow capacity factor is set to k and the storage capacity factor is set to $k^{0.75}$ as proposed by [8] to avoid the possibility of heavily congested links. For this study, we also downscale the DRT fleet size.

The analysis carried out in this study consists of two parts; an initial analysis that does not consider stochasticity, and a subsequent analysis, where each sampling scenario is repeated 10 times with different random seeds.

Preliminary analysis: Here, the output of the random sampling for a range of 20 sample populations between 0.01 and 1 are analysed. We check to ensure that the spatial distribution of the DRT trips are consistent across all sample sizes. Subsequently, we examine the following output statistics: average travel time, average wait time, average detour factor (the ratio of passenger experienced total travel time and direct travel time), and average vehicle occupancy.

Final analysis: Due to the stochastic nature of MATSim, it is necessary to run multiple simulations with different random seeds and analyse the distributions of the output statistics [1]. This is essential when downscaling in order to exclude the effects of a random sample fleet size and population selection and to estimate the variance of the simulation results due to the stochastic nature of the simulation. From the results of the preliminary analysis, which are described in detail below, a few sample sizes are selected, and the sample simulations are run repeatedly with different random seeds. The selected k values for the final analyses are 0.25, 0.35, 0.5, 0.65 and 0.75.

The main output statistic investigated here is the vehicle occupancy, specifically, the share of vehicles in different vehicle states (either idle, moving but with zero passenger, carrying one or up to four passengers) over the course of the day. The coefficient of variation (CV) is computed for each simulated population sample size and the full population as shown in Equation 1.

$$CV = s/m \tag{1}$$

where m and s is the mean and standard deviation of the output statistics of the repeated simulation runs.

Comparisons between full population and sampled populations are made by estimating the relative bias between the output statistic c_{ki} of the sampled repeated runs and the average mean m_f of the output statistic of the full population. This is shown in Equation 2.

$$r_{ki} = (c_{ki} - m_f)/m_f \tag{2}$$

where r_{ki} is the relative bias for each repeated simulation run and c_{ki} is the output statistic of a k-sample random seed simulation i, and m_f is the average of the output statistic of all the random seed runs for the full population.

Furthermore, the relative bias computed is averaged over the number of runs and the standard deviation is computed. The resulting values are checked; if the mean relative bias is close to zero and the standard deviation is close to $CV_f/\sqrt{k}$, the output statistic is considered robust to downscaling.

The results of the 1% population sample are outliers, varying greatly from the results of the other sample sizes. Consequently, they have not been included in Fig. 1. The average travel time and average wait time for this sample size are 40 and 210 minutes, respectively, the average detour factor is 68 and the average occupancy is 1.23. One reason for these high values is the sparseness of the demand. For this sample size, there are 120 vehicles to serve about 6 500 agents in absolute numbers. This small number of agents is as widely distributed over the model region as the full population, and since no trips are to be rejected, unless the trip cannot be carried out within the simulation time, large wait times can be observed. This suggests that, linearly scaling the fleet size for such a small sample size results in some artefacts in the system.

From Fig. 1a-c, a non-linear relationship can be observed between the output statistics and the sample sizes. Larger deviations from the result of the 100% population can be seen in sample sizes less than 20%, with a decrease in the deviation, the higher the sample size. For the average travel time, the absolute difference between the sample populations and full population ranges between 0.1% and 15%. From a sample size of 20% upwards, the percentage change is less than or about 1%. Similarly, for the average wait time, there is a decrease in percentage change with increasing sample size. However, it is only from sample sizes of 40% and above that one can observe a percentage change of less than 10%; and from 75% sample sizes upwards, the percentage change is close to 1%. For the detour factor, the percentage change is lower than 10% for sample sizes of 35% upward, and about 1% for sample sizes from 75% and above.

In contrast, vehicle occupancy (see also Fig. 1d) increases with increasing sample size without reaching a stable value. From 50% sample size and above, the percentage change is less than 10%, and only at 95% sample size, the percentage change is less than 1%.

Depending on the output statistics, the deviation in the results of the downscaled population from the full population differs for different sample sizes, with decrease in the marginal error with increasing sample size, as can be especially seen from Fig. 1d fitted log function $f(k) = 0.55 + 0.12log_{10}(k)$. However, it is important to note that the above results do not take into consideration the stochasticity of the simulation.

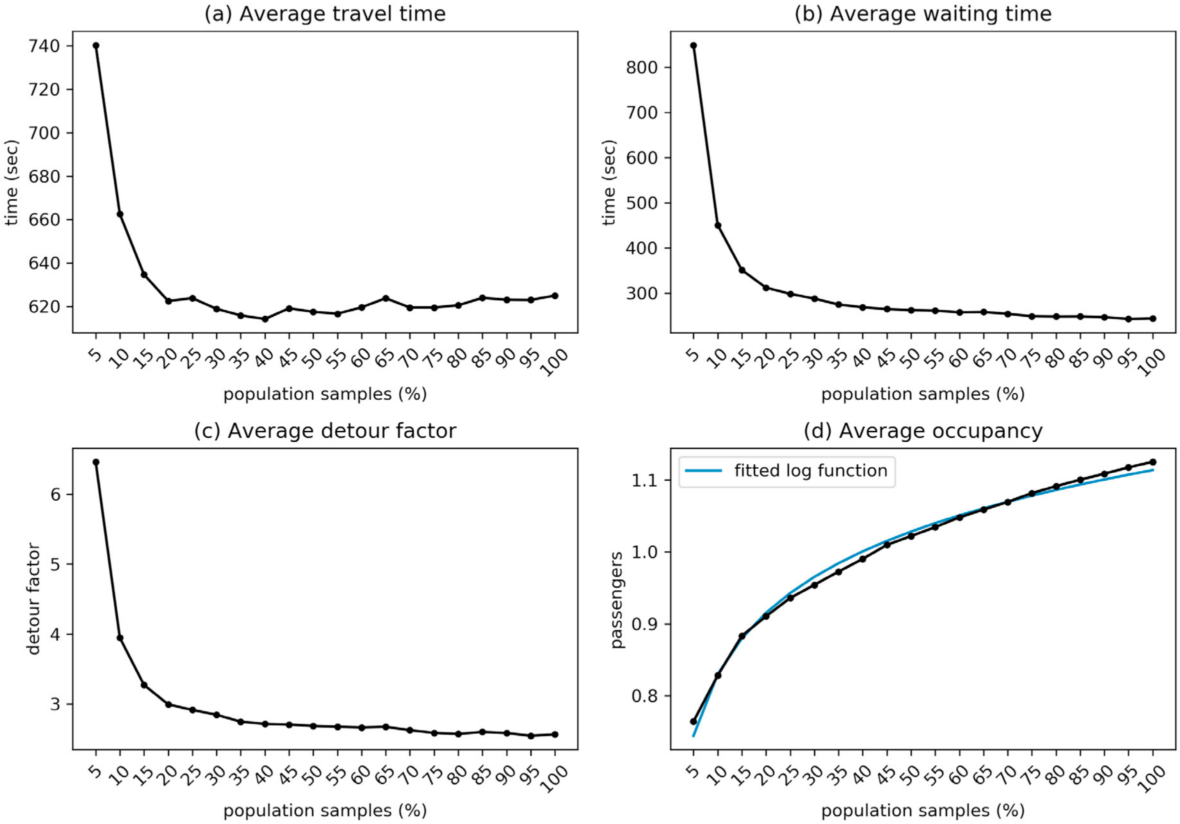

Fig. 1. (a) shows average travel time per population size, (b) shows average wait time per population size, (c) shows the average detour factor (ratio of passenger total travel time to direct travel time) per population size and (d) shows the average vehicle occupancy per population size with a fitted log function $f(k) = 0.55 + 0.12log_{10}(k)$

3.3. Final analysis

Since average occupancy is the most important measure that we consider in this study, for the next analysis, we dive deeper into observing the impact of population sampling on vehicle occupancy looking at the different states of the vehicle over the whole day. We check for variability in the results, while determining the measure of the similarity of the downscaled and the full population for selected k samples. The results of the population samples are also scaled up for this comparison. In the figures presented below, the full population statistic is highlighted in gold colour, while other k samples correspond to the plot legend. Fig. 2 shows the average values of the repeated runs for the sample sizes distributed over the whole day. One can immediately observe that sample sizes from 65% have closer results to the 100% population, although this varies with the different states of the vehicle.

Next, the variability in the results is examined. While the coefficient of variation (CV, refer to Eqn. 1) is very low for no passenger and 1 passenger states, Fig. 3 (note that the scales vary on the y-axis) shows that the CV increases for higher vehicle occupancy. Further, very high values can be observed for early and late hours of the day, with higher values for the smaller sample sizes. The peak in the early and late hours is likely due to a low number of observations

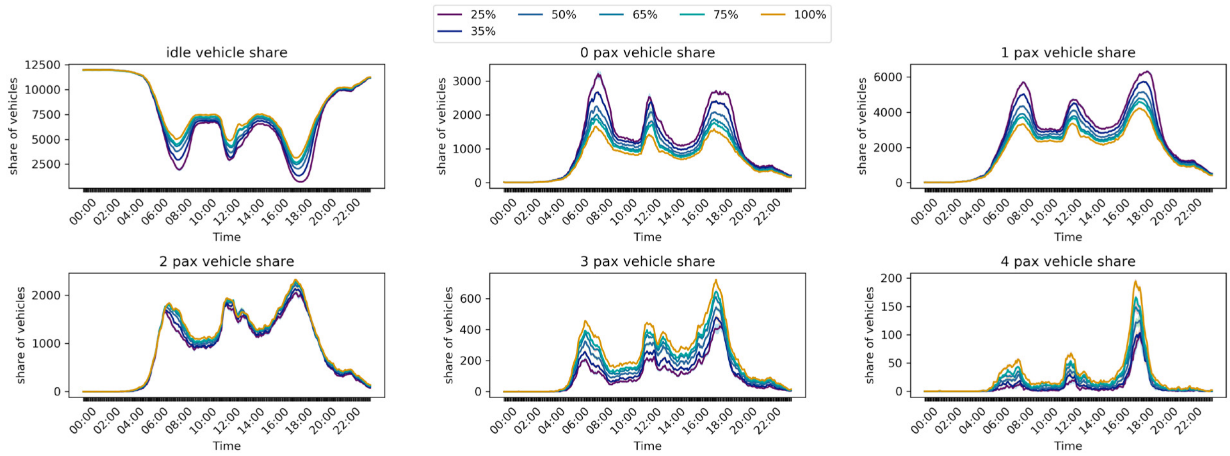

Fig. 2. Share of vehicles operating in different vehicle states for different population sizes

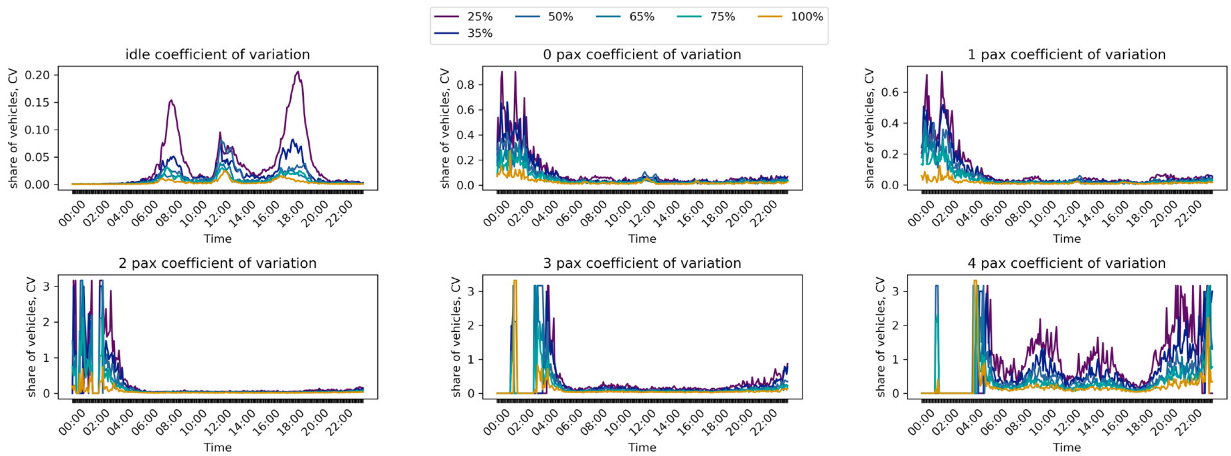

Fig. 3. Coefficient of variation of different random seeds for share of vehicles operating in different vehicle states for different population sizes

during this time period. A larger number of repeated simulation runs would need to be tested, especially for sample sizes below 65%.

Finally, the relative bias of the observed distributions for the different sample population sizes in respect to the full population is explored. Fig. 4 shows the relative bias for the different vehicle states for different periods of the day. The results which are in 5-minute time bins have been grouped into morning peak (5 AM - 8:59 AM), afternoon peak (11 AM - 1:59 PM), evening peak (4 PM - 7:59 PM), and off-peak representing all the hours in between.

The findings reveal that there is a relative bias in all sample sizes, and this bias, based on the vehicle state, varies. The results are similar for the peak periods. For the off-peak period, we observed a relative bias of more than 150% for higher vehicle occupancy states. Consequently, for better clarity, the off-peak period is not shown in the figure. For vehicles with no passengers or one passenger, one can see that the bias is relatively low, compared to the other vehicle states. Any of the tested sample sizes in these two states seem reasonable, with a relative bias less than 20% and even as low as 1% for 65% sample size and above. However, one can observe an almost u-shaped curve relatable to interdependencies in the system and differences between random seeds such as trip density at different locations. For vehicle states with more than one passenger, the nonlinear decrease in the relative bias with increasing sample

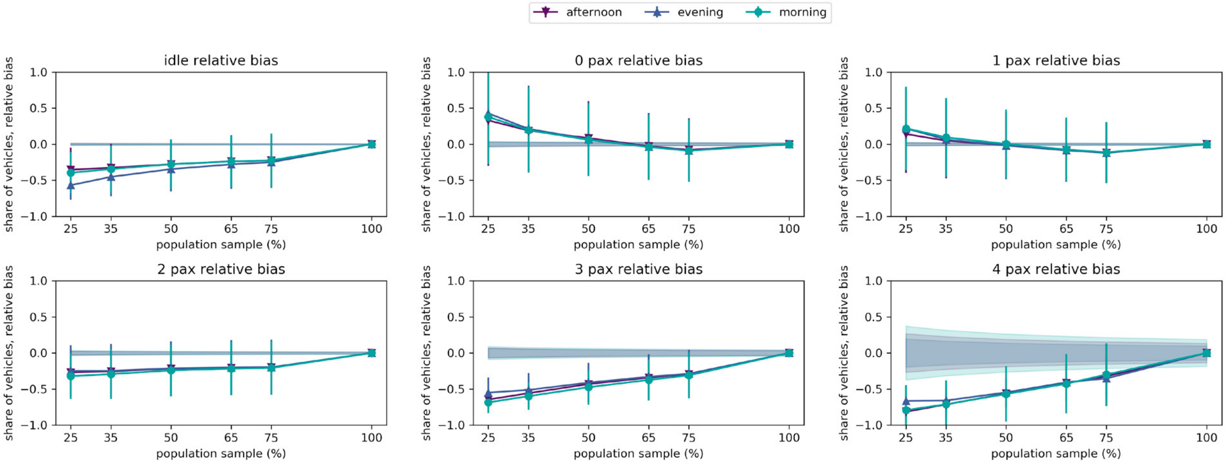

Fig. 4. Average relative bias and its standard deviation (error bars). The shaded area signifies $\pm CV_f / \sqrt{k}$ boundaries

size is as expected. However, one should note that the bias is very large and consistently underestimates the vehicle occupancy, with the bias increasing with a higher vehicle occupancy value and as the sample size decreases. Even with 75% sample size, the bias remains as high as 35%. This demonstrates two things. Firstly, the need to always look at such relative biases when determining the appropriate sample size and the tradeoff with simulation run time which would be discussed next. Secondly, that further investigation is needed to determine how to both measure, and reduce these biases. This future work should include testing higher sample sizes and testing more than 10 repeated runs of these population samples to better estimate the coefficient of variation.

3.4. Simulation runtime

The inclusion of a DRT Service in a MATSim scenario requires higher computational power due to the dispatching algorithm used by the DRT service. Larger fleet sizes and demand increases the simulation time and memory needed. In the full population where 12 000 vehicles are used to serve about 700 000 trips, the complete simulation run time for the population samples [25%, 35%, 50%, 65%, 75% and 100%] are 3, 5, 8, 14, 17 and 35 hours, respectively. The simulation has been run on the ETH high performance computing cluster (Euler) with 8 cores and 192 GB memory. MATSim v13, as sourced from [www.https://github.com/matsim-org/matsim-libs/tree/13.x] with the corresponding DRT extension version has been used in this study.

4. Conclusion

The effects of downscaling populations in large-scale agent-based transport simulation varies, depending on the output statistic of interest and the transport scenario being modelled. Previous studies that have only examined these effects for conventional modes have recommended sample values between 5% and 50%. The most recent among them did a comprehensive analysis for the car mode and suggests a safe value of 25%. With the consideration of on-demand shared mobility services, a critical question is how the vehicle occupancy should be scaled. This study has revealed different results in the effects of downscaling for a ride-sharing scenario, wherein vehicle occupancy is a key output statistic of interest.

Our findings show that a linear upscaling of the output statistics can produce an overestimation or underestimation of the results. Particularly for vehicle occupancy, a logarithmic relationship was seen with the sample sizes underestimating the average vehicle-occupancy of the full population. This is likely due to the effect of trip density, whereby, as the sample size decreases, the trip density decreases, leading to loss of trip heterogeneity. That is, the more sparse an area, the more difficult it is to pick up additional passengers under realistic wait time/detour time constraints as well as fleet size constraints. This is consistent with the results of Zwick et al. [12] who also found a logarithmic relationship between trip density and pooling rate.

Furthermore, the findings suggest that when downscaling, especially in cases where you have the possibility for vehicle pooling, it is necessary to check the tradeoffs between sample sizes and the computation time. For example, for a 25% sample size which takes about 3 hours for four iterations, compared to 35 hours for a 100% population, the relative bias was very high (potentially up to 80%). Therefore, it is necessary to consider a much higher sample size, at least for producing robust results. For example, the 75% sample size have been shown to have a relative bias of about 35%, while its simulation time is about half of that of the full population. This begs the question of how much error margin one is willing to consider in order to reduce the simulation time. Thus, it is highly recommended to continue this line of research to determine a way to correct the sampling bias.

Lastly, the simulations used to produce the results presented in this paper are from a larger study that is still in progress. One should keep in mind that the biases observed in this paper could be due to a number of reasons. Factors such as network effects, differing trip densities from random sampling, study's methodologies, or even scenario-specific settings could influence the sign and magnitude of the biases discussed. A pertinent and specific example is that the demand and fleet size used in this study's scenario were held constant for this initial investigation. Subsequently, it could be possible that the results discussed are specific to the simulated study area and the assumed demand. Future work should focus on how to reduce the relative bias while considering the aforementioned effects.

References

[1] Ben-Dor, G., Ben-Elia, E., Benenson, I., 2021. Population downscaling in multi-agent transportation simulations: A review and case study. Simulation Modelling Practice and Theory 108, 102233.

[2] Bischoff, J., Maciejewski, M., Nagel, K., 2017. City-wide shared taxis: A simulation study in Berlin, in: 2017 IEEE 20th international conference on intelligent transportation systems (ITSC), IEEE. pp. 275–280.

[3] Hörl, S., 2020. Dynamic Demand Simulation for Automated Mobility on Demand. Ph.D. thesis. ETH Zurich.

[4] Horni, A., Nagel, K., Axhausen, K.W., 2016. The multi-agent transport simulation MATSim. Ubiquity Press London.

[5] Kaddoura, I., Schlenther, T., 2021. The impact of trip density on the fleet size and pooling rate of ride-hailing services: A simulation study. Procedia Computer Science 184, 674–679.

[6] Kagho, G.O., Hensle, D., Balac, M., Freedman, J., Twumasi-Boakye, R., Broaddus, A., Fishelson, J., Axhausen, K.W., 2021. Demand Responsive Transit Simulation of Wayne County, Michigan. Transportation Research Record 2675, 702–716.

[7] Llorca, C., Moeckel, R., 2019. Effects of scaling down the population for agent-based traffic simulations. Procedia Computer Science 151, 782–787.

[8] Nicolai, T.W., 2012. Using MATSim as a travel model plug-in to UrbanSim. SustainCity. Deliverable D 7.

[9] Parry, H.R., Bithell, M., 2012. Large scale agent-based modelling: A review and guidelines for model scaling. Agent-based models of geographical systems , 271–308.

[10] Ruch, C., Lu, C., Sieber, L., Frazzoli, E., 2021. Quantifying the Efficiency of Ride Sharing. IEEE Transactions on Intelligent Transportation Systems 22, 5811–5816.

[11] Zhuge, C., Shao, C., Wang, S., Hu, Y., 2019. Sensitivity analysis of integrated activity-based model: Using MATSim as an example. Transportation Letters 11, 93–103.

[12] Zwick, F., Kuehnel, N., Moeckel, R., Axhausen, K.W., 2021. Ride-Pooling Efficiency in Large, Medium-Sized and Small Towns -Simulation Assessment in the Munich Metropolitan Region. Procedia Computer Science 184, 662–667.