Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 62 (2022) 719–726

Transportation Research Procedia

www.elsevier.com/locate/procedia

24th Euro Working Group on Transportation Meeting, EWGT 2021, 8-10 September 2021, Aveiro, Portugal

Uncertainty and Variability Analysis of Agent-Based Transport Models

Lasse Bienzeisler[a,*], Torben Lelke[a], Oskar Wage[b], Lena-Marie Huck[a], Bernhard Friedrich[a]

[a]*Institute of Transportation and Urban Engineering, TU Braunschweig, Hermann-Blenk-Str. 42, 38108 Braunschweig, Germany*
[b]*Institute of Cartography and Geoinformatics, Leibniz University Hannover, Appelstr. 9a, 30167 Hannover, Germany*

Abstract

This paper presents an analysis of the output variability of agent-based transport models. We simulated a MATSim model of the city of Hanover multiple times with identical input and evaluated the resulting travel times on different level of aggregation. On a global level, we observed minor variations of travel times. However, the results show an increased variation when examining the output on the level of districts or for individual agents. A recommendation for estimating the required number of simulation runs for a stable output of travel time for the purposed aggregation level is derived from our case study.

Keywords: MATSim; Output Variation; Agent-based Transport Simulation

1. Introduction

With the increasing relevance of agent-based simulations, various approaches have been developed, with MATSim (Horni et al., 2016) emerging as one of the most frequently used open-source simulation frameworks. MATSim is based on utility maximization. Individual mobility decisions on trip purpose, destination, mode, and time choice are calculated by econometric discrete choice models to reproduce a fine-grained traffic demand. The ability to simulate each agent individually enables the consideration of complex linkages across multiple trips. While competing with all other agents for space-time slots on the transport infrastructure, each agent repeatedly optimizes its daily activity schedule. Optimization is performed in an iterative cycle with a predefined fraction of agents randomly changing their plans at each iteration. The framework evaluates the new plan using a scoring function after the subsequent simulation step (Horni et al., 2016).

* Corresponding author.
E-mail address: l.bienzeisler@tu-braunschweig.de

10.1016/j.trpro.2022.02.089

The variability of the agents' choice in MATSim is based on a series of pseudo-random numbers determined by a random seed (Paulsen et al., 2018). However, due to the coevolutionary algorithm, the choices are not executed pseudo-randomly in every run of the simulation (Horni et al., 2011). Thus, the resulting models are non-deterministic. The uncertainty in the evaluation of different simulation runs is a well-known problem that needs to be considered when conducting simulation case studies (Rasouli and Timmermans, 2012). Investigating different model parameters on different levels of aggregation (*LOA*), we observed substantial variations of different measures. However, there is a lack of systematic uncertainty analysis of MATSim simulations to facilitate more educated decision-making. The prerequisites for stable simulation results with a desired reliability have not yet been investigated for different aggregation levels.

2. Related Work

Quantification of the reliability of decisions based on mathematical models is a prominent topic in the field of transport modelling. With the emergence of models for increasingly complex problems that are hard to interpret, researchers have started efforts to generalize definitions of uncertainty and analysis methods. Walker et al. (2003) presented a theoretical framework for systematic uncertainty analysis in model-based decision support. Therein, uncertainty is defined as 'any departure from the unachievable ideal of completely deterministic knowledge of the relevant system'. The authors differentiate between three dimensions of uncertainty, which they define as uncertainty of *location*, *nature*, and *level*.

Multiple studies have examined variations in activity-based micro-simulations, such as the established Albatross (Arentze and Timmermans, 2004) or Feathers models (Bao et al., 2015). According to Baustert (2021), the most commonly analyzed uncertainty location in these models is the *simulation error*. Reasons of this are the relative ease with which this location can be addressed and the often stochastic nature of these models. Castiglione et al. (2003) studied the minimum number of runs needed to achieve robust average results. Cools et al. (2011) assessed the impact of micro-simulation errors on the average daily number of trips per person as well as the average daily distance traveled per person. Their results show minimal variation, especially for aggregated values.

Agent-based micro-simulations, such as MATSim, are particularly prone to model uncertainties since they often rely on discrete choice models to perform mode choice and trip assignment. According to Horni et al. (2016) and Horni et al. (2011), the coevolutionary algorithm of MATSim is the major location for uncertainties in the simulation and infers different types of uncertainty introduced by time, route, and destination choice modules. Caused by the random seed, distinct uncertainty is introduced for ever iteration of the simulation. Different random numbers may lead the optimization algorithm to find other local optima. Moreover, MATSim contains a random variability in how the replanning of plans is handled. Horni et al. (2011) demonstrated that the results of simulations can change significantly between multiple runs. In their work, they studied the impact of varying random seeds with a focus on link loads in two different MATSim scenarios. They considered that the variation in daily link loads is generally low. However, considering hourly values, the coefficient of variation increases. In their literature review they also concluded that average results generated from micro-simulations become stable after 'a relatively small number of simulation runs' (Horni et al., 2011, p. 8). These findings were probated by Paulsen et al. (2018). Chapter 48 of the MATSim book (Flötteröd, 2016) also describes the challenges in MATSims output evaluation due to the influence of the choice of one specific random seed and elaborates the need for further research in this particular area. Thus, we strive to add additional levels of investigation to this discussion by analyzing the output's travel time variation of MATSim and expanding our studies to regard different *LOA*.

3. Methodology

To investigate the variability of MATSim simulation outputs, we set up a simulation case study for an 10 % model of the city of Hanover, Germany (Bienzeisler et al., 2020). Using the referenced configuration parameters, we repeatedly simulated the Hanover input model with 750 iterations. The public transport system was implemented as a network mode. In addition, commercial traffic was included in the model using the freight extension of MATSim (Zilske et al., 2012) separated by different branches. We simulated 16 simulation runs with the same input parameters to explore inconsistencies across the simulation outputs.

After Paulsen et al. (2018) concentrated their work on the variation of link loads using different random seeds, we focused on the variation of travel times. Travel time distributions are a model characteristic that can be used for calibration or validation. Thus, the evaluated dimension of the travel time t per private agent (p) or commercial traffic vehicle (ct) was defined as the sum of all trip durations per day. We considered the changing travel times t^r per run r in the set of 16 runs R per agent $a \in$ all agents A to explore the effects of the uncertainties from the random choice parts of the MATSim algorithm. We assigned the corresponding home district $d \in$ all districts of Hanover D to each agent a_p. Three aggregation levels of the analyzed travel times were introduced as a set of travel times $t^r \in LOA$. The evaluation of travel times was carried out separately for each LOA and each simulation run $r \in R$:

- LOA_1: Global average travel time of Hanover: $\overline{t^r}$ with t^r for $a \in A$
- LOA_2: Average travel time for each district d of Hanover: $\overline{t^r}$ with t^r for $a \in d$
- LOA_3: Travel time for each agent a of Hanover: t^r for all $a \in A$

To quantify the variation of the travel time, we applied the coefficient of variation $c_v(t^r)$, which is defined as the standard deviation of the sample divided by the sample mean, on our defined LOA.

4. Analysis of the Variation of Travel Times

To obtain a first understanding of the variation of travel times across the simulation runs, we started our work by comparing the frequency distributions of all occurring travel times per agent of the private traffic for each simulation run separately. Travel times were grouped in bins of 1 minute, each with their corresponding frequency per run. For a better comparison of the resulting 16 travel time distributions, we have combined the histograms in the 3D bar plot shown in Figure 1. Each bar represents the frequency of occurrence of a travel time group per simulation run.

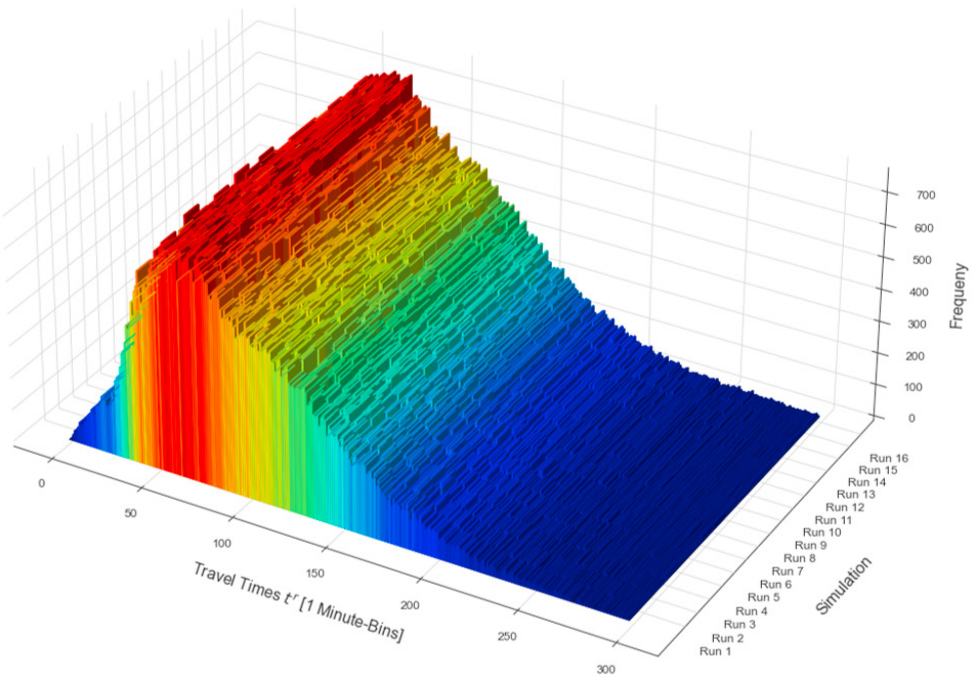

Fig. 1: Combined Histograms of travel time t_r in bins of 1 minute with $r \in R$.

The rough surface depicted in the figure provides a visual indication that the simulation generates different distributions of travel times. Especially in the range of frequently occurring values, different patterns can be observed.

However, the diagram also highlights that there are runs with a similar distribution of travel times, where the surface of the plot is constant and smooth. This can be observed for runs 11 and 12.

To determine the deviation of the simulation runs, we calculated the corresponding Root Mean Square Error (RMSE) and thus compared all runs to each other (see Figure 2). In most cases, the RMSE varies from 12.73 to a maximum of 20.20. Notice that there are runs with a RMSE of 0. This indicates an identical distribution of the travel times for the combination of these specific runs. This observation is consistent with the first visual analysis. The simulations replicated exactly the same travel times for run 4, 9, 11, 12, 14, and 16.

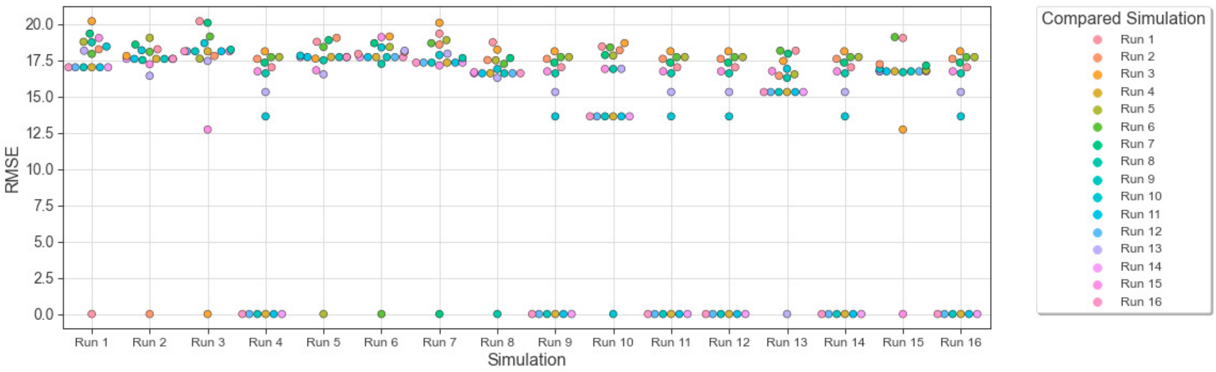

Fig. 2: RMSE-Analysis of the travel time distribution of all simulation runs R.

Another finding from our data is that the simulation outputs can settle in several discrete states for each agent. The number of states s per agent a is defined as the number of different travel time values for an agent occurring over all simulation runs. Figure 3 shows the number of different states s_a occurring over all simulation runs as a cumulative distribution plot.

A large group of private agents, 21.9 % (n = 16.312) has one state, i.e. one constant travel time over all runs. In these cases, the travel time distribution does not oscillate and the specific state reoccurs in every simulation run. For commercial vehicles, this applies for 3.4 % (n = 211) of the agents.

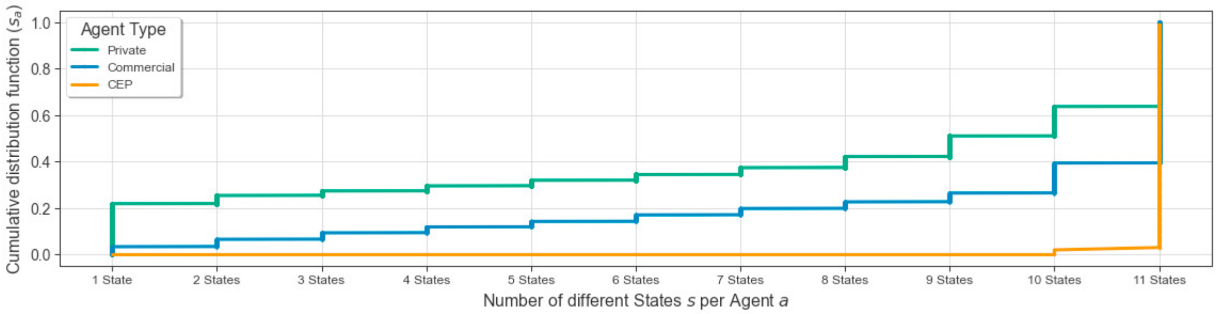

Fig. 3: Empirical cumulative distribution with the number of different states s_a across all agents grouped by agent type.

As shown, the output for commercial traffic vehicles differs more. In particular couriers, express, and parcel service (CEP) vehicles do not settle in discrete states, but show individual travel times for each run. However, the sample size is significantly smaller (n_p = 74.394, n_{ct} = 6.148, n_{CEP} = 98). To explore this characteristic of the freight traffic in MATSim in detail, we plotted the frequency distribution of travel times per simulation run in Figure 4. The more homogeneous distribution of travel times t^r for commercial traffic vehicles compared to CEP vehicles per run is evident.

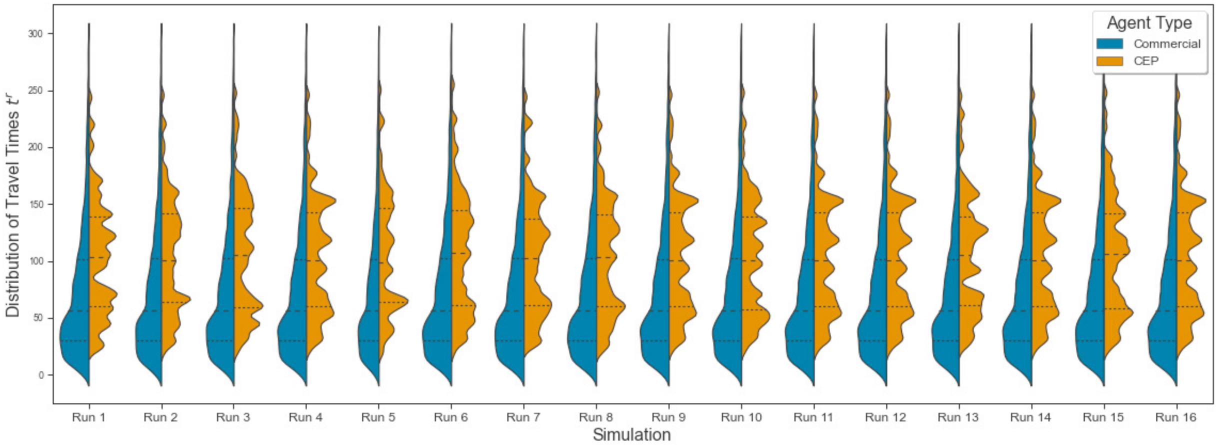

Fig. 4: Distribution of travel time t^r for commercial agents differentiated by type with $r \in R$.

To analyze the variability of travel times t^r in detail, we introduced the coefficient of variation $c_v(t^r)$ for different *LOA* as a measure of variability and applied it to our data set. Figure 5 illustrates the characteristics that led to a particularly high $c_v(t^r)_{LOA_3}$ in our simulation case study.

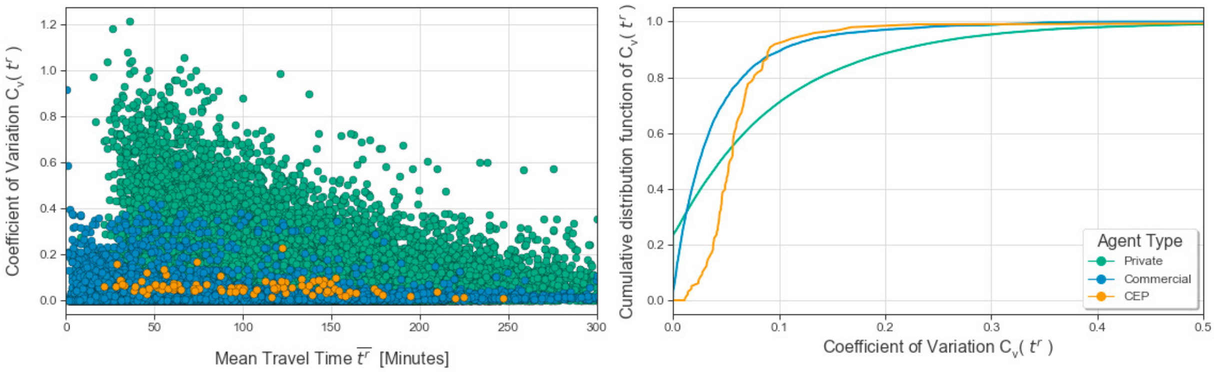

Fig. 5: Distribution of the variation coefficient $c_v(t^r)$ in relation to the average travel time combined with the empirical cumulative distribution function of the variation coefficient $c_v(t^r)$ with $r \in R$.

The various travel time distributions observed previously are evident in the variation of measured travel time values across all agents and simulation runs. The scatter plot indicates that higher $c_v(t^r)$ values usually occur at lower average travel times. Since the analysis of travel time frequencies shows that most of the agents' travel time tend to decline within this range of lower travel times, the observed clustering can be partly explained by the correspondingly larger sample size. It is apparent that several agent's travel times varies considerably between the simulation runs. The maximum values $c_v(t^r)$ differ significantly between agent types, i.e. $c_v(t^r_p)_{(max)} = 1.212$, $c_v(t^r_{ct})_{(max)} = 0.917$ and $c_v(t^r_{cep})_{(max)} = 0.230$. In total, only nine agents show a value of $c_v(t^r) > 1$. A $c_v(t^r) > 0.5$ can be observed for 741 agents (0.1 %).

Comparing agent types, the travel times of private agents show up the highest rate of deviation. This trend is also evident in the cumulative frequency distribution of $c_v(t^r)$. Analogous to the number of different states per agent, 71.0 % of the agents of the individual traffic ($n = 52.820$) have a value of $c_v(t^r_p) > 0.1$ across all simulation runs. For CEP-vehicles, it is close to 89.6 % ($n = 88$). The corresponding results indicate that all of these vehicles change their travel time in every simulation run. However, this variance is smaller compared to the other agent types and the resulting travel times are more consistent.

Table 1: Resulting distribution of the coefficient of variation $c_v(t^r)$ with $r \in R$.

Level of aggregation		Coefficient of variation $c_v(t^r)$	$\overline{c_v(t^r)}$	Standard deviation $\sigma^{c_v(t^r)}$	$Q1^{c_v(t^r)}$	$Median^{c_v(t^r)}$	$Q3^{c_v(t^r)}$
LOA_1							
	Total private agents of Hanover	0.0015	-	-	-	-	-
	Total commercial agents of Hanover	0.0009	-	-	-	-	-
LOA_2							
	Private agents per districts	-	0.0058	0.0031	0.0037	0.0049	0.0064
LOA_3							
	Individual agents of private traffic	-	0.0818	0.1074	0.0028	0.0449	0.1156
	Individual agents of commercial traffic	-	0.0437	0.0579	0.0092	0.0250	0.0546

After we were able to show that different travel time distributions occur using identical simulation input, we started investigating the thresholds for stable simulation results. We determined the value of $c_v(t^r)$ for each agent on LOA_3 differentiated by type. CEP vehicles were included in commercial traffic. For a better comparability, we averaged $c_v(t^r)$ across all agents. Evaluating LOA_2 on district level, we only included private agents because commercial vehicles usually start at specific companies with their tour and are therefore not so widely distributed over the simulation area. The travel times of the corresponding agents were averaged per district and the variability of this average value was examined and a mean value of $c_v(t^r)$ with $t^r \in d$ was calculated across all districts. LOA_1 is the variation of the global average travel time across all simulation runs differentiated by individual and commercial agents. The results are summarized in Table 1. For LOA_2 and LOA_3 statistical parameters of the distribution of $c_v(t^r)$ are provided since a single $c_v(t^r)$ value was calculated for each district or agent of Hanover.

Our results support our initial assumptions and the findings from the literature review. The global mean of the average travel times of all private agents from Hanover remains almost constant over all simulations runs $c_v(t_p^r)_{LOA_1} = 0.0015$. The variation of the commercial agent travel times are smaller with a $c_v(t_{ct}^r)_{LOA_1} = 0.0009$. Observed variation at district level increases slightly $c_v(t_p^r)_{LOA_2} = 0.0058$ and the analysis of each agent individually results in the highest observed variation of travel times $c_v(t_p^r)_{LOA_3} = 0.0818$. As a comparison of the variation of the aggregated travel times per district and the corresponding variation of the agents living in this district, we grouped the $c_v(t_p^r)_{LOA_3}$ values by the agent's home district (Figure 6).

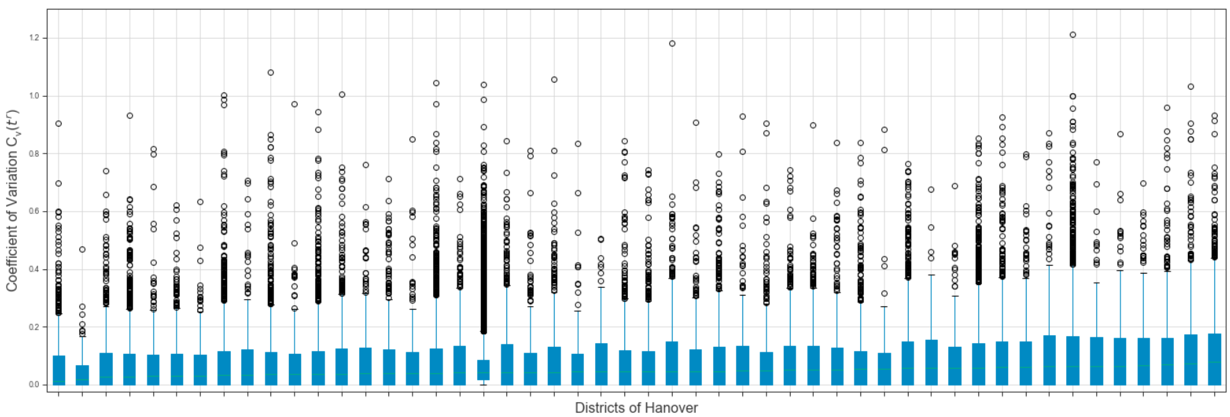

Fig. 6: Boxplot of variation of individual travel times $C_v(t^r)$ grouped by the agent's home location with $r \in R$.

Subsequently, we compared the individual agent travel time variability on LOA_3 with the variability of the aggregated travel times on LOA_2. The mean distribution of travel times for all agents living in the corresponding district varies between $c_v(t_p^r)_{LOA_{1(min)}} = \overline{0.048}$ and $c_v(t_p^r)_{LOA_{1(max)}} = \overline{0.114}$. The corresponding results for $c_v(t_p^r)_{LOA_2}$ are 0.007

and 0.005. Although the agents of a district show a variation of their corresponding travel times over all simulation runs, the aggregate travel time of all residents of the district varies less. The results of our case study show a compensation of the variations of travel times of individual agents on the aggregate dimension of districts. Thus, the more aggregated evaluation values are stabilizing rather fast at one level. The results imply that a prediction about these global parameters, especially on LOA_1 and LOA_2, can be made using an average value of only a few simulation runs. For practical work with MATSim, it is of interest how many simulation runs are necessary to determine the adequate value for the corresponding LOA with a desired accuracy.

5. Prediction of Required Number of Simulation Runs

To allow the derivation of generally valid indications from our results, we investigated how many simulations are necessary to arrive at robust mean values at the three aggregation levels defined. We applied the convergence of subsequent mean values $t_n \longrightarrow t_c$ to our data set by forming a moving mean value t_n with a progressing number of simulations. As soon as the deviation of the calculated mean value to the convergence mean value t_c was less than one percent, we considered the obtained mean value to be robust. However, the 16 simulation runs we performed were not sufficient to achieve a robust mean value. Despite this, to predict the number of simulations at which a robust mean is reached, we used our observed travel time distributions for each agent to generate artificial simulation results. This process was continuously repeated to replicate the observed travel time distribution. We consider this methodology to be valid because running a large set of simulations with MATSim to explore the needed number of simulation runs is not practical due to the comparatively long computation times.

The calculated distributions are summarized in Figure 7. Our first investigations indicate that the travel time values on LOA_1 and LOA_2 are already robust after one iteration. Thus, this robustness occurs for aggregated results. For $LOA_{3,p}$ 31 simulations were in average sufficient to reach a robust mean. At the maximum 36 runs were necessary. Additionally, the graph shows the development of the mean value convergence for commercial vehicles and, as a subset of this, for CEP-vehicles. The function of $LOA_{3,ct}$ develops similar to $LOA_{3,p}$ with a wider range of variation, even though the the robust mean value was in average reached earlier after 20 runs.

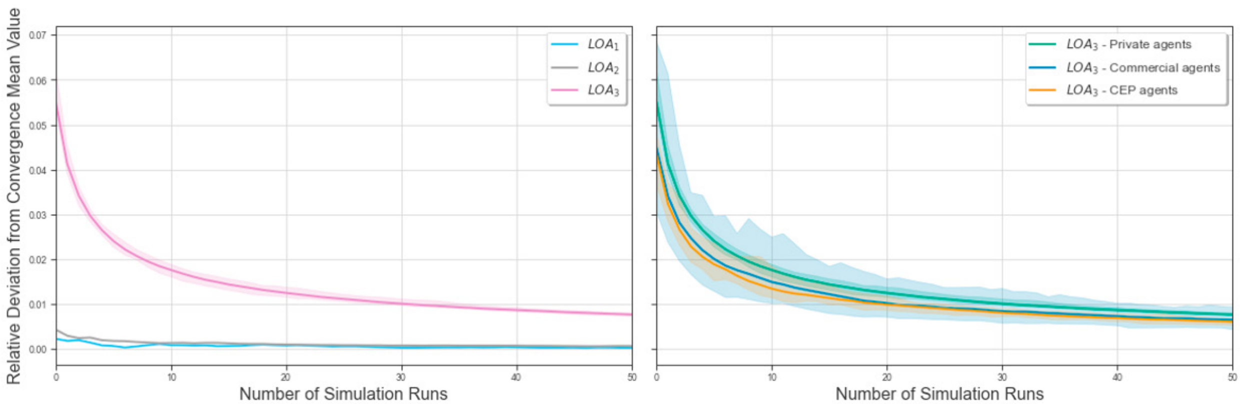

Fig. 7: Relative deviation from calculated convergence mean values for different LOA.

6. Conclusion and Future Work

The travel times of a MATSim simulation vary despite constant input parameters. We developed a recommendation for the needed number of simulation runs according to different aggregation levels. Our aim was to obtain results with a desired reliability of one percent deviation from the predicted travel time values. Accordingly, the variation of the travel times decreases as the aggregation level increases, while global aggregated parameters such as the average travel time remain approximately constant throughout the simulation. By analyzing the converging average, we were able to show that a single simulation is sufficient for an aggregated evaluation of travel times. These results are in line with previous contributions in this research area, since aggregated macroscopic data is often used to validate MATSim model results (Kagho et al., 2020). However, the analysis of the coefficient of variation also showed varying travel times of individual agents per simulation run. This can be particularly important when evaluating simulations focused on specific population groups with comparatively small sample sizes. A possible evaluation case applies for CEP traffic. These vehicles are part of commercial transport and thus have a small number of vehicles compared to the private traffic. Our simulations illustrated that the travel times of the freight agents tend to be relatively constant, although the travel times of the CEP vehicles still varies. Outliers can change the overall result due to the small size of the sample. For these sample sizes our results lead us to recommend to average at least the results of two simulation runs to reduce the variability of the evaluated travel times.

MATSim simulation runs are computationally expensive. Due to this, MATSim models are often scaled down. The variation of the individual agent travel times on LOA_3 thus has a higher influence on the aggregated values and leads to an inherent error. Consequently, our findings support the work of Llorca and Moeckel (2019), who observed different travel time distributions for smaller scale factors.

The objective of our future work is to provide an overview of the variance of a MATSim model in relation to the defined level of aggregation to allow more accurate evaluations with MATSim. Travel times vary depending on the agent types. Thus, it is appropriate to investigate attributes causing a corresponding variability and finally predicting the expected error for certain groups of agents. In addition, our artificial generation of travel time distributions, respectively simulation runs, must be validated with further simulation runs in order to be able to determine the predicted values more precisely.

Acknowledgement

The scientific research published in this article is granted by the Federal Ministry of Education and Research Germany for projects USEfUL and USEfUL XT (grant ID 03SF0547 & 03SF0609). The authors cordially thank the partners and funding agency.

References

Arentze, T., Timmermans, H., 2004. Albatross: A learning based transportation oriented simulation system. Transportation Research Part B: Methodological 38, 613–633.

Bao, Q., Kochan, B., Bellemans, T., Janssens, D., Wets, G., 2015. Investigating micro-simulation error in activity-based travel demand forecasting: a case study of the feathers framework. Transportation Planning and Technology 38, 425–441.

Baustert, P., 2021. Development of an uncertainty analysis framework for model-based consequential life cycle assessment: Application to activity-based modelling and life cycle assessment of multimodal mobility. Ph.D. thesis. Built Environment.

Bienzeisler, L., Lelke, T., Wage, O., Thiel, F., Friedrich, B., 2020. Development of an agent-based transport model for the city of hanover using empirical mobility data and data fusion. Transportation Research Procedia 47, 99–106.

Castiglione, J., Freedman, J., Bradley, M., 2003. Systematic investigation of variability due to random simulation error in an activity-based microsimulation forecasting model. Transportation Research Record 1831, 76–88.

Cools, M., Kochan, B., Bellemans, T., Janssens, D., Wets, G., 2011. Assessment of the effect of micro-simulation error on key travel indices: evidence from the activity-based model feathers. Proceedings of the 90th Annual Meeting of the Transportation Research Board .

Flötteröd, G., 2016. Matsim as a monte carlo engine. The Multi-Agent Transport Simulation MATSim , 327–336.

Horni, A., Charypar, D., Axhausen, K.W., 2011. Variability in transport microsimulations investigated with the multi-agent transport simulation matsim. Arbeitsberichte Verkehrs- und Raumplanung 692.

Horni, A., Nagel, K., Axhausen, K.W., 2016. The Multi-Agent Transport Simulation MATSim. Ubiquity Press, London, GBR.

Kagho, G., Balac, M., Axhausen, K., 2020. Agent-based models in transport planning: Current state, issues, and expectations. Procedia Computer Science 170, 726–732.

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

Paulsen, M., Rasmussen, T.K., Nielsen, O.A., 2018. Output variability caused by random seeds in a multi-agent transport simulation model. Procedia Computer Science 130, 850–857.

Rasouli, S., Timmermans, H., 2012. Uncertainty in travel demand forecasting models: literature review and research agenda. Transportation Letters 4, 55–73.

Walker, W., Harremoës, P., Rotmans, J., Sluijs, J., Asselt, M., Janssen, P., Kraus, M., 2003. Defining uncertainty: A conceptual basis for uncertainty management in model-based decision support. Integrated Assessment 4.

Zilske, M., Schröer, S., Nagel, K., Liedtke, G., 2012. Adding freight traffic to matsim. VSP Working Paper 12-02 .