Variability in Transport Microsimulations Investigated With the Multi-Agent Transport Simulation MATSim

Working Paper

Author(s):
Horni, Andreas; Charypar, David; Axhausen, Kay W.

Publication date:
2011-08

Permanent link:
https://doi.org/10.3929/ethz-a-006686329

Originally published in:
Arbeitsberichte Verkehrs- und Raumplanung 692

Variability in Transport Microsimulations Investigated With the Multi-Agent Transport Simulation MATSim

Date of submission: 2011-08-1

Andreas Horni
Institute for Transport Planning and Systems (IVT), ETH Zürich, CH-8093 Zürich
phone: +41-44-633 31 51
fax: +41-44-633 10 57
horni@ivt.baug.ethz.ch

David Charypar
Institute for Transport Planning and Systems (IVT), ETH Zurich, CH-8093 Zurich
phone: +41-44-633 35 62
fax: +41-44-633 10 57
charypar@ivt.baug.ethz.ch

Kay.W. Axhausen
Institute for Transport Planning and Systems (IVT), ETH Zurich, CH-8093 Zurich
phone: +41-44-633 39 43
fax: +41-44-633 10 57
axhausen@ivt.baug.ethz.ch

Words: 6582 words + 4 figures = 7582 word equivalents

ABSTRACT

Transport microsimulations are stochastic. Randomness is, for example, introduced by the error terms of discrete choice models, a common component in utility-based microsimulations. This leads to random variability in results at all resolution levels. This paper's objectives is an analysis of this variability. As a very common and important aggregate measure in transport planning network link volumes are analyzed, based on MATSim simulation experiments.

Constrained by modeling and simulation costs, recent large-scale, high-resolution microsimulations are cross-sectional models. When looking at aggregate levels relevant to planning, for common statistics there is relatively little variability over multiple simulation runs. However, these models do not properly account for temporal variability. This is problematic because temporal variability measured in reality is substantial. Thus, considering extension of these cross-sectional microsimulation models to longitudinal models will be necessary in the near future. To support this, the paper also documents first insights about temporal variability and temporal correlations in microsimulations.

PROBLEM DESCRIPTION AND RESEARCH GOAL

Variability Analysis

Many transport microsimulations are based on utility maximization implemented by econometric discrete choice models *(1)*. These models contain systematic and random parts to reproduce observations, i.e, measured population choice distributions. In addition to other randomness sources, clearly, these random parts could potentially introduce substantial randomness, making variability analyses for microsimulations necessary.

Until recently, the utility function of MATSim was deterministic, i.e., it did not contain random error terms. Nevertheless, some randomness was introduced by the co-evolutionary algorithm and the mobility simulation, as described later. Now, as part of the recent destination choice integration for discretionary activities, the random error terms have been added, finally making MATSim fully compatible with discrete choice theory. Variability issues must now, at latest, be investigated for MATSim as for any other travel demand simulation.

The main objective of this paper is, on one side, a general analysis of variability in transport microsimulations, with emphasis on theoretical background. On the other side, random variability over multiple simulation runs of agents' utilities and simulated network link volumes is analyzed for MATSim in the Zurich scenario, a frequently used and well calibrated model implementation. This illustrates theoretical considerations, but is interesting in its own right as well, as link volumes are also a very common and important measure of model validation and policy evaluation. Thus, the results are also relevant for simulation practice.

The goal of this paper is well summarized by *(2)*: "It would be useful to conduct analyses similar to those presented here with other model systems, both to examine the transferability of the conclusions and to provide analysis specific to those models for future reference as they are used in application."

Temporal Variability: Toward a Longitudinal Model

Most recent large-scale transport microsimulations are cross-sectional models. They are primarily designed to capture inter-personal variability and some of them intra-day dynamics as well. But intra-personal (i.e., temporal) variability beyond a single day is missing in these models. Travel demand, however, also features substantial mid- to long-term intra-personal variability (see e.g., *(3, 4)*). Its sources are manifold and cannot be recapped in this paper.

Clearly, to model both inter-personal *and* intra-personal variability, a longitudinal model is optimal. The main reason, why todays large-scale transport microsimulations are designed as cross-sectional and not as a longitudinal model, are probably the very high computation costs already incured by cross-sectional modeling. With MATSim, for example, simulating Switzerland (7 million person days) takes several days, even on large high performance computers *(5, 6)*. The simulation of this paper's 30 runs took 30 $runs \times 4\ days/run = 120\ days$ of runtime, where 30 runs are the minimum statistically. Other reasons might include very high modeling costs and gaps in research for longitudinal microsimulation models.

But with ever-increasing computer power, it will be feasible to run longitudinal models in the near future. Thus, while looking at model variability, this paper's secondary goal is assessing strategies to extend MATSim and other microsimulations by multi-day dynamics.

RANDOM VARIABILITY IN MICROSIMULATIONS

There are different types of variability in microsimulations. In this section *random* variability is investigated. Other types of variability are detailed in a later section. The stochastic nature of microsimulations demands that results are given based on *multiple runs* performed with varying random seeds. The fluctuations between these runs form the random variability. To report microsimulation results, ordinary statistical measures like standard deviation, sampling error level, or even, better confidence intervals should be applied (Section *Random Variability, Sampling Error and Confidence Interval*). Important when dealing with variability is its dependency on the aggregation level as shown in Section *Random Variability and Aggregation*. This section here concludes by scrutinizing handling of random variability in microsimulation practice, in previous work and in this paper.

Random Variability and Sampling: Microsimulation as Sampling Tool

Not all measured behavioral variability *is* systematic and not all systematic variability can be identified, or observed, as such. Some decisions are inherently random, meaning they are performed purely by chance; for other decisions, the modeler just lacks knowledge about decision makers' idiosyncratic rationales. Thus, as mentioned above, discrete choice models usually contain a *systematic* and a *random* part *(1)*. The models' application is based on randomly drawing from random error distributions.

Thus, results based on discrete choice models are essentially *random variables*. Parameters of their distributions (such as mean or standard deviation) are usually estimated based on *random sampling*. Most utility-based microsimulations are based on discrete choice models. This means that microsimulation results, e.g., link volumes, are also *random variables*, and, as expressed by *(7)*, microsimulations are "fundamentally an exercise in sampling". For microsimulations, the population is the set of all possible microsimulation runs, applying different random seeds. This set is infinite. A *random sample*, accordingly, is a random sub-set of runs. One run represents one realization of a random variable.

Random Variability, Sampling Error and Confidence Interval

Parameter estimates (population statistics), generated by random sampling, are subject to a *sampling error*, also known as *standard error*. The sampling error depends on sample size and *population variability*. While the modeler specifies sample size, population variability needs to be estimated. Finding a reference point on this variability using MATSim is the goal of this paper.

Confidence intervals, i.e., interval estimates, are the preferred means to report statistical estimates. Sampling error—and thus variability—also plays a central role in the confidence interval. This is described in detail below, as this paper is also intended to be a general basis for further variability analyses. For the sake of illustration, the *mean* is chosen as an example parameter. Similar applies for other parameters.

Sampling Error

Assuming a probability distribution given by the density function $f(x)$ with finite mean μ and *finite variance* σ^2, the standard error or sampling error σ_s for the mean of $f(x)$ is given as (see also Figure 1):

$$\sigma_s = \frac{\hat{\sigma}}{\sqrt{n}}$$

where $\hat{\sigma}$ is estimated sample standard deviation of $f(x)$ and n is sample size. Sampling error is the standard deviation of the sampling distribution. Sampling distribution $f_s(\bar{x})$ is a theoretical construct generated by the individual means of *infinitely* many samples of size n drawn from $f(x)$. According to the central limit theorem, $f_s(\bar{x})$ is Gaussian for all $f(x)$ with finite variance. Derivation of the above formula is given in *(8)* and repeated here, as it is a central concept in microsimulations. Assuming that sample means are independent realizations of the random variable M, standard error is the standard deviation of M:

$$M = \frac{1}{n}(X_0 + X_1 + X_2 + ... + X_n)$$

after rearranging:

$$M = \frac{X_1}{n} + \frac{X_1}{n} + ... + \frac{X_n}{n}$$

M has variance:

$$Var_M = Var\Big(\frac{X_0}{n} + \frac{X_1}{n} + ... + \frac{X_n}{n}\Big)$$

with $Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)$, where $Cov(X,Y) = 0$ for independent variables:

$$Var_M = Var\Big(\frac{X_0}{n}\Big) + Var\Big(\frac{X_1}{n}\Big) + ... + Var\Big(\frac{X_n}{n}\Big)$$

with $Var(aX) = a^2 Var(X)$:

$$Var_M = \frac{1}{n^2}Var(X_0) + \frac{1}{n^2}Var(X_1) + ... + \frac{1}{n^2}Var(X_n)$$

Applying $Var(X_i) = \hat{\sigma}^2$ and rearranging gives:

$$Var_M = \frac{1}{n^2}(\hat{\sigma}^2 + \hat{\sigma}^2 + ... + \hat{\sigma}^2)$$

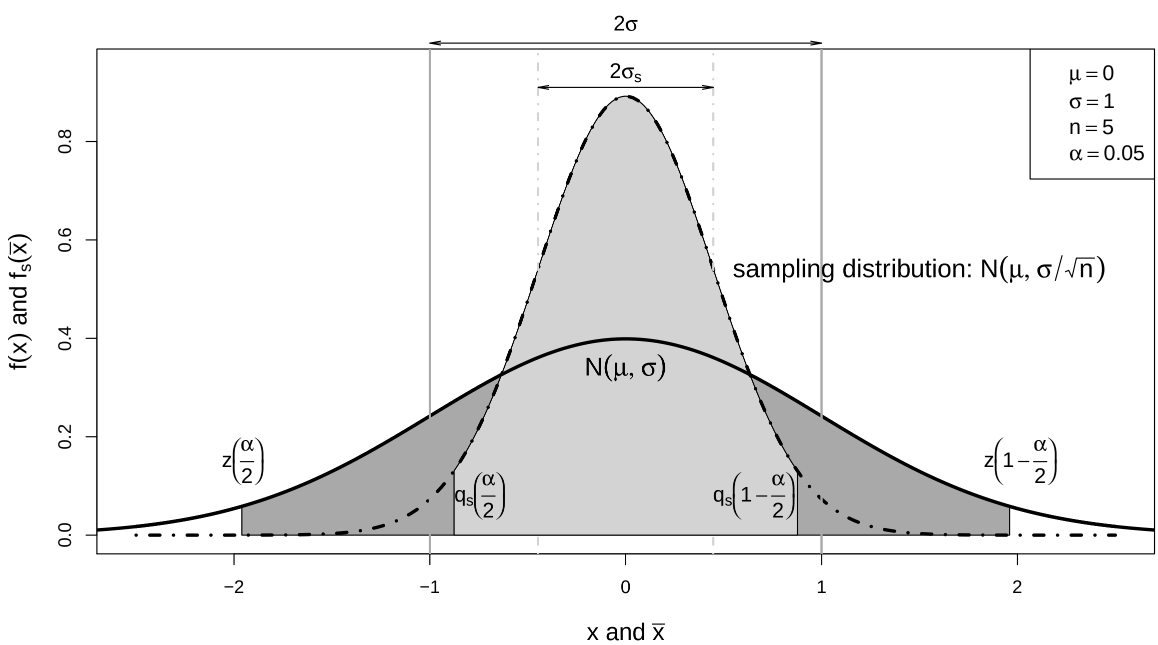
$$Var_M = \frac{1}{n^2}n\hat{\sigma}^2$$

Standard deviation of M, i.e., the standard error of the sampling distribution is:

$$\sigma_s = \sqrt{Var_M} = \sqrt{\Big(\frac{1}{n^2}n\hat{\sigma}^2\Big)} = \frac{\hat{\sigma}}{\sqrt{n}}$$

As mentioned above, sampling error is dependent on sample size $\sqrt{n}$ and population variability $\hat{\sigma}$ investigated in this paper.

FIGURE 1 Sampling Distribution, Sampling Error, and Confidence Interval



Confidence Interval

The confidence interval CI for the parameter θ of $f(x)$ is usually given as:

$$CI = [\hat{\theta} \pm \psi]$$

where $\hat{\theta}$ is an estimate of θ and ψ is the margin of error. In our case $\theta := \mu$.

The margin of error ψ is given as:

$$\psi = q(\alpha)\frac{\hat{\sigma}}{\sqrt{n}}$$

where $1 - \alpha$ is the confidence level, $q(\alpha)$ is the α-quantile of $f(x)$, n is sample size and $\hat{\sigma}$ is the sample standard deviation, which quantifies variability present in the sample. For large n (> 30), $q(\alpha)$ can be approximated by the quantile of the standard normal distribution $z(\alpha)$, according to the central limit theorem.

It is now apparent why sampling error appears in the confidence interval CI. Both the confidence interval and the sampling distribution make a statement about the estimated parameter $\hat{\theta}$. Sampling error simply transforms the quantiles of standard normal distribution $z(\alpha)$ to the respective quantiles of the *sampling distribution* $q_s(\alpha)$.

Ideally, microsimulation results should be accompanied by a confidence interval. For a given error level, the required number of runs n can be derived. This is straight-forward at high aggregation levels. At *low levels*, however, this is non-trivial. For example, the investigation in this paper encompasses 123 links, each with 24 hourly volumes. Every hour on every link has its own variability and averaging does not necessarily lead to a meaningful statement. Essentially, for every link and every hour, a confidence interval should be given. Furthermore, it is not yet clear which of these interval defines the required number of runs n. Methods to analyze, summarize and present large number of confidence intervals for the microsimulation context need to be developed in the future. In this work, the coefficient of variation (as defined later) is used to report results variability.

Random Variability and Aggregation

Random variability is dependent on aggregation level. Acknowledging this is important for the variability assessment and the choice of a resolution level in policy studies. As link volumes are an aggregate, influence of aggregation on variability is also directly relevant for this paper's results.

The confidence interval increases with the specific attribute's random variability. For behavioral models, individual variability at the person level is usually large, stemming from a large decision space for every decision-maker spanned by the choice dimensions: time, route, mode, destination, and more recently, activity chain choice. With increasing aggregate size, variability decreases in relative terms; i.e., absolute variability grows, but, relative variability (in relation to the estimate parameter) decreases. In general, the higher the aggregation level, the fewer runs are required, as shown by a generic example:

Let us assume that decision makers face two alternatives. The choice of person i for one of these alternatives can be described with a Bernoulli variable X_i which takes the values 1 for one alternative and 0 for the other alternative. The choice probability for the first alternative shall be p, for the other alternative $1 - p$. The mean is $\mu_i = p$ and the standard deviation is $\sigma_i = \sqrt{p(1-p)}$.

For an aggregate of $\tilde{n}$ decision-makers, each described by X_i the following holds.

Mean of an Aggregate

The mean of this aggregate is a random variable X_{avg} with $\mu_{avg} = \frac{1}{\tilde{n}}\tilde{n}p = p$ and standard deviation

$$\sigma_{avg} = \sqrt{Var\left(\frac{1}{\tilde{n}}\sum_{i=0}^{\tilde{n}} X_i\right)}$$

Assuming independent choices with $Cov(X, Y) = 0$ this gives:

$$\sigma_{avg} = \sqrt{\frac{1}{\tilde{n}^2}\sum_{i=0}^{\tilde{n}} \mathrm{Var}(X_i)}$$

$$\sigma_{avg} = \sqrt{\frac{1}{\tilde{n}^2}\tilde{n}\,\mathrm{Var}(X_i)}$$

$$\sigma_{avg} = \sqrt{\frac{1}{\tilde{n}^2}\tilde{n}p(1-p)}$$

$$\sigma_{avg} = \frac{\sqrt{p(1-p)}}{\sqrt{\tilde{n}}} = \frac{\sigma_i}{\sqrt{\tilde{n}}}$$

The standard deviation of a single person's decision is σ_i. The standard deviation of an aggregate of decisions is smaller by $\tilde{n}$, i.e., variability decreases with aggregates' size, meaning that fewer random runs n are required to reach a given error level for the aggregate than for an individual person.

Sum of an Aggregate

The sum of an aggregate is a random variable X_{sum} with $\mu_{sum} = \tilde{n}p$ and standard deviation

$$\sigma_{sum} = \sqrt{Var\Big(\sum_{i=0}^{\tilde{n}} X_i\Big)} = \sqrt{\tilde{n}p(1-p)}$$

A sum of Bernoulli trials is described by the Binomial distribution. Showing that the required number of runs is reduced with larger aggregates for sums is more complicated than for aggregates' averages. The variance of the sum grows linearly with $\tilde{n}$. The standard deviation of this sum grows with $\sqrt{\tilde{n}}$. However, standard deviation can be normalized with the estimated parameter using the following argument. When defining a confidence interval for a population statistic, the margin of error ψ is reasonably chosen *relative* to the this statistic. In other words, the margin of error is given as a relative percentage of the estimate.

Here, normalizing the standard deviation by the mean gives:

$$\sigma_{sum,normalized} = \frac{\sigma_{sum}}{\mu_{sum}} = \frac{\sqrt{\tilde{n}p(1-p)}}{\tilde{n}p} = \sqrt{\frac{1-p}{\tilde{n}p}} \tag{1}$$

The normalization for an *individual* decision described by X_i gives:

$$\sigma_{i,normalized} = \frac{\sigma_i}{\mu_i} = \sqrt{\frac{1-p}{p}}$$

Joining the last two equations gives:

$$\sigma_{sum,normalized} = \frac{\sigma_{i,normalized}}{\sqrt{\tilde{n}}}$$

It can be seen that, with respect to the mean, relative normalized standard deviation $\sigma_{sum,normalized}$ decreases for the aggregates compared to individual decisions described by $\sigma_{i,normalized}$. Thus, required number of runs n decreases with increasing aggregate size for both the average and the sum. Note, that here, the variability itself (quantified by the standard *deviation*), and not only the standard *error* is reduced. Clearly, the applicability of these statements is perfect for independent variables and looses validity with increasing correlation between the observations.

Handling Random Variability: State of Practice and Previous Work

Large-scale microsimulation results are often given on the basis of one single run *(9)*, due to very high computation costs. Strictly speaking, this does not represent a valid point estimate let alone an interval estimate. Nevertheless, the procedure is productive, as policy decisions based on a single microsimulation run are preferable to those lacking this information.

Furthermore, relying on a single simulation run is defensible as long as results are given at the appropriate aggregation level. As shown earlier, aggregation generally reduces variability, meaning that even for results based on one single run, aggregation helps reduce implied confidence intervals such that they might be acceptably small.

However, using a single simulation run and relying exclusively on aggregation to control sampling error is problematic, especially in the context of spatial correlations. Doing aggregation over, e.g., an area including both rural and urban sub-areas with very different infrastructure levels is not productive. Increasing the aggregate's size (to reduce sampling error), simultaneously introduces variability, necessitating even larger aggregates. This can be a problem, as aggregation reduces model resolution. Many current planning questions (e. g. road pricing) require a certain model resolution. Concluding, this means that relying on a single simulation run definitely has its limits.

For a few microsimulations, variability issues have been investigated or discussed *(10, 11, 12, 2, 13, 14, 15, 16)*. The investigations focus on the required number of microsimulation runs to reach "stable results". Random seeds are mutated where inputs are held constant. The papers conclude that sampling error is essentially a *non-issue* for these simulators and the investigated resolution levels, i.e., only a relatively small number of simulation runs are required for reliable results.

Random Variability in this Paper

This paper investigates whether or not previous studies' general findings can be confirmed.

Amount of variability introduced by the random term at a person level is controlled by estimation procedure and is relatively large. At the population level, amount of variability is expected to be relatively small, according to the *Random Variability and Aggregation* section. In general, the amount of variability "transferred" from the individual level to aggregate levels (such as link volumes) decreases. However, microsimulations contain many non-linear components such that small changes at one level may have very large effects on a different level. Additionally, applied aggregations are spatially heterogeneous because they are usually done on a network. Thus, the resulting amount of variability on an aggregate level cannot be estimated in a deductive manner, i.e., it is not known a priori. In the example above, even the probability p is unknown. Instead, experiments are required for quantification, achieved in this paper by running multiple simulation runs with different random seeds and constant inputs.

FURTHER TYPES OF VARIABILITY IN MICROSIMULATIONS

Previous sections focused on inter-run variability (random variability) of microsimulation results. There are, however, also other variability types, incorporated in microsimulations by mechanisms other than random sampling. Apart from the temporal variability, these types are not the main focus of this work. They are described briefly here, for a comprehensive overview and because they are important for the overall understanding of microsimulation variability issues.

Systematic Variability

It is essential to note the systematic variability between decision makers (inter-personal variability). Systematic differences in choice making are usually modeled by using socio-demographics as explanatory variables. They are *observed* by the models in contrast to random variability, which is *unobserved* (but measured) variability. For variability analyses, it is important to note that systematic variability does *not* contribute to the inter-run variability handled by sampling.

Temporal Variability

Another potentially important component of variability is temporal variability (intra-personal variability). While the intra-day dynamics are modeled in recent microsimulations, mid- to long-term variability is missing. Temporal variability can be seen as the result of temporal changes in the choice situation and persons' inherent motivations, as detailed later.

In terms of modeling it remains to investigate whether temporal variability is substantial, or choices are stable (i.e., repetitive). At a disaggregate level, it is already clear that temporal variability is substantial (see references above). To contribute in answering this question at a an aggregate level, hourly Swiss traffic count data are analyzed in this paper. It is shown that temporal variability is substantial and should be taken into account when modeling variability.

The next question is, how to model temporal variability. The optimal model clearly is longitudinal. This, however, incurs very high modeling and simulation costs.

Aiming for small modeling costs, the next logical step toward a longitudinal model could be simulation of multiple cross-sectional model runs with fixed inputs and varying random seeds. Temporal variability could thus be included in individual random error terms, as temporal variability is unobserved in cross-sectional models. However, this approach has two serious drawbacks.

First, any cross-sectional model estimated using data from one specific day actually includes a certain amount of temporal variability because persons' decisions are not perfectly synchronized. However, people behave differently in winter than in summer, for example; to capture *individual* temporal variability, one would also need to collect data for different periods of the year, quite similar as with a longitudinal model.

Second, people not only behave differently over time, but behavior is also influenced by general rhythms of life according to different seasons, the global economic situation, weather, etc. In a more abstract sense, this can be interpreted as *temporal correlations* between persons. These correlations substantially influence *aggregate* results' variability. In mathematical terms, this reads as follows. Given, for example, two random variables X_0 and X_1 representing an arbitrary time-dependent decision of individual 0 and individual 1, i.e., $X_0 = f_0(t)$ and $X_1 = f_1(t)$, the variance of two random variables is $Var(X_0 + X_1) = Var(X_0) + Var(X_1) + 2Cov(X_0, X_1)$. The covariance is non-zero for correlated variables; the covariance is greater than zero if variables are equidirectional. There are many transport-related decisions where individuals tend to have a positive correlation, i.e., $Cov(X_0, X_1) > 0$. This is caused by general life rhythms. There are also decisions where correlation is negative i.e, $Cov(X_0, X_1) < 0$. An example might be the avoidance of demand peaks, such as not visiting certain skiing resorts during school holidays. By analyzing the count data, shown later, it can be seen that the positive correlation predominates, increasing temporal aggregates' variability.

To summarize, cross-sectional models cannot adequately capture temporal variability. Exactly as it needs a network model to capture spatial correlations correctly, it needs model components reproducing general life rhythms to capture temporal correlations correctly. In other words, a longitudinal model is inevitable.

This conclusion also helps resolve the following controversy. For cross-sectional microsimulations, the simulated day can be interpreted in different ways, which is also an issue for MATSim developers. Some modelers interpret outcomes as just an arbitrary working day when data was collected. Others think, that outcomes represent the average working day and argue that typically the model inputs represent averages over a longer time period and that the incorporated choice models are estimated on data, not being truly longitudinal but still being collected at different days over a longer time period.

However, in a non-linear context, as given for microsimulations, one must adequately accounted for temporal variability; it is not the same if inputs or outputs are averaged, i.e, $f(\bar{x}) \neq f(\bar{x})$. In light of the problems with cross-sectional models formulated above, the authors prefer the first interpretation.

Endogenous and Exogenous Variability

In modeling, the distinction between endogenous and exogenous variability is very important. A comprehensive "world-model" has only endogenous variability. Clearly, no model can be comprehensive from its inception. At early stages of model development, some components must thus to be given exogenously. The goal is to successively incorporate them into the model. Model output variability is the product of input variability and model variability. Analyzing exogenous and endogenous variability gives the modeler a first idea how much variability to expect at the output. Cross-sectional microsimulation models, for example, clearly produce less variability than longitudinal models if inputs are held stable.

Variability of the Choice Situation

This section is not directly relevant for modeling, but it completes the variability analysis and may facilitate the further development of microsimulations.

Above, choice situation dynamics are mentioned. Choice situation is dependent on the decision maker's internal state and the state of the choice environment. This distinction is natural and common, although, strictly speaking, persons are also an inherent part of choice environment.

Following these logic, an active (person) and a passive component (environment) are present in the decision-making process. The decision-maker *perceives* the environment and makes a choice. The choice process is thus always composed of an action according to the person state and a "re-action" to the environment state. Whether all actions are also re-actions in the long run, i.e., whether the environment triggers *all* actions, is a philosophical question and will not be further discussed here.

Accordingly, behavioral variability is the result of variability of the person state and the environment. Person state variability is induced by personal motivations changing over time, such as needs, preferences or also personal experience. Environment variability is made up of temporal changes and spatial heterogeneity, including feedback from other transport system participants. As far as choice situation variability is systematic (observed), it does not contribute to inter-run variability.

VARIABILITY IN MATSIM

MATSim—In Brief

Before MATSim's variability is analyzed in detail below, a short introduction to the simulation framework is given.

MATSim is an activity-based, extendable, open source, multi-agent simulation toolkit implemented in JAVA and designed for large-scale scenarios and is a co-evolutionary model. A good overview of MATSim is given in *(17)*. In competition for space-time slots on transportation infrastructure with all other agents, every agent iteratively optimizes its daily activity chain by *trial and error*. Every agent possesses a fixed amount of day plans memory, where each plan is composed of a daily activity chain and an associated utility value (in MATSim, called *plan score*).

Before plans are executed on the infrastructure in the network loading simulation *(*e. g., *18)*, a certain share of agents (here 20%) is allowed to select and clone a plan and to subsequently modify this cloned plan.

If an agent ends up with too many plans (here set to "5 plans per agent"), the plan with the lowest score (configurable) is removed from the agent's memory. One iteration is completed by evaluating the agent's day described by the selected day plans.

If an agent has obtained a new plan, as described above, then that plan is selected for execution in the subsequent network loading. If the agent has *not* obtained a new plan, then the agent selects from existing plans. The selection model is configurable. In many MATSim investigations, a model generating a logit distribution is used. However, for this paper, agents will select the plan with the highest score.

Computation of plan score is compatible with micro-economic foundations. The basic MATSim utility function was formulated in *(19)* from the *Vickrey* model for road congestion as described in *(20)* and *(21)*. Utility of a plan described in detail in *(19)* is computed as the sum of all activity utilities plus the sum of all travel (dis)utilities.

Endogenous Variability

Endogenous choice dimensions currently consist of time *(22)*, route *(23)* and destination choice for discretionary activities *(24)*. Usually, these choices are modeled by drawing from a choice model composed of a systematic and a random part. In MATSim, the utility function for destination choice contains explicit random error terms. This introduces random variability as described above. The utility function for route and time choice does not (yet) contain a random error term. Nevertheless, the mobility simulation implicitly introduces randomness. Furthermore, a certain amount of randomness (i.e., unobserved heterogeneity) implicitly enters the model as *algorithmic variability* as follows. The co-evolutionary algorithm implemented in MATSim introduces random variability in two ways. The first source is algorithmic difficulties. For large-scale systems, finding the global optimum is not trivial. Starting from different initial points given by different random seeds, one might get stuck in local optima. Second, the co-evolutionary algorithm essentially assigns limited resources to persons in a random manner. This means, for example, that two identical persons with the same start and end location may end up with different routes or start times, according to the random order in which they undergo the replanning. Essentially, this means that a random term is added *implicitly* to the choices. The meaning of this variability is not yet fully understood in MATSim.

Further variability could possibly be introduced by infrastructure constraints. In MATSim, opening times are taken into account.

Exogenous Variability

Day chain structures and individual desired activity durations are exogenously assigned. They are derived—in an ad-hoc manner—from a PUS (see e.g., Balmer *et al. (6)*), here the National Travel Survey for the years 2000 and 2005 *(25)*. Spatial distribution of the populations' home and work locations are also given exogenously by the Swiss Census of Population 2000 *(26)*.

Constraints are taken into account when generating input; e.g., chains containing work activities are not assigned to children. However, apart from that, person attributes, such as household type or income are not yet taken into account. In other words, little variability is introduced by socio-demographics.

In MATSim, except for constraints and network, environment has no influence on choices; there is, for example, no weather or season modeled.

As it was done in previous studies, in this paper, as a first step, it is investigated how much *endogenous* variability is present. In other words, inputs are held constant while the random seeds are varied. Random seeds in this work influence time and route choice (both implicit) and destination choice (explicit). All simulation random seeds are varied simultaneously.

METHOD

Model variability is examined using the Zurich simulation scenario. The (aggregate) temporal variability measured in the real transport system is assessed and compared to simulation results using the annual Swiss road count data.

Real-world Scenario: Zurich Scenario

The Zurich scenario is frequently used in MATSim development, as well as in projects in Swiss planning practice *(*e.g., *6, 27)*. Simulation scenario demand is derived from the Swiss Census of Population 2000 *(26)* and the National Travel Survey for the years 2000 and 2005 *(25)*. A 10% sample of car traffic (including cross-border traffic) crossing the area delineated by a 30 km circle around Bellevue, a central location in Zurich is drawn, resulting in almost 68'000 simulated agents. Work now in progress will look at different sampling rates.

The activity location data set, comprising more than 10^6 home, work, education, shopping and leisure locations, is based on the Federal Enterprise Census 2001 *(28)* and the Swiss Census of Population 2000. The network from the Swiss National Transport Model *(29)* is used, which consists of 60'492 directed links and 24'180 nodes. A single day is simulated with 3.35 average number of trips per agent. In total, 25'896 shopping activities and 40'971 leisure activities are performed. The choice setting comprises the three dimensions, time, route and destination choice for discretionary activities.

30 simulation runs of the Zurich scenario are performed with identical input, but varying random seeds, corresponding to the *method of replication* as described in *(30)*.

In this paper, the relative sample standard deviation expressed as a percentage is used. Except for the agents' utilities this is identical with the coefficient of variation (CV) mainly used in previous studies. The relative sample standard deviation is applied to not underestimate variability of the utilities, which can be negative.

Road Count Data

MATSim focuses on "regular" workdays. Thus, the count data are prepared as follows. A couple of filtering steps are applied (see also *(6)*): only Tuesdays, Wednesdays and Thursdays are included, while any public holidays are excluded. The days between Christmas and New Year are also filtered out and finally, only count values greater than zero are included. 600 unidirectional links are measured for Switzerland and 123 for the center of Zurich (defined here as the area within a 12 km radius around the Bellevue).

RESULTS

Cross-Sectional Random Variability at Different Aggregation Levels

Utilities

At person level, the average CV of the agents' executed plan utilities is approximately 3%. At population level, as expected, there is little variability between simulation results; mean utility (averaged over agents) of all executed plans of the final iteration 200 has a CV of 0.087 %. This shows empirically that aggregation actually reduces variability, as derived earlier.

Link Volumes

For this analysis, the 123 links with count stations are used. As mentioned earlier, link volumes are used here as they are a very important measure in transport planning. Link volumes represent an aggregate where the *sum* of the aggregate is computed. Thus, the conclusions of Section *Random Variability and Aggregation* are applicable.

The CV for the volumes (identical with relative sample standard deviation), is plotted as percentage *per link*. To clarify, a single point in the box plot represents the random variability of a single network link, meaning, that, to compute the relative standard deviations, every link is compared only with *itself*. In the scatter plots, daily and hourly link volumes are also plotted for every link compared with itself. The abscissa represents the *average* value over multiple runs or multiple iterations, where the ordinate represents the individual values.

Variability of *daily* volumes is shown in Figures 2(a) and 3(a). Consistent with previous work, relatively little variability exists at this resolution level. The simulated variability is smaller than the measured variability shown in Figure 3(b). One reason for that might be the missing temporal variability as discussed earlier.

Variability for *hourly* volumes is shown in Figures 2(b) and 3(c). One—respectively three—different hours are included, but values for other hours are very similar. In Figure 3(c), most variability is present for the time slot between 11-12. This is plausible as during this time period, share of discretionary activities is higher than for the other two hours and because in this paper, destination choice is performed only for discretionary activities.

In the *hourly resolution*, relatively high variability is observed. Initially, this is surprising, as previous studies conclude that the sampling error is a non-issue. A direct comparison with previous studies, however, is difficult. Random variability depends on the spatial and temporal resolution and the choice dimensions included in the model. For example, taking only route choice and daily volumes into account strongly reduces the degrees of freedom in the model, while every degree of freedom usually introduces randomness. The microsimulator also normally introduces randomness, but this is true for all models. In *(12, 11, 2)* only daily measures are investigated. Furthermore, in *(11)*, while including many choice dimensions, only population level is researched. In *(10, 15)* while evaluating hourly measures, only route choice is applied. Hackney *(*page 128ff *13)* applied only time and route choice and results are given for daily measures.

In conclusion, future research is needed along the following lines. Clearly, the first is verification of the newly implemented destination choice module. Also analyses incorporating only one single choice dimension should be done with MATSim. Additionally, the effect of running sample populations must be investigated. When calculating the CV of a certain measure, any scaling factor γ cancels out because γ is applied to both numerator ($\hat{\sigma}$) and denominator

($\hat{\mu}$). However, the fact that one agent decides for multiple persons is still true and represents a discretization error. Its effect on variability should be analyzed. Figure 2(b) shows that low-volume links tend to have larger relative inter-run variability. Thus, investigation is needed to ascertain whether an additional weighting by the absolute link volumes would be appropriate, especially when simulating population samples where one agent represents many persons.

Another potential source for the substantial inter-run variability is the substantial *intra*-run variability in Figures 2(c), 2(d), and 3(d). A large intra-run variability could indicate that the system has reached a utility plateau with many user equilibria close to each other, or that it has not yet reached equilibrium although the score is stable. Intra-run variability might also be created by the replanning modules based on random mutation. Note that to accelerate convergence of the destination choice module, the replanning share is comparatively large here (20% compared to 10% used earlier).

In addition to potential influence on inter-run variability, a large intra-run variability raises several problems for future work in its own right. For MATSim, strategies to reduce the replanning share or range when approaching equilibrium should be researched. Methods to assess the distance to an equilibrium state, which are being developed for MATSim, should be applied and further researched. Finally, further research on the existence and uniqueness of user equilibria in large-scale microsimulations is important to understand microsimulation variability better.

Temporal Variability and Temporal Correlations

In Figure 4, analysis of measured (i.e., counted) link volumes is given for both the whole year and single months. As above, sample standard deviation is plotted *per link*. I.e, a single point in the box plot represents temporal variability of a single network link, either for the whole year, or for a specific month. The hours 11-12 and 17-18 are shown as examples; similar patterns can be observed for all hours. Daily volumes are also reported.

The plots show that temporal variability in reality is substantial. It can also be seen that *temporal correlations* actually have a substantial influence on link volume variability as derived earlier. Yearly values show a larger variability than monthly values, meaning that a general rhythm of life (guided by, for example, the seasons) introduces substantial variability and should be taken into account explicitly in the model.

CONCLUSIONS AND OUTLOOK

This paper contributes to the ongoing research on microsimulation variability. The focus is on random variability and on temporal variability but other variability types are also discussed.

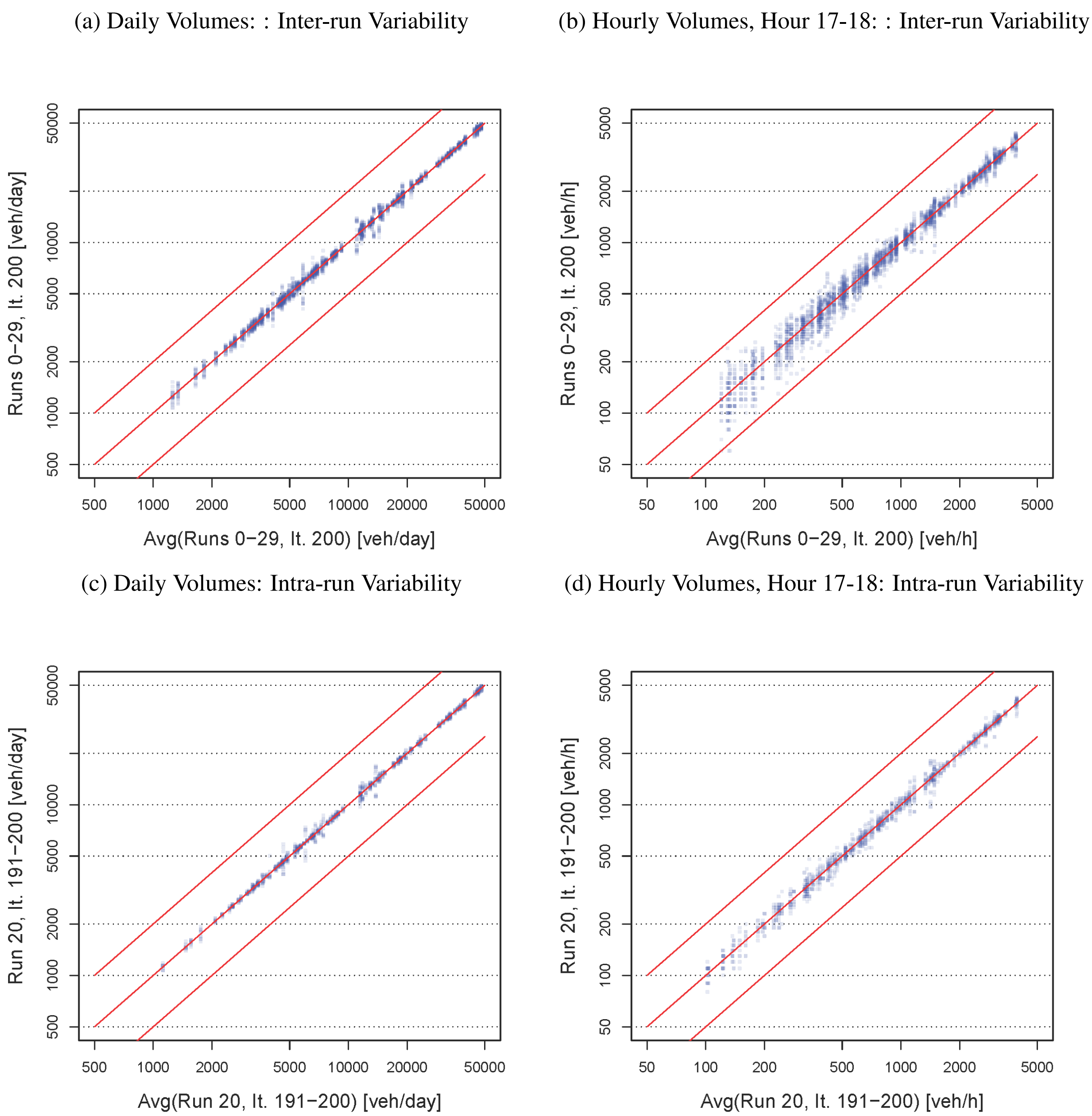
Results of this investigation are in line with previous work. *Daily* link volumes and agents' utilities show little variability such that actually few runs are necessary to achieve stable results.

However, *hourly* volumes show substantial variability. This is initially surprising but not implausible. The resolution is higher and/or there are more degrees of freedom in this experiment than in previous studies, suggesting that a higher variability must be expected. Nevertheless, verification work should be done in the future.

General, but also MATSim-specific, future research problems are identified, concerning primarily the population sampling rate and the MATSim intra-run variability.

Finally, the knowledge base for the improvement of normally cross-sectional large-scale

FIGURE 2 Simulated Link Volumes

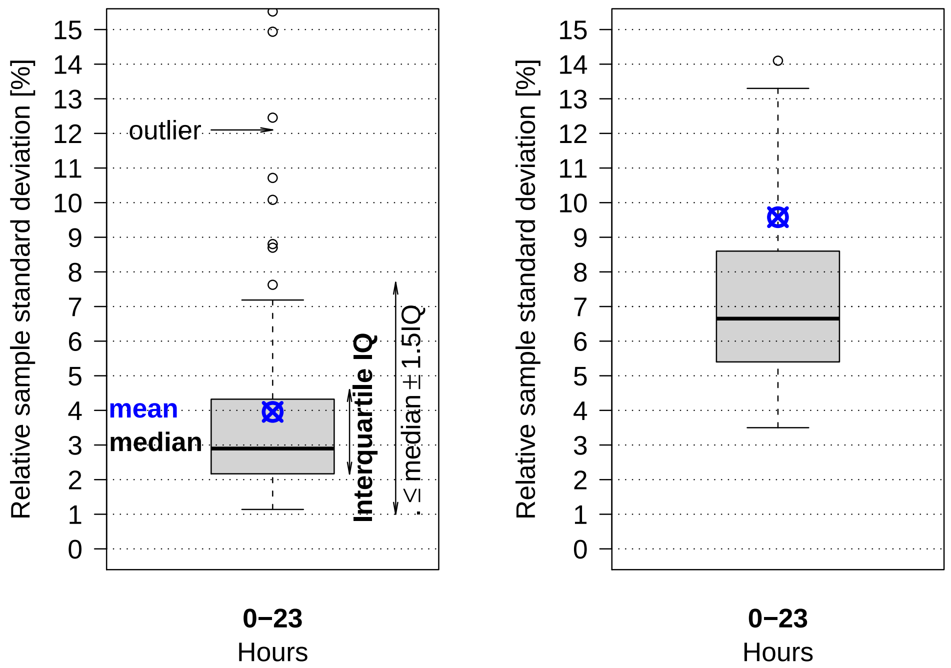


transport microsimulations toward longitudinal models is extended to eventually facilitate temporal variability modeling.

FIGURE 3 Simulated and Measured Link Volumes

(a) Simulated Daily Volumes: Inter-run Variability, Runs 0-29, Iteration 200

(b) Measured Daily Volumes: Temporal Variability Over One Year in the Region of Zurich.



(c) Simulated Hourly Volumes: Inter-run Variability, Runs 0-29, Iteration 200

(d) Simulated Hourly Volumes: Intra-run Variability (Run 20, Iterations 191-200)

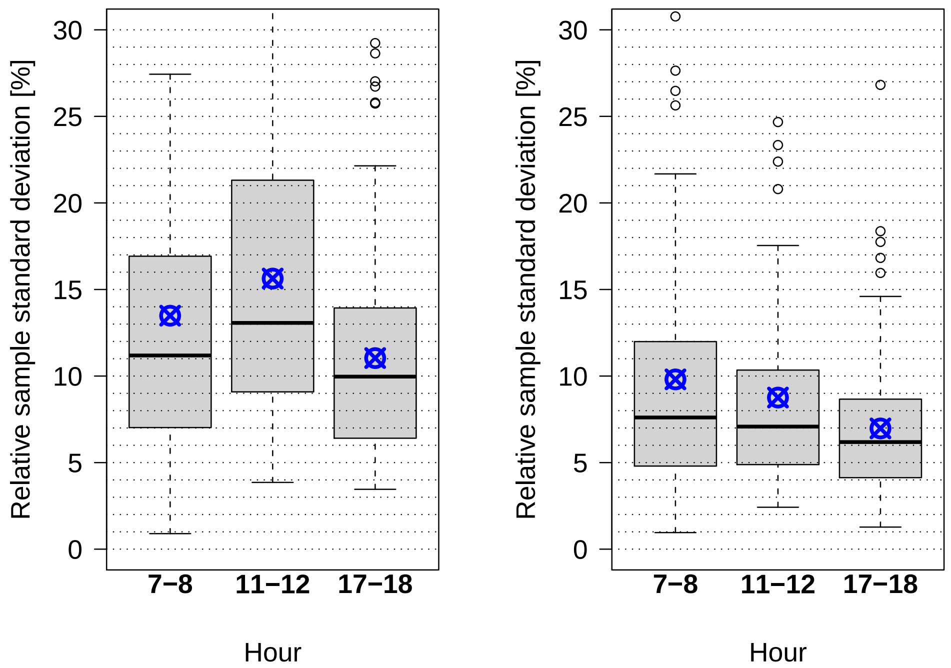
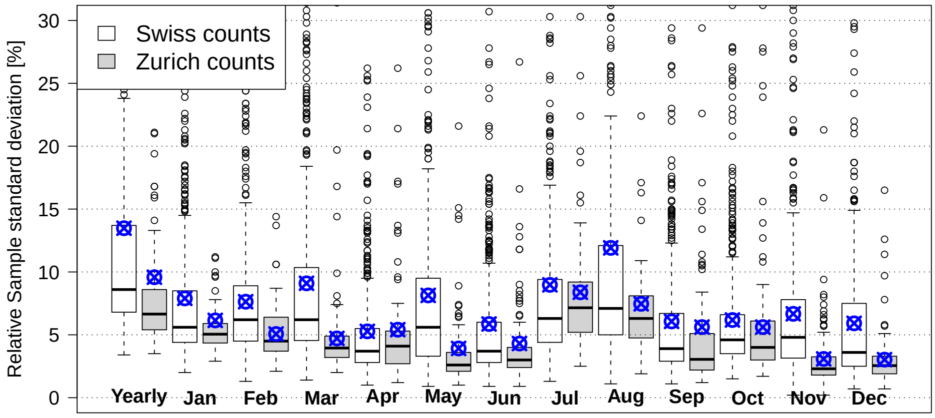
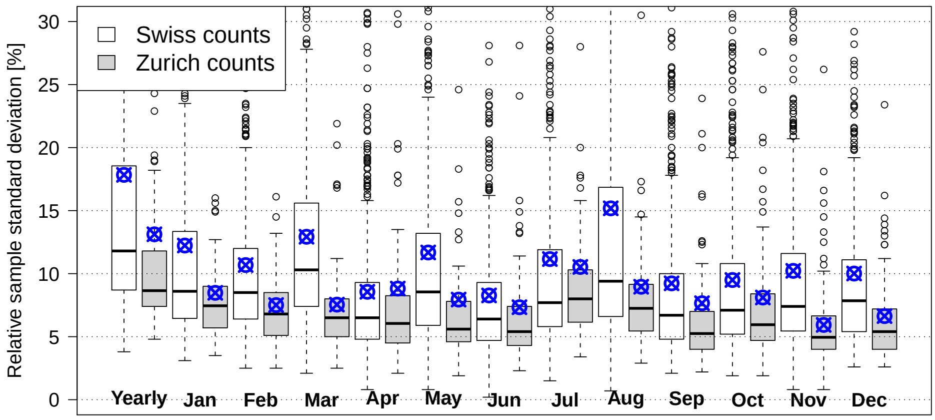


FIGURE 4 Measured Volumes

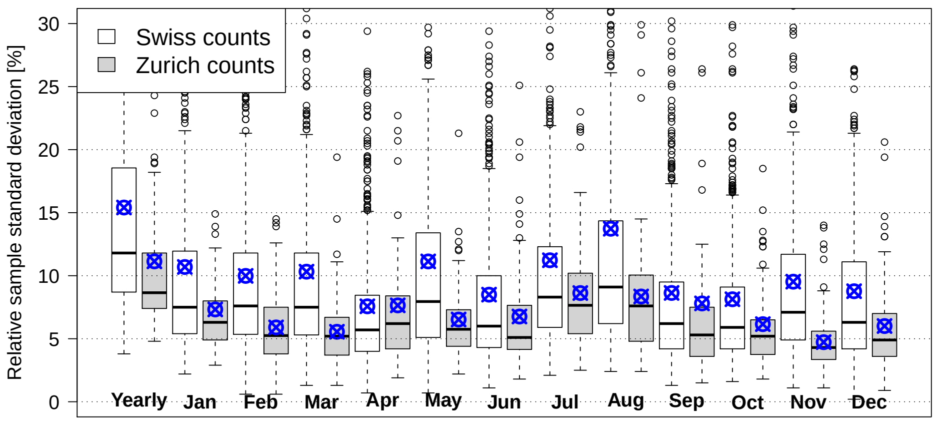
(a) Daily Volumes



(b) 11:00-12:00



(c) 17:00-18:00



REFERENCES

1. McFadden, D. (1978) Modeling the choice of residential location, in A. Karlqvist (ed.) *Spatial Interaction Theory and Residential Location*, 75–96, North-Holland, Amsterdam.

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

3. Kitamura, R., T. Yamamoto, Y. O. Susilo and K. W. Axhausen (2006) How routine is a routine? an analysis of the day-to-day variability in prism vertex location, *Transportation Research Part A: Policy and Practice*, **40** (3) 259–279.

4. Schlich, R. and K. W. Axhausen (2003) Habitual travel behaviour: Evidence from a six-week travel diary, *Transportation*, **30** (1) 13–36.

5. Balmer, M., K. Meister, R. A. Waraich, A. Horni, F. Ciari and K. W. Axhausen (2010) Agenten-basierte Simulation für location based services, *Final Report*, F&E Förderung: Science to Market, **KTI 8443.1 ESPP-ES**, Datapuls AG, IVT, ETH Zurich, Zurich, February 2010.

6. Balmer, M., A. Horni, K. Meister, F. Ciari, D. Charypar and K. W. Axhausen (2009) Wirkungen der Westumfahrung Zürich: Eine Analyse mit einer Agenten-basierten Mikrosimulation, *Final Report*, Baudirektion Kanton Zurich, IVT, ETH Zurich, Zurich, February 2009.

7. Wolf, D. A. (2001) The role of microsimulation in longitudinal data analysis, *Canadian Studies in Population*, **28** (2) 313–339.

8. Hutchinson, T. P. (1993) *Essentials of Statistical Methods in 41 Pages*, Rumsby Scientific Publishing, Adelaide.

9. Hale, D. (1997) How many NETSIM runs are enough?, *McTrans Newsletter*, **11** (3) 4–5.

10. Ziems, S. E., S. Bhargava, J. Plotz and R. M. Pendyala (2011) Stochastic variability in microsimulation modeling results and convergence of corridor-level characteristics, paper presented at the *90th Annual Meeting of the Transportation Research Board*, Washington, D.C.

11. Cools, M., B. Kochan, T. Bellemans, D. Janssens and G. Wets (2011) Assessment of the effect of micro-simulation error on key travel indices: Evidence from the activity-based model feathers, paper presented at the *90th Annual Meeting of the Transportation Research Board*, Washington, D.C.

12. Veldhuisen, J., H. J. P. Timmermans and L. Kapoen (2000) Microsimulation model of activiy patterns and traffic flows: Specification, validation tests, and monte carlo error, *Transportation Research Record*, **1706**, 126–135.

13. Hackney, J. K. (2009) Integration of social networks in a large-scale travel behavior microsimulation, Ph.D. Thesis, ETH Zurich, Zurich.

14. Vovsha, P., E. Petersen and R. Donnelly (2002) Microsimulation in travel demand modeling: Lessons learned from the New York best practice model, *Transportation Research Record*, **1805**, 68–77.

15. Lawe, S., J. Lobb, A. W. Sadek, S. Huang and C. Xie (2009) TRANSIMS implementation in Chittenden County, Vermont: Development, calibration, and preliminary sensitivity analysis, *Transportation Research Record*, **2132**, 113–121.

16. Gibb, J. and J. L. Bowman (2007) Convergence of an activity-based travel model system to equilibrium: Experimental designs and findings, paper presented at the *11th National Transportation Planning Applications Conference*, Daytona Beach, Florida, May 2007.

17. Balmer, M., K. W. Axhausen and K. Nagel (2006) An agent-based demand-modeling framework for large scale micro-simulations, *Transportation Research Record*, **1985**, 125–134.

18. Cetin, N. (2005) Large-scale parallel graph-based simulations, Ph.D. Thesis, ETH Zurich, Zurich.

19. Charypar, D. and K. Nagel (2005) Generating complete all-day activity plans with genetic algorithms, *Transportation*, **32** (4) 369–397.

20. Vickrey, W. S. (1969) Congestion theory and transport investment, *The American Economic Review*, **59** (2) 251–260.

21. Arnott, R., A. de Palma and R. Lindsey (1993) A structural model of peak-period congestion: A traffic bottleneck with elastic demand, *The American Economic Review*, **83** (1) 161–179.

22. Balmer, M., B. Raney and K. Nagel (2005) Adjustments of activity timing and duration in an agent-based traffic flow simulation, in H. J. P. Timmermans (ed.) *Progress in Activity-Based Analysis*, 91–114, Elsevier, Oxford.

23. Lefebvre, N. and M. Balmer (2007) Fast shortest path computation in time-dependent traffic networks, paper presented at the *7th Swiss Transport Research Conference*, Ascona, September 2007.

24. Horni, A., K. Nagel and K. W. Axhausen (2011) Hig-resolution destination choice in agent-based demand models, *Working Paper*, **682**, IVT, ETH Zurich, Zurich.

25. Swiss Federal Statistical Office (2006) *Ergebnisse des Mikrozensus 2005 zum Verkehrsverhalten*, Swiss Federal Statistical Office, Neuchatel.

26. Swiss Federal Statistical Office (2000) Eidgenössische Volkszählung 2000, `http://www.bfs.admin.ch/bfs/portal/de/index/infothek/erhebungen__quellen/blank/blank/vz/uebersicht.html`.

27. Horni, A., D. M. Scott, M. Balmer and K. W. Axhausen (2009) Location choice modeling for shopping and leisure activities with MATSim: Combining micro-simulation and time geography, *Transportation Research Record*, **2135**, 87–95.

28. Swiss Federal Statistical Office (2001) *Eidgenössische Betriebszählung 2001 - Sektoren 2 und 3, GEOSTAT Datenbeschreibung*, Swiss Federal Statistical Office, GEOSTAT, Neuchatel, `http://www.bfs.admin.ch/bfs/portal/de/index/dienstleistungen/geostat/datenbeschreibung/betriebszaehlung05.Document.111424.pdf`.

29. Vrtic, M., P. Fröhlich and K. W. Axhausen (2003) Schweizerische Netzmodelle für Strassen- und Schienenverkehr, in T. Bieger, C. Lässer and R. Maggi (eds.) *Jahrbuch 2002/2003 Schweizerische Verkehrswirtschaft*, 119–140, Schweizerische Verkehrswissenschaftliche Gesellschaft (SVWG), St. Gallen.

30. Benekohal, R. F. and G. Abu-Lebdeh (1994) Variability analysis of traffic simulation outputs: Practical approach for TRAF-NETSIM, *Transportation Research Record*, **1457**, 198–207.