Variability in transport microsimulations investigated for MATSim
Preliminary results

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

Publication Date:
2011

Permanent Link:
https://doi.org/10.3929/ethz-a-006482318

Rights / License:
In Copyright - Non-Commercial Use Permitted
Variability in Transport Microsimulations Investigated for MATSim: Preliminary Results

Andreas Horni
David Charypar
Kay.W. Axhausen

Transport and Spatial Planning  
April 2011

11th Swiss Transport Research Conference
Monte Verità / Ascona, May 11 – 13, 2011
Abstract

As part of the recent integration of destination choices for discretionary activities, mechanisms to explicitly treat unobserved heterogeneity in MATSim were developed. This potentially introduced a large amount of randomness to the model as so far the MATSim utility function was deterministic. Adding randomness to a model clearly makes necessary variability analysis. In this paper first steps in assessing the variability of MATSim results are undertaken. The results are preliminary and need comprehensive verification and elaboration.

Keywords
Variability analysis, autocorrelation, microsimulation
1 Problem Description and Research Goal

MATSim is based on utility maximization implemented by econometric discrete choice models (McFadden, 1978). However, until recently the utility function of MATSim was deterministic. As part of the recent integration of destination choices for discretionary activities the random error term has been added, finally making MATSim compatible with discrete choice theory. This potentially introduced a large amount of randomness to the model. Obviously, it is now at latest required to perform a variability analysis or in other words a sampling error analysis.

1.1 Sampling Error Analysis: State-of-Practice and State-of-Art

In practice large-scale microsimulation results are not seldom given on the basis of one single run (Hale, 1997) owed to the very high computation costs (see e.g., (Balmer et al., 2009)). E.g. in MATSim simulating 7 Million person days takes minimally $4.5 \times 50 = 225$ iterations $\sim 10$ days (Balmer et al., 2010, see). This procedure is justifiable as policy decisions based on a single microsimulation run are still preferable to those missing this peace of information. Furthermore, relying on a single simulation run is not necessarily a problem as long as the results are given at the appropriate aggregation level as aggregation in general reduces variance, i.e., reduces the sampling error as given by Equation 1.

However, when trying to reach a certain confidence, exclusively relying on aggregation done for a single simulation run is problematic. This is particularly true when very heterogeneous elements are aggregated. Making the aggregates very large, may lead to aggregation of e.g., rural and urban areas with very different infrastructure levels. This introduces again a large amount of systematic variance. Furthermore, aggregation clearly reduces model resolution. This is may problematic as many current planning questions (e.g. road pricing) require a certain minimal model resolution. Thus, aggregation is strictly limited.

Concluding, this means that relying on a single simulation run is at least questionable. In any case, be it that the results are produced by a single run or multiple runs, sampling issues need to be analyzed for different spatial and temporal resolutions.

A few papers exist, where sampling issues have been investigated in the context of microsimulations such as Ziems et al. (2011); Cools et al. (2011); Veldhuisen et al. (2000); Castiglione et al. (2003); Hackney (2009). The investigations focus on the required number of microsimulation runs to reach stable results to reach a given level of confidence. Thereby only the random seeds are mutated whereas the inputs are held fixed. Castiglione et al. (2003) concludes his investigations: “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. Performing such analyses for MATSim is the first goal of this paper. This continues the work of Hackney (page 128ff [2009] who has investigated small ensembles of MATSim runs, where unobserved heterogeneity was not yet included in MATSim.

The conclusion of the papers cited above is that sampling error is essentially a non-issue for the simulators in question and the investigated temporal and spatial resolution levels. I.e., only a relatively small number of simulation runs (i.e., samples) is required to reach results given a certain practically useful confidence level. This means that the simulation results only feature little variability (as derived from Equation 1). A follow-up analysis must thus compare the variability in the microsimulation with the variability measured in the transport system. Getting some first insights regarding this, is the second goal of this paper.

### 1.2 Sources of Variability in Microsimulations

Travel demand is characterized by substantial intra- and interpersonal variability (see e.g., Buliung et al. [2008], Kitamura et al. [2006], Susilo and Kitamura [2005], Schlich [2001], Schlich and Axhausen [2003], Pas and Koppelman [1987], Pas [1988], Jones and Clarke [1988], Hanson and Huff [1988a,b], Huff and Hanson [1986], Hanson and Huff [1982], Burnett [1977], Golledge [1970]). This variability is due to changing endogenous motivations of the individuals (such as needs, preferences or personal experience) and due to a varying environment, including the feedback behavior of the other transport system participants.

For modeling purposes variability is often divided into a systematic and a random part (see e.g., Vovsha et al. [2002]). The microsimulation model variability can thus be categorized as follows:

- **Systematic observed variability**: It has rational sources in reality which are identified by the model at hand (observed). It is included either as model input or directly in the model formulation or both.

- **Random variability**:
  - inherently random (i.e., unobservable) variability: A certain share of the individuals’ decisions are made on the basis of chance.
  - unobserved, but actually systematic variability: The decision is actually made on a rational basis. But the modeler misses knowledge about the idiosyncratic rationales of the decision makers.
  - true random model variability: This variance is caused by algorithmic difficulties. For large-scale systems finding the global optimum is not trivial. One might get stuck local optima dependent on the random seeds chosen. This variability is present even if no other sources of random exist.
While true random model variability reduces the model quality in general and should be mini-
mized, all the other variabilities mentioned above are necessary parts of the model. The random
parts of the variability are included in the model by means of random sampling mechanisms. As
sampling is the constitutive process of microsimulations [Wolf (2001a,b)], they are a productive
tool in this context, and they are increasingly applied in transport modeling.

Before the variability included in MATSim is analyzed in detail in Section 1.4, a short introduc-
tion of the simulation framework is given.

1.3 MATSim—In Brief

The Multi-Agent Transport Simulation Toolkit, MATSim [MATSim-T, 2011] is an activity-
based, easily extendable, open source, multi-agent simulation toolkit implemented in JAVA
and designed for large-scale scenarios. A good overview of MATSim is given in Balmer
et al. (2006).

MATSim is a co-evolutionary model. While being in a competition for space-time slots on the
transportation infrastructure or in destination facilities with all the other agents, every agent
iteratively optimizes its daily activity chain by trial and error. Every agent therefore possesses a
memory of a fixed number of day plans, where each plan is composed of a daily activity chain
and an associated utility value (in MATSim called plan score).

In every iteration prior to the simulation of the network loading (e.g., Cetin, 2005), every agent
selects a plan from its memory. This selection is dependent on the plan utility. In many MATSim
investigations, a model that generates a logit model is used. Before the plans are executed on the
infrastructure a certain share of the agents (usually 10 %) is allowed to clone the selected plan
and to subsequently modify this cloned plan. Thereby three choice dimensions are included for
this paper: time choices [Balmer et al., 2005], route choices [Lefebvre and Balmer, 2007], and
destination choices.

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 memory of these agents.

An iteration is completed by evaluating the agent’s day described by the selected day plans.
The computation of the plan score is compatible with micro-economic foundations. The basic
MATSim utility function was formulated in [Charypar and Nagel, 2005] which is derived from
the Vickrey model for road congestion as described in Vickrey (1969) and Arnott et al. (1993).
The utility of a plan (described in detail in Charypar and Nagel, 2005) is computed as the sum
of all activity utilities plus the sum of all travel (dis)utilities.
1.4 Sources of Variability in MATSim

In the MATSim scenarios applied so far, work and home locations are fixed as they are given with high geographic resolution by the Swiss Census of Population 2000 (Swiss Federal Statistical Office 2000). Variability in terms of model input is introduced by variety of the day chain structures and the desired activity durations. These data are derived from a PUS, namely from the National Travel Survey for the years 2000 and 2005 (Swiss Federal Statistical Office 2006). Model variability is introduced by the time, route and destination choice modules. For the investigations of this paper car traffic only is modeled. Person attributes other than the home and work locations, such as household type or income are not included in the MATSim replanning modules in a robust manner. I.e., no variability is introduced by such kind of person attributes yet.

One goal of this paper is to investigate the amount of variability generated by the modules while the model input is held constant. However, this is only the starting point for a more in-depth future analysis, which also must include the explicit modeling of temporal variability as described in the next section.

1.4.1 Temporal Variability and Microsimulations Applied as Cross-Sectional Methods

While interpersonal variety in terms of travel behavior is researched to a great extent, intrapersonal or temporal variety has long been neglected in transport literature. The assumption was that individual travel behavior is dominated by repetition. This lack of research stems from high costs for longitudinal studies but also from methodological difficulties in assessing the temporal variability. The assessment of variety is intrinsically tied to the measures chosen, where the measurement resolution is particularly important. Clearly, a higher measurement resolution in general generates larger variability. This makes the assessment of variability methodologically demanding.

Not being paralyzed by this lack of research, people have started building models that are essentially based on cross-sectional data with the “typical (work) day” as the unit of analysis. In these models clearly averages are built over multiple persons. In other words, intrapersonal variability diffuses into interpersonal variability (which is incorrect) and also white noise (unobserved heterogeneity). Many microsimulations and also MATSim belong to this group of models.

Recent research has shown that the intrapersonal, i.e. temporal, variability is large. This renders cross-sectional models problematic as large aggregates are necessary to reach stable results.
Furthermore, the spatial (i.e., interpersonal) and temporal dimension are substantially correlated (known as autocorrelation) (see Figure 1), which, of course, influence the total variability of the results. In mathematical terms this reads. Given two random variables $X_0$ and $X_1$ representing an arbitrary time-dependent decision of individual 0 and 1, i.e., $X_0 = f_0(t)$ and $X_1 = f_1(t)$, the variance of two random variables is $\text{Var}(X_0, X_1) = \text{Var}(X_0) + \text{Var}(X_1) + 2\text{Cov}(X_0, X_1)$. The covariance is non-zero for correlated variables, where the covariance is greater than zero if variables are equidirectional. There are a lot of transport-related decisions for which individuals tend to have a positive correlation, i.e., $\text{Cov}(X_0, X_1) > 0$. This is given by global rhythms of life. There are also decisions for which correlation is negative i.e., $\text{Cov}(X_0, X_1) < 0$. An example might is the avoidance of demand peaks, such as not visiting certain skiing resorts during school holidays. If the positive or negative correlation part dominates is a question for future work.

1.5 Discussion

Small Numbers of Simulation Runs:
As mentioned earlier previous work comes to the conclusion that the sampling error is negligible for the respective simulators and the levels of temporal and spatial resolution analyzed. In consequence, the simulation results only feature little variability. To be more specific, the results are subject to only little unobserved heterogeneity. However, if in reality a large variability is measured and if the unobserved part of this "true" variability is substantial, then small numbers of required simulation runs means that unobserved heterogeneity is not captured adequately by these models. I.e., then, small numbers of required simulation runs are essentially a weakness of the model. On the other side, if the variability is observed, then it can be introduced to the model as input, for example in the form of different day plans. In this case low required numbers of simulation runs are desirable.

Temporal Variability and Microsimulations Applied as Cross-Sectional Methods:
In the opinion of the authors the strength of microsimulations is that not only averages can be captured but also the fluctuations (see also Esser and Nagel (page 704 2001) or Newman and Barkema (page 11 1999). However, as mentioned above, if microsimulations are used as cross-sectional methods, intrapersonal (i.e, temporal) variability diffuses into interpersonal variability and also white noise (unobserved heterogeneity). This transformation is random in nature. This makes it impossible to capture the temporal variability consistently by the random error terms. The results might show the same amount of variability as measured in reality but its effects on the network might are completely different. Thus, the presence of temporal variability (and even more so the temporal correlations) actually means that results for the
average working day requires to simulate multiple days. However, MATSim is not yet there. Thus, it is investigated here how much variability is produced by simulating one single day as done up to now.

2 Method

The sampling error is examined by using the 10% Zurich simulation scenario as described in Section 2.1. For illustration purposes a small-scale scenario as described in Section 2.2 is used in addition. The (aggregate) temporal variability in the transport system is assessed and compared to simulation results by using annual count data. The investigations done in this paper as well as future investigations profit from establishing and systematically applying a microsimulation sampling terminology. First steps in this regard are accomplished in Section 2.3.

2.1 Real-world Scenario: 10% Zurich Scenario

The 10%-Zurich scenario is frequently used for the further development of MATSim but also for projects in Swiss planning practice (e.g., Balmer et al. 2009, Horni et al. 2009). The demand of the simulation scenario is derived from the Swiss Census of Population 2000 (Swiss Federal Statistical Office 2000) and the National Travel Survey for the years 2000 and 2005 (Swiss Federal Statistical Office 2006). A 10% sample of the car traffic (including cross-border traffic) that crosses the area delineated by a 30 km circle around the center of Zurich (Bellevue) is drawn, which results in almost 68’000 agents simulated. The activity location data set, comprising more than 10^6 home, work, education, shopping and leisure locations, is computed from the the Swiss Census of Population 2000 and the Federal Enterprise Census 2001 (Swiss Federal Statistical Office 2001). The network from the Swiss National Transport Model (Vrtic et al. (2003)) is used, which consists of 60’492 directed links and 24’180 nodes. An average week day is simulated, where the number of trips per agent averages at 3.35. In total 25’896 shopping activities and 40’971 leisure activities are performed. Comparable data is available in most countries from official sources, such as censuses, national travel diary studies and commercial sources, such as navigation network providers, yellow pages publishers or business directories.

The choice setting comprises the three dimensions, time, route and destination choice for discretionary activities. The random seeds of all choice dimensions are varied simultaneously in this work.
2.2 Small-Scale MATSim Simulation Scenario

To be able to simulate a very large number of simulation runs for illustration purposes a small-scale synthetic MATSim scenario is used. The configuration of the scenario is depicted in Figure 5. The influence of temporal correlations is illustrated with this scenario.

1’000 persons, living in the two residence zones (location 1 and 2) perform one shopping activity with a desired duration of 90 minutes per day. The shopping activity can be performed in the home zones (location 1 and 2) or in the city zone (locations 5-9). 5 consecutive working days are simulated. A variable share of persons does a working activity in the city zone (at location 3) with shares of 0.9, 0.9, 0.9, 0.8, 0.5 (from Monday to Friday). On Fridays all workers have a desired working activity duration of only 7 instead of 9 hours. Time, route and destination choices are performed.

2.3 Sampling Terminology for Transport Microsimulations

Transport microsimulations are applied in activity-based, i.e., disaggregate contexts. Units of simulation, able to make decisions, are the persons or small homogeneous groups of persons. Prominent decision dimensions are time, route, mode and destination choice, where more recently also activity chain choice is researched. This spans a huge range of possible combinations of decisions. The number of combinations is essentially infinite as time choice is continuous.

A single microsimulation run represents one of these combinations. In sampling terminology, this is the sampling unit. Variables of interest may are the total vehicle miles traveled in the study area or the total daily load on a specific link. One simulation run provides one realization of these variables. The sample contains \(n\) simulation runs. The population parameters (such as the mean) of these variables are estimated by a population statistic. This statistic is accompanied by the standard error, also called sampling error. The sampling error is the standard deviation of the sampling distribution. For the case where the statistic is the mean the sampling error \(\epsilon_s\) can be estimated by

\[
\epsilon_s \sim \frac{\hat{\sigma}}{\sqrt{n}} \tag{1}
\]

where \(\hat{\sigma}\) is the statistic for the standard deviation and \(n\) is the sample size. The derivation of this formula is for example given in [Hutchinson (1993)].
3 Results

In the following sections results of preliminary experiments and analyses are presented. Comprehensive checks are planned for future work.

3.1 Traffic Count Data

As mentioned earlier the assessment of temporal variability of the transport system is non-trivial. In this paper we use traffic count data. Clearly, in general model assessment must not only base on count data, as also count data for a specific time features a substantial amount of random noise. I.e., random influences such as the weather naturally have a significant effect on counted values. However, the large amount of data provides a large number of sampling points. The data is prepared as described in detail in Balmer et al. (e.g., 2009). Additionally, the days between Christmas and new year are filtered out. Using Swiss count data with hourly resolution the volumes of an average working day (Tuesday to Thursday) are calculated. ... respectively ... unidirectional links are measured for Switzerland and the Zurich region respectively. Only count values greater than zero are included.

In Figure 1 the analysis of counted link volumes is given for different spatial and temporal domains, i.e., complete Switzerland and the region of Zurich (area with 12 km radius around the Bellevue in Zurich) is evaluated for the complete year and for single months. The (population) standard deviation is given with respect to the mean for the respective temporal and spatial domain. E.g., monthly volumes are compared to the monthly average and not to the annual average. A substantial variability can be observed, where the annual variability is always larger than the variability of single months. This indicates substantial temporal autocorrelations.

The variability of the Swiss volumes are greater than the values for the Zurich region. Different regions of Switzerland show different annual rhythms. E.g., skiing resorts have high loads in winter while Alpine passes are closed then. This can be interpreted as spatial autocorrelation combined with temporal autocorrelation.

3.2 Small-Scale Scenario

Two different configurations are simulated. For configuration 2 the averaging is done on the input side. E.g., an average share of workers is simulated, where for every day 20 % of the workers has a reduced desired working duration of 7 instead of 9 hours. In configuration 1 temporal correlations are explicitly taken into account by applying the shares of working activities and desired work durations per day. The hourly arrivals and departures plotted in Figure 2 and 3.
respectively clearly illustrate that both configurations generate similar averages but different variability. Configuration 1 shows substantially larger variability as discussed in Section 1.4.1.

3.3 Zurich Scenario

A set of 10 simulation runs of the Zurich scenario are performed with identical input but varying random seeds. This corresponds to the Method of replication as described in Benekohal and Abu-Lebdeh (1994). At the moment a 10\% random draw of the complete population is simulated. Work started recently will look at different sampling rates (including oversampling).

3.3.1 Analyses

**Population Level:** At population level only little variability between simulation results exists. The mean score (averaged over agents) of all executed plans of the final iteration 100 averages at 180.871 with a sample standard deviation of 0.230 (i.e., 0.127\% of the mean).

**Link Level:** In Figure 4 the sample standard deviations of the link volumes for three different hours of the day are shown. When looking at a single iteration (here the final iteration 100) (Figure 4(a)) one can see substantial variability. However, when looking at the link volumes which are averaged over 5 iterations (here 96-100)(Figure 4(b)) this variability is at least halved. This means that results are subject to a relatively large intra-run variability. This can either mean that the system is far from a relaxed state or that the replanning modules add too much variability per iteration in general. In this work the replanning shares are quite high (20\% of the agents perform destination and route choice with time adaptations accordingly and additionally 20\% of the population performs time choice), which potentially introduces a large intra-run variability. In any case much intra-run variability is undesirable and makes further model improvements necessary. In this vein, this paper also fulfills first steps of model verification.

**Person Level:** In Figure 6 the sample standard deviations of the plans executed in iteration 100 are shown. The vast majority of agents shows a relatively small variability.

4 Conclusions and Outlook

This paper undertakes first steps in evaluating the variability of MATSim results. This is especially important now as recently mechanisms to explicitly treat unobserved heterogeneity
Figure 1: Standard Deviation [%] of Swiss Annual Count Data.

(a) 07:00-08:00

(b) 11:00-12:00

(c) 17:00-18:00
Figure 2: Arrivals for Configuration 1 and 2 (Iteration 100)

(a) Configuration 1

(b) Configuration 2

were added to the MATSim destination choice utility function. The results are preliminary and need comprehensive verification and elaboration.

An evaluation of annual Swiss count data shows substantial temporal and spatial variability. Microsimulation procedures must be developed to efficiently model not just the average statistics but also this variability.

The MATSim results also show substantial variability. A large intra-run variability is observed. As the development of the destination choice module is in an early stage the sources of the variability are not yet fully identified. It might be that some of the variety is introduced by the replanning configuration (e.g., large replanning share), which is of course undesirable.

Recommendations on required numbers of random runs for practice will be developed, when the verification of the destination choice module is finished. Future work will include the
investigation of different sampling rates (including oversampling of the population). Data for finer analyses will be available by a recent GPS survey conducted at the institute at the moment. In the far future mode choice and activity chain choice will be included.
Figure 4: Standard Deviation of Simulated Link Volumes

(a) Iteration 100

(b) Averaged Iteration 96-100
Figure 5: Small-scale Scenario

Legend:
- Network node
- Network link with id
- Work locations
- Shop locations
- Home & shop locations

Figure 6: Standard Deviations of Executed Plans Scores
5 References


Variability in Transport Microsimulations Investigated for MATSim: Preliminary Results

April 2011


