Output variability caused by random seeds in a multi-agent transport simulation model

Paulsen, Mads; Rasmussen, Thomas Kjær; Nielsen, Otto Anker

Published in:
Procedia Computer Science

Link to article, DOI:
10.1016/j.procs.2018.04.078

Publication date:
2018

Document Version
Publisher’s PDF, also known as Version of record

Link back to DTU Orbit

Citation (APA):
Output variability caused by random seeds in a multi-agent transport simulation model

Mads Paulsen\textsuperscript{a,}\textsuperscript{*}, Thomas Kjær Rasmussen\textsuperscript{a}, Otto Anker Nielsen\textsuperscript{a}

\textsuperscript{a}DTU Management Engineering, Technical University of Denmark, Bygningstorvet 116B, 2800 Kgs. Lyngby, Denmark

Abstract

Dynamic transport simulators are intended to support decision makers in transport-related issues, and as such it is valuable that the random variability of their outputs is as small as possible. In this study we analyse the output variability caused by random seeds of a multi-agent transport simulation model (MATSim) when applied to a case study of Santiago de Chile. Results based on 100 different random seeds shows that the relative accuracies of estimated link loads tend to increase with link load, but that relative errors of up to 10\% do occur even for links with large volumes. Although the proportion of links having large relative errors is roughly the same for all of the investigated seeds, it is shown the variations of individual link loads between seeds largely dominate variations between the two last iterations within a seed.

© 2018 The Authors. Published by Elsevier B.V.
Peer-review under responsibility of the Conference Program Chairs.

Keywords: Transport Simulation, Stochasticity, Simulation Variability, MATSim

1. Introduction

Uncertainty in transport models is a well-known problem that needs to be addressed when using their outputs\textsuperscript{1}. Whereas the uncertainty of traditional analytical transport models is caused by uncertainty of input parameters, models that rely on simulation have additional variability in their output due to the stochasticity of the series of pseudo-random numbers that are drawn. These series are determined by the so-called random seed.

Although some types of transport models use stochasticity/simulation for choice set generation, it is most widely used in dedicated transport simulators. The open-source software MATSim\textsuperscript{2} is one example of such large-scale transport simulator used for analysing transport scenarios all across the world. As transport simulators are generally intended to facilitate policy support, it is of great interest to examine the output variability caused by random seeds in MATSim in order to determine whether this can potentially overshadow the impacts of suggested infrastructure investments under consideration.

This paper investigates the variability in output caused by random seeds for an open-data scenario of Santiago de Chile\textsuperscript{3,4,5}. This is done by running the model repeatedly with different seeds using otherwise the same default

\textsuperscript{*}Corresponding author. Tel.: +45 45 25 65 98.
\textit{E-mail address}: mads@dtu.dk

1877-0509 © 2018 The Authors. Published by Elsevier B.V.
Peer-review under responsibility of the Conference Program Chairs.
10.1016/j.procs.2018.04.078
configurations and inputs. In the remainder of the paper, some additional background is initially given in section 2, followed by an introduction to the applied methodology in section 3. Section 4 presents how random numbers are generated in MATSim, while the case study and results are presented in sections 5 and 6, respectively. Section 7 discusses the findings and gives suggestions for further work.

2. Background

While a lot of studies have dealt with uncertainties in transport models in general — in particular uncertainties due to uncertainties of input — only a few studies have previously focused on the pure stochastic variability caused by the selection of random seeds in activity-based models.

Veldhuisen et al. concluded that Monte Carlo errors are negligible when considering aggregate results of the microsimulation model RAMBLAS.

Castiglione et al. used an approach where the partial mean of the first \( n \) seeds is compared to the final mean \( (n = 100) \) for the so-called San Francisco Model. Although this is also interesting, the partial mean is dependent on the order in which the seeds are used. Also it does not generally answer how large errors are when using a single seed, as only one of such cases is investigated (the first partial mean).

Lawe et al. investigated the sensitivity to variations in random seed based on five different random seeds in TRANSIMS. Although based on a small sample, the coefficient of variation of link loads were small on average for all investigated links.

Also with TRANSIMS, Ziemssen et al. follows the same approach as in Castiglione et al. using 20 seeds, but extends the analysis by changing and keeping the random seeds of different parts of the simulation. Coefficients of variations of link loads for two investigated links were 1.7% when changing the random seed of the entire model.

Cools et al. investigated the variability of demand for each mode when using FEATHERS. Based on 200 different seeds the coefficient of variation was generally below 1%.

Bekhor et al. investigates the variability that arises from random components of the model specifications of the Tel Aviv Transportation Model. It also studies variability solely caused by random seeds by running various settings with three different seeds. The number of different seeds used is too low to quantify the findings, though.

Several studies by Nagel et al. touches upon the issue of simulation errors caused by selection of random seeds. It is noted and visualised that such variability exists based on two different random seeds, but an in-depth investigation of the variation has not been published, although unpublished work by Raney et al. is mentioned.

One of the studies by Nagel et al. mentions the need to further investigate the variability caused by random seeds in MATSim specifically. This is supported by a chapter of the MATSim book, stating that quantities of interest from the output of the model can be found by averaging over a series of runs with different seeds. This is encouraged due to the broken ergodicity of the model, meaning that once a random seed is selected some till then possible outcomes of the model are no longer reachable.

3. Methodology

In order to evaluate the variations caused by random seeds in a multi-agent transport simulator, a number of measures will be applied in this study. The first is the coefficient of variation, \( c_v \), defined as the sample standard deviation divided by the sample mean, both considered across all seeds \( s \) in the set of seeds \( S \).

We also introduce two additional measures. With \( x^l_s \) denoting the link load of link \( l \) within the set of links \( L \) when using seed \( s \in S \), we introduce the following notation with Iverson brackets to denote the empirical probability of the link, \( l \in L \), having a link load that is more than \( q \) times off its sample mean across all seeds, \( \bar{x}_l \),

\[
\rho_q^l = \frac{\sum_{s \in S} \left[ \frac{|x^l_s - \bar{x}_l|}{\bar{x}_l} > q \right]}{|S|}, \quad \text{for } l \in L, \quad q \geq 0. \tag{1}
\]
Likewise, we denote the proportion of links having link loads further than ℚ times away from their corresponding empirical mean values when the seed, s ∈ S, was used by,

\[ r_q^s = \sum_{l \in L} \left| \frac{x_l^{s,l} - \hat{x}_l}{\bar{x}_l} > q \right| \frac{1}{|L|}, \quad s \in S, \quad q > 0. \tag{2} \]

By considering \( r_q^l \) and \( r_q^s \) across all \( l \in L \) and \( s \in S \), respectively, a series of values, \( r_q^l \) and \( r_q^s \), is obtained.

Furthermore, we introduce the terms within-seed variation, \( W_l \), and the between-seed variation, \( B_l \), for a link \( l \in L \),

\[
W_l = \frac{1}{|S|} \sum_{s \in S} \left( x_l^{s,l} - x_l^{s,l} \right)^2, \quad l \in L,
\]

\[
B_l = \frac{1}{|S| - 1} \sum_{s \in S} \left( x_l^{s,l} - x_l^{s,l} \right)^2, \quad l \in L.
\]

The calculation of the within-seed variation utilises that the (ex post) expected value of iteration \(|l| - 1\) is the value obtained in the last iteration, \(|l|\). For the between seed variation, we consider that each seed might converge to a different solution, why the expected value is found as the mean across all seeds \( s \in S \) in iteration \(|l|\), i.e. \( x_l^{s,l} = \frac{1}{|S|} \sum_{s \in S} x_l^{s,l} \). We can compare \( B_l \) and \( W_l \) by calculating \( \tilde{R}_l \),

\[
\tilde{R}_l = \sqrt{\frac{B_l + W_l}{W_l}}, \quad l \in L.
\]

If this is close to 1, then the between-seed variation is negligible, whereas the opposite is true for large values of \( \tilde{R}_l \).

4. MATSim: A Multi-Agent Transport Simulator

MATSim is an activity- and individual-based simulation model, where each agent aims at obtaining the highest possible value of their scoring function. It uses a co-evolutionary algorithm (Algorithm 1) to reach its final output.

**Algorithm 1** Co-evolutionary, population based search

1. **Initiation**: Generate at least one plan for every agent.

2. **Iterations**: Repeat the following many times.
   (a) **Plan Selection**: Select one plan for every agent.
   (b) **Plan Scoring**: Obtain a score for every agent’s selected plan by executing all selected plans simultaneously in a mobility simulator and attach a performance measure (score) to each executed plan.
   (c) **Plan Innovation**: Generate new plans for some of the agents by mutating existing plans or re-routing.

The algorithm utilises on random numbers on several occasions. Firstly, they are used to determine the order in which to handle in-going links of each node in every time step of the mobility simulator. Here any link is chosen with a probability proportional to its capacity.

Secondly, they are used to determine which agents use which plan mutation strategy and plan selection strategy in each iteration. In each iteration 30% of agents are selected to generate a new plan for their choice set – half of these by re-routing (new shortest path searches), the other half by substituting the mode of a discrete-uniformly randomly selected sub-tour by another mode from the pool of possible modes. If the generation of the plan forces the choice set to exceed the maximum number of plans, the plan with the worst score is removed from the choice set.

Thirdly, the 70% of agents who do not generate a new plan, draw a plan, \( P^* \), from their choice set with discrete uniform probability, and selects this plan with probability \( 1, 0.01 \cdot e^{\frac{U_P - U_{PC}}{2}} \), where \( U_P \) and \( U_{PC} \) are the latest
scores of the drawn plan, \( P^* \), the expected value is found as the mean across all seeds obtained in the last iteration, why the expected value is obtained.

The calculation of the within-seed variation utilises that the (ex post) expected value of iteration \( I \) and \( S \) is the sum of all likelihoods in the iteration, i.e. \( \sum_{l \in L} W_l \). We can compare, for instance, \( r \) and \( \hat{r} \), was used by, and the currently selected plan, \( P_C \), respectively. If \( P^* \) is not chosen, the agent will execute the current plan again in the next iteration.

Finally, in order to have a larger diversity on the routes suggested when re-routing, the utility of money is drawn from a log-normal distribution when initiating a router.\(^{17}\)

The random numbers are thus predominantly used for choosing between reasonable discrete alternatives. Therefore, MATSim does not at seem to be particularly vulnerable to variability arising from random seeds a first glance.

With the role of random numbers in MATSim established, we now turn to discuss how pseudo-random numbers are created in MATSim. They are to a large extent based on Java’s built-in Linear Congruential Generator (LCG). The traditional LCG\(^{18}\) is of the form,

\[
X_{n+1} = aX_n + c \mod m,
\]

where \( X_n \) is the latest draw, \( X_{n+1} \) is the next draw, and \( a, c \) and \( m \) are parameters having the values \( m = 2^{48}, a = 25,214,903,917 \), and \( c = 11 \) in Java. In order to get a uniformly distributed number \( X_{n+1} \) is divided by \( m \).

MATSim offers a few extensions to Java’s built-in random number generation. First of all, when a random number generator is instantiated, it immediately draws and throws away the first 100 draws, as these are found not to be sufficiently random.\(^{15}\) Secondly, whenever a new random number generator is created from an existing one, the value of an internal counter is multiplied by 23 and added to the latest used seed before discarding the next 100 draws of the newly created instance. The initial random number generator is established with the globally assigned random seed as \( X_0 \), meaning that this number completely determines the sequence of all other random numbers used in the run.

5. Case Study

The model has been run for 100 different initial global random seeds on the Santiago de Chile open data scenario\(^{3,4,5}\). The scenario includes a toll system on selected roads, and schedule-based public transport running on dedicated public transport links. The default “out-of-the-box” configurations set by the developers of the scenario have been used, including a population sample of 10% corresponding to 665,201 agents, a network consisting of 22,981 unidirectional links for car traffic, and 100 iterations per seed. Choice sets for every agent are locked after 80 iterations, at which point MSA is enabled for the (at most) five plans per agent. Considered modes for substitution of a sub-tour are walk, public transport, and – if available to the agent – car.

6. Results

The results show that coefficients of variation of link loads (dedicated public transport links excluded), \( c_{v_l} \), are generally quite low with 65.7 % of the links not exceeding 5% (see Figure 1). The distribution has a skewness of 16.7 and has some resemblance to a log-normal distribution, which is also shown in the figure.

The correlations between the mean (\( \bar{x}_l \)), the standard deviation (\( \hat{d}_v \)), and the coefficient of variation (\( c_{v_l} \)) across all links \( l \in L \) are shown in Table 1. It is seen that there is a strong correlation between \( \bar{x}_l \) and \( \hat{d}_v \) (0.849) indicating
that using \( c_{vl} \) as a measure seems valid. However, since the correlation between \((\bar{x}_l)i\) and \(c_{vl}^{\text{cv}}\) is negative (-0.113), \(c_{vl}^{\text{cv}}\) is generally expected to be a little smaller for links with high flows.

In Figure 2 it is seen that the coefficient of variation generally is about 1% for the 50 busiest links, but that some links have a significantly higher \(c_{vl}^{\text{cv}}\). The figure also shows that the ranges for individual links seem to be large, indicating that for some seeds the error can be notably higher than a few percent.

As mentioned in section 3, \( R_{q}^{l} \) holds the empirical probabilities of each link having a relative error larger than \( q \) when using a single seed. The corresponding cumulative distribution function is found in Figure 3 for various values of \( q \).

### Table 1. Sample Pearson correlation coefficient between the mean, the standard deviation, and the coefficient of variation of link loads.

<table>
<thead>
<tr>
<th></th>
<th>Mean ((\bar{x}_l)i)</th>
<th>Standard Deviation ((\hat{\sigma}^{vl}))</th>
<th>Coefficient of Variation (c_{vl}^{\text{cv}})</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean ((\bar{x}_l)i)</td>
<td>1</td>
<td>0.849</td>
<td>-0.113</td>
</tr>
<tr>
<td>Standard Deviation ((\hat{\sigma}^{vl}))</td>
<td>0.849</td>
<td>1</td>
<td>-0.115</td>
</tr>
<tr>
<td>Coefficient of Variation (c_{vl}^{\text{cv}})</td>
<td>-0.113</td>
<td>-0.115</td>
<td>1</td>
</tr>
</tbody>
</table>

Fig. 2. Boxplots of link loads \(\left\{x_{l}^{l}\right\}\) for the 50 busiest links and their corresponding coefficient of variation \(c_{vl}^{\text{cv}}\).

Fig. 3. Empirical cumulative distribution function of \(R_{q}^{l}\) for various values of \( q \).
indicating that for some seeds the error can be notably higher than a few percent. Standard Deviation (\(\hat{\sigma}\)) Coefficient of Variation (\(cv\)).

Table 1. Sample Pearson correlation coefficient of variation of link loads.

As mentioned in section 3, the correlation between the mean, the standard deviation, and the coefficient of variation generally is about 1 % for the 50 busiest links, but that some of the links had a relative error exceeding 5 % in at least half of the runs. Figure 4, however, shows that the extreme errors are primarily seen for less busy links. An interesting phenomenon occurs for links with a link load between 1,000 and 10,000, though, as the probabilities of 15 %, 10 %, and especially 5 % and 2.5 % errors seem to be larger than for the links with a mean link load just below 1,000.

In the above we have found that at the link level, there is a risk of getting relatively high deviations when only using a single seed value. In Figure 5 we consider each seed as a whole across all links, and find the corresponding proportion of links having a relative error larger than \(q\), \(r^S_q\). It is seen that the seeds generally perform equally well, but that about 24 % of links are expected to have a link load deviating at least 5 % from their empirical mean.

So far we have established that between-seed variation is in fact considerable, however, without comparing it to other types of known variations. This is done by calculating \(\hat{R}_l\) as described in section 3. The empirical cumulative distribution function of \(\hat{R}_l\) across all \(l \in L\) (Figure 6) shows that between seed variation greatly dominates the within seed variation for almost every link. This trend is even more evident for larger links, as \(\hat{R}_l\) is at least 4 for 90.2 % of links with a link load above 10 (see Figure 7).

Variability of mode markets shares has also been investigated, but was found to be negligible with the largest coefficient of variation being \(8 \cdot 10^{-4}\). Variability of travel speeds, on the other hand, showed a pattern that was nearly indistinguishable from that of link loads.
The results of this study indicate that the potential relative error when using a single seed is generally a decreasing function of the amount of traffic on the link. However, among the largest links there seems to be a few links that are particularly error-prone. This should give rise to some concern as in many cases such busy links are the most interesting for policy support purposes. A geographical analysis of the links with a high coefficient of variation would therefore be a natural extension of the current work.

The finding that the between-seed variations largely dominate the within-seed variations suggests that broken ergodicity might in fact be a problem that needs to be addressed in transport simulators dealing with discrete choices of agents. Especially, since the tests in this paper were conducted using a low number of iterations (it is not uncommon for MATSim applications to use as much as 1,000 iterations) and thus were expected to have a relatively high level of within-seed variation.

One way to overcome this problem could be to average the results across multiple runs with different seeds, however at the theoretical risk that the resulting (averaged) solution is not an equilibrium solution. Alternatively, it may be favourable to present the results as a distribution by running the model multiple times with different seeds as proposed in the literature\(^{15}\). Especially since the results of this paper suggest that the information gained by such additional runs exceeds that of running supplementary iterations of a single seed.

Representing results as distributions instead of point estimates would be of particular interest for project appraisal purposes. Here the additional information would contribute to a further understanding of the uncertainty of the responses to a proposed project. Potentially this could answer whether the effects of infrastructural changes risk being overshadowed by between-seed variations. Running tests under such project appraisal-like circumstances with a base network and an alternative network is thus recommended for future work.

Another method previously proposed in the literature\(^{19,20}\) aims at producing a more stable model outcome in the first place. This is, however, beyond the scope of this paper.
8. Acknowledgements

A special thanks to the developers of the Santiago de Chile scenario Benjamin Kickhöfer and Alejandro Tarachini, as well as the creators of MATSim, Kai Nagel and Kay Axhausen. Without the open access of the scenario and model this study would not have been possible.

References