Disaggregation of static OD-matrices for dynamic MATSim simulations

2025-04-15

Franz-Xaver Rupprecht [1], Gunnar Flötteröd [2], Yusak Susilo [1]

(1) Department of Landscape, Spatial and Infrastructure Sciences, Institute for Transport Studies (IVe), BOKU University, Vienna, Austria. Corresponding author: franz.rupprecht@boku.ac.at
(2) Department of Communication and Transport Systems, Linköping University, Linköping, Sweden & The Swedish National Road and Transport Research Institute, Stockholm, Sweden.

Other than classic four-step models (FSM), MATSim can simulate human behavior at the individual-level (Horni et al., 2016). This type of simulation is based on individual-level travel demand data that is typically derived from upstream activity-based models. However, these models have been observed to be expensive and challenging to implement (Rasouli & Timmermans, 2014).

The goal of this work is to provide a tool that makes travel demand data from existing FSMs (Origin-Destination (OD) matrices) accessible for MATSim simulations. Over the past decades, many cities have invested significant resources in implementing FSMs. Hence, the ability to use this data for agent-based transport models will help to increase their acceptance.

The challenge is to disaggregate the trips captured in OD-matrices into distinguishable and home-anchored trip chains that can later be assigned to the agents of a given synthetic population. We call such trip chains round trips and use the Metropolis-Hasting's algorithm (MHA) to sample a list of many such round trips at each iteration of the algorithm (Hastings, 1970). In this case a round trip consists of a sequence of departure locations (with the first one being the home location) and a sequence of departure times.

To evaluate how well a sample (list of round trips) reconstructs a given OD-matrix (target OD-matrix), we enter all trips of this list into a sample OD-matrix, which is compared to the target OD-matrix. By calculating the deviation $E(x)$ between the two matrices, the likelihood $b(x)$ of the sample x is determined:

$$b(x) = e^{E(x)}$$

The larger the deviation between a sample and the target, the smaller its likelihood will be. Since the probability of the MHA to accept a sample as a valid next step is proportional to the likelihood of that sample, we see that over the course of many thousands of iterations, the algorithm delivers samples that perfectly reconstruct the target OD-matrix. A detailed specification of the sampling machinery is given in Flötteröd (2025).

As shown in last year's presentation, we applied the approach to the city of Vienna, using the static OD-matrix from the PTV Visum model of VOR as the target OD-matrix (PTV Planung Transport Verkehr GmbH, 2022; Verkehrsverbund Ost-Region (VOR) GmbH, 2024). The scenario has 250 zones, an hourly temporal resolution, at each iteration a list of $100{,}000$ round trips is sampled, and up to this point only OD-matrices containing car trips are considered.

In addition to the likelihood term for OD reproduction, we added four more terms of the same mathematical structure, each of which reproducing an additional dataset. This results in the following likelihood function:

$$b(x) = e^{\,E_{OD}(x) + E_{HL}(x) + E_{WL}(x) + E_{HA}(x) + E_{WA}(x)}$$

As described above $E_{OD}(x)$ represents the deviation between the sample and the target OD-matrix, $E_{HL}(x)$ represents the deviation between the spatial distribution of home locations in the sample and the target data, and $E_{WL}(x)$ doing the same for work locations. As with the OD data, the location data was taken for the PTV Visum model of VOR. $E_{HA}(x)$ and $E_{WA}(x)$ ensure the correct reproduction of the start time and duration of home and work activities (longest out-of-home activity). The target data for activity start times and durations was extracted from the Austrian national household travel survey Österreich Unterwegs (Bundesministerium für Verkehr, Innovation und Technologie, 2016).

Figure 1 shows the high reproduction quality of the OD-matrix, the spatial distribution of home locations, and the spatial distribution of work locations. Figure 2 shows the plausible time structure of the travelling, which results from the correct reproduction of activity times and durations.

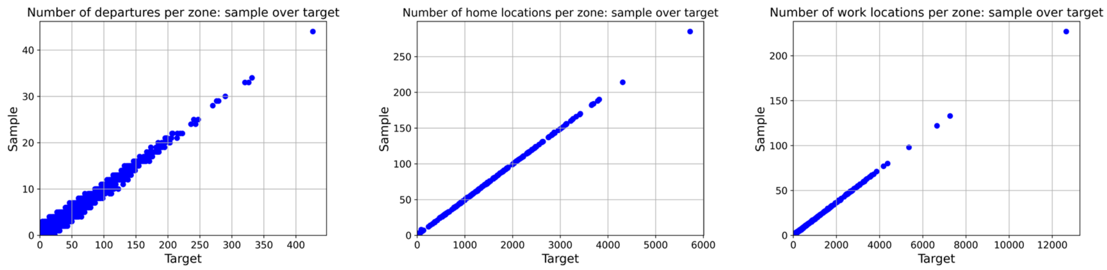

Figure 1: Number of trips per OD relation in the sample vs. in the target (left); Number of home locations per zone in the sample vs. in the target (center); Number of work locations per zone in the sample vs. in the target (right)

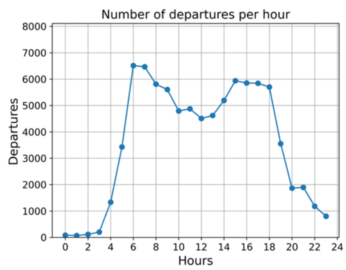

Figure 2: Number of departures per hour

The results prove that the method is suitable for disaggregating travel demand data. The reproduction quality is adjustable and more than satisfactory, even when reproducing multiple datasets at the same time. Unlike population synthesis pipelines (Hörl & Balac, 2021), the Metropolis-Hasting's based synthesis of round trips incorporates all datasets in one model and solves it simultaneously. The advantage of this is threefold: first, it improves the traceability of assumptions, second, it reduces error propagation, and third, it clearly decouples the sampling process from the specification of the model.

Until now, the main limitation of the approach has been its computational cost. The results presented here involve 10 million iterations of the algorithm and still do not indicate full stationarity in all the statistics analyzed. The runtime of this experiment was about 9 days on a 3.0 GHz AMD EPYC 7302 CPU single-threaded. Ongoing work hence focuses on speeding up the algorithm.

Compared to the state of the work presented at last year's MATSim user meeting, we have increased the spatial resolution by a factor of ten and replaced assumptions on the time structure of travel with real data. Apart from that, ongoing work focuses on extending the method to all modes of transport included in the upstream FSM (walking, cycling, public transport, car) and creating a MATSim-compatible plans file. We are optimistic that by the time of the conference we will be able to present a comparison of the PTV VISUM model and its extension to MATSim in terms of network flows, modal split and person-level travel times.

References

Bundesministerium für Verkehr, Innovation und Technologie. (2016). *Österreich unterwegs 2013/2014*.

Flötteröd, G. (2025). *An operational alternative to origin/destination matrices*. 13th Symposium of the European Association for Research in Transportation, Munich, Germany, 2025.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, *57*(1), 97–109. https://doi.org/10.1093/biomet/57.1.97

Hörl, S., & Balac, M. (2021). Synthetic population and travel demand for Paris and Île-de-France based on open and publicly available data. *Transportation Research Part C: Emerging Technologies*, *130*, 103291. https://doi.org/10.1016/j.trc.2021.103291

Horni, A., Nagel, K., & Axhausen, K. W. (2016). *The Multi-Agent Transport Simulation MATSim—Book*. Ubiquity Press. https://doi.org/10.5334/baw

PTV Planung Transport Verkehr GmbH. (2022). *PTV Visum 2022—Manual*.

Rasouli, S., & Timmermans, H. (2014). Activity-based models of travel demand: Promises, progress and prospects. *International Journal of Urban Sciences*, *18*(1), 31–60. https://doi.org/10.1080/12265934.2013.835118

Verkehrsverbund Ost-Region (VOR) GmbH. (2024). *Verkehrsverbund Ost-Region*. Verkehrsverbund Ost-Region. https://www.vor.at/en/