

23rd EURO Working Group on Transportation Meeting, EWGT 2020, 16-18 September 2020,
Paphos, Cyprus

An efficient approach to create agent-based transport simulation scenarios based on ubiquitous Big Data and a new, aspatial activity-scheduling model

Dominik Ziemke^{a,b,*}, Billy Charlton^a, Sebastian Hörl^c, Kai Nagel^a

^aTechnische Universität Berlin, Transport Systems Planning and Transport Telematics, Salzufer 17-19, 10587 Berlin, Germany

^bHasselt University, IMOB (Transportation Research Institute), Wetenschapspark 5 bus 6, 3590 Diepenbeek, Belgium

^cETH Zürich, Institute for Traffic Planning and Systems, Stefano-Franscini-Platz 5, 8094 Zürich, Switzerland

Abstract

Agent-based transport simulation models are a particularly useful tool to analyze demand-oriented transport policies and new mobility services, which have both gained significant attention lately. Since travel diaries, a traditional source to create the transport demand in agent-based transport models, are often hard to procure and not policy-sensitive, alternative approaches to creating travel demand representations for simulation scenarios are sought. In this study, a particularly efficient approach based on Big Data and a new, aspatial activity-based demand model with comparatively low input data requirements is established. Home, work, and education locations are informed based on mobile-phone-based origin-destination matrices. Other activity locations are modeled within the scope of the coevolutionary algorithm of the agent-based transport model, which is also responsible for finding suitable travel options of the modeled individuals. As a result, a comparatively lightweight process chain to create an agent-based transport simulation scenario is established, which is transferable to other regions. A basic quality evaluation of the created tool chain is carried out against a well-validated transport simulation model of the same region.

© 2020 The Authors. Published by Elsevier B.V.

Peer-review under responsibility of the scientific committee of the 23rd EURO Working Group on Transportation Meeting..

Keywords: Agent-based transport simulation; transport model; big data; cell-phone data; activity-based demand models

1. Introduction

MATSim (Horni et al., 2016) is an agent-based transport simulation framework. The individual traveler (*agent*) is the central unit of analysis and maintained during the whole modeling process. Each agent has one or more daily activity-travel patterns that they try to follow and, in doing so, compete for space-time slots with all other agents on the transport infrastructure. All agents try to optimize their individual travel behavior by applying modifications along

* Corresponding author. Tel.: +49-30-314-21383 ; fax: +49-30-314-26269.

E-mail address: dziemke@vsp.tu-berlin.de

different possible choice dimension (e.g. changing routes, changing modes of transport), trying out these modified plans in the simulation, and evaluating them based on the concept of utility maximization. Since all agents undergo this process simultaneously, this approach constitutes a co-evolutionary algorithm, which eventually leads to a stochastic user equilibrium.

As agents with all relevant properties, full knowledge of their intentions, and a memory of their previous experience are maintained during the whole process, dependencies and constraints between different trips or trips and activities (e.g. timing constraints, availability of transport modes based on previous travel and status thereof, e.g. battery load) are inherently resolved. This makes the simulation framework particularly interested for the analysis of novel, often more demand-oriented transport policies like user-stratified restrictions (e.g. environmental zones with distinction by vehicle emission class), tolls or user pricing (e.g. [Kaddoura, 2015](#)) or the analysis of novel transport services (e.g. MaaS or autonomous taxis) and technologies (e.g. electric vehicles) (e.g. [Bischoff and Maciejewski, 2014](#)).

As pointed out, MATSim's iterative algorithm – by which agents optimize their behavior in reaction changed conditions of the simulated reality – is used to analyze the effects of transport policies. Different choices that agents can take to adjust their behavior are set by the modeler in the form of *innovative strategy modules*. While such an innovative strategy can be simple (e.g. random alteration along a choice dimension), the utility-based evaluation (*scoring*) that operates on the simulated results of the agents' choices (simulated travel) ensures behavioral soundness.

This process of travel behavior adaption may not only be applied to analyze policies, but also in the process of the initial creation of a valid agent-based transport model for a given region. While MATSim is not a travel demand generation model in the sense of creating travel schedules based on demographic/socio-economic data, it should be regarded a *demand-adaptation model* whose abilities reach to some extent into the domain of typical activity-based demand generation model (or activity-scheduling model) (like CEMDAP ([Bhat et al., 2004](#)) or FEATHERS ([Bellemans et al., 2010](#))).

Properties of the transport demand whose corresponding decisions can be altered by an appropriate innovative strategy module during the simulation do not have to be provided endogenously as input. While in traditional route assignment models, all travel decision except route choice needs to be modeled correctly before applying the route assignment model, in MATSim additional choices like departure times or locations of discretionary activities can be modeled endogenously. Conversely, those travel demand choices that are not covered within the MATSim loop have to be provided endogenously ([Horni et al., 2016](#), p.19). In the current development stage, activity sequences and preferred durations of activities must always be provided as input as no readily applicable corresponding modules exist in MATSim.

A classic data source to create daily activity-travel plans is a travel survey. Based on a set of trip diaries of the study region, the plans of all members of the synthetic population (as representatives of the real-world population) can be created. With the above distinction between component of travel decisions that are exogenous to MATSim vs. those which can be modeled endogenously, trip diaries actually contain information more than needed. As least all transport-related properties, like route and mode choice can be typically modeled in MATSim. On the other hand, travel diaries come with the disadvantage that all information taken from diaries is by definition not sensitive to policies. Furthermore, trip diaries are normally available for only a very small fraction of the population. Also, in many places, travel diaries are subject to strong usage restrictions because of data privacy regulations, which often results on a strong coarsening of all spatial information.

In **this paper**, an efficient approach to creating a transport simulation scenario with a high policy sensitivity and low input data requirements is described. It circumvents aforementioned issues related to more traditional input data types and explores the use of Big Data as partial substitutes. The result of this study is a process chain is developed, described, applied in a prototype scenario and, based on initial results, evaluated.

2. Methodology

The creation process of a policy-sensitive transport simulation scenario in the present study (illustrated in Fig. 1) consists of the four steps of (1) creation of a population, (2) using cell-phone-based origin-destination matrices to derive work and education locations for all members of the population who are workers or take part in the education system, (3) applying a new, aspatial activity-scheduling model with comparatively frugal input data requirements,

and (4) taking advantage of MATSim's demand adaptation capabilities, notably also for the selection of locations for discretionary activities.

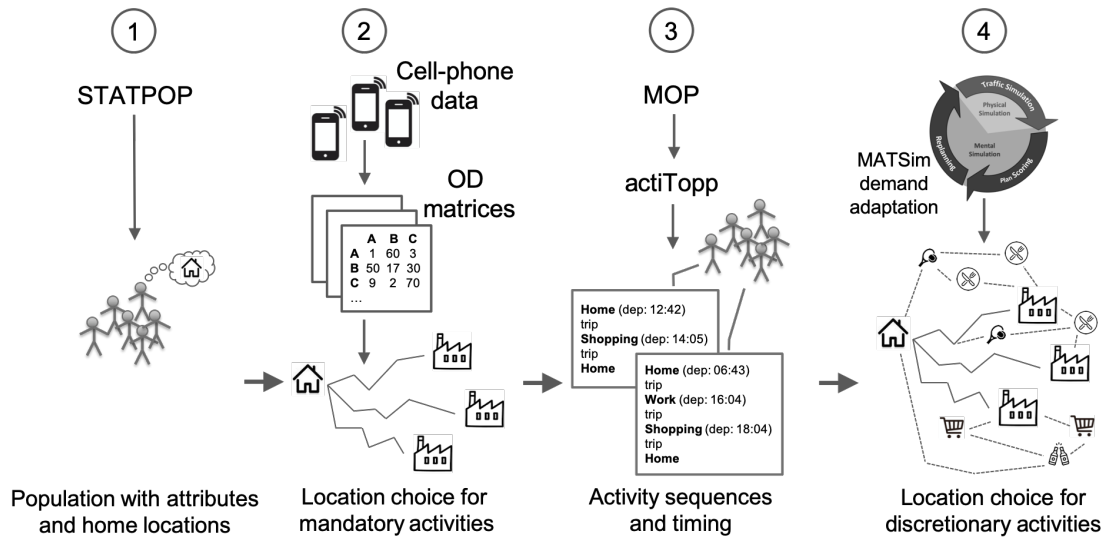


Fig. 1: New process of efficient transport simulation scenario creation based on cell-phone data and the actiTopp activity-scheduling model.

This approach is put into practice as a prototype for the metropolitan area of Zurich in Switzerland. This region is particularly suitable because multiple MATSim scenarios (among them the very first MATSim scenario) have been created for this region with a high standard of scenario validation, such that scenario-to-scenario comparisons can be carried out. Based on these existing scenarios, also the transport supply side is already readily set up and can be transferred to this study. Also, various other input data exist at high quality standards and are, at the same time, easier to procure than elsewhere. Last but not least, the cell-phone network operator Swisscom provided mobile-phone-based origin-destination matrices for Switzerland within the context of this project.

2.1. Population with relevant personal attributes and home locations

To create a population for the transport model, one can either translate a full population census or apply synthetic population generation methods (e.g. Müller and Axhausen (2011)) based on a population microsample and corresponding aggregate statistics of the whole population. Both approaches are established in the MATSim context. Since this step is not the focus of our analysis, it is welcomed that the STATPOP dataset (Bundesamt für Statistik, 2016), a full population census of the whole Swiss population could be used to derive the model population directly.

By this, the full population of the study area with properties of the individuals like gender, age, household context, and, in particular, home location specified to the level of municipality is known. The actiTopp model, which will be applied to model activity schedules for each member of the study area's population in a subsequent step (cf. Sec. 2.3), requires additional commute distances to work or education. These will be derived from mobile-phone-based data as explained in the next section (cf. Sec. 2.2).

2.2. Work and education locations based on ubiquitous Big Data

Origin-destination (OD) matrices from mobile-phone data records are used to derive work and education locations for each member of the study area's population. As such, this project is related to research by Zillske and Nagel (2013) who found that it is possible to obtain traffic patterns from mobile phone sightings.

For the present project, the Swiss mobile-phone network operator Swisscom provided 24 origin-destination matrices, which record movements of people on the municipality level on averaged working days on a month-by-month basis. Since the fall season is generally regarded to be 'average' in various respects (e.g. not a vacation time, not many

public holidays, most people following their regular work and education schedules), the matrix for the month of October 2016 is selected for this modeling step. From this OD matrix, the municipality-to-municipality movements of the four time slices from 6:00 to 10:00 are extracted and listed by origin municipality. For each member of the synthetic population designated as working or attending an education facility, a movement with the the person's home municipality as origin is selected randomly from the corresponding list. The person's workplace/education municipality is set as the destination of the selected movement. By this procedure, it is ensured that the outgoing morning traffic of each municipality in the study area is in agreement with people's movements as captured in the mobile-phone-based OD matrices.

2.3. Activity sequences and duration based on a new, aspatial activity-scheduling model

To model activity sequences and activity durations, the new and comparatively data-frugal activity-based demand model (ABDM) *actiTopp* (Hilgert et al., 2017) is used. Like MATSim, it is an open-source software written in Java, which facilitates model integration significantly as standard dependency-management technologies can be applied. Like other activity-based demand models, *actiTopp* is based on the concept of utility-based regression models and stepwise modelling. In contrast to other models, *actiTopp* models the activity participation of individuals over the course of a full week. The model is estimated based on data from the German Mobility Panel (MOP, IfV, accessed 27 February 2019), a national household travel survey for Germany, and is therefore (without model updates) representative for the travel patterns of the whole population of Germany.

Both the modeling capabilities and the input data requirements of *actiTopp* are lower than those of other activity-based demand models. *actiTopp* does neither have a notion of travel nor space, but models pure activity sequences with corresponding durations. Trips between subsequent activities are unspecified beyond the fact that they take place. Locations are not modeled. On the input side, *actiTopp* only requires age, gender, occupation status, number of cars in household, number of children in household, home location setting, and commuting distance to work or education (if applicable).

Both the limited modeling scope and the low input data requirements make *actiTopp* a very good fit for MATSim. On the one hand, travel-related properties are modeled by MATSim as pointed out in Sec. 1. In the case of the application of other activity-based demand models there would be overlaps in the modeling scope as these ABDMs model trip-related attributes like trip modes, which should from the perspective of MATSim rather be modeled MATSim-internally to guarantee a reasonable exploration of the search scope of potential travel alternatives. While these overlaps can be handled by a corresponding design of the interplay between the ABDM and MATSim, it is still an advantage if such an overlap in modeling scopes does just not exist, as it is the case with *actiTopp*. Finally, *actiTopp*'s low input data requirements are an advantage from an application point of view.

2.4. Discretionary activities based on MATSim's destination choice module

actiTopp models activity sequences of individuals and corresponding durations (cf. Sec. 2.3). Neither trips nor locations are modeled. In the present approach, the agents' home locations are known based on population input data. Locations of mandatory activities (i.e. attending work and education) have been derived from mobile-phone-based OD matrices (cf. Sec. 2.2). The locations of discretionary activities (also referred to as secondary activities, e.g. shopping or taking part in leisure activities) is, however, unspecified until now.

Horni et al. (2012) have developed an extension to MATSim which models the choice of facilities for such discretionary activities. This approach, which has not been applied extensively since initial studies, is revisited for the present approach as it perfectly suits the modeling task at hand. Horni's location choice approach fixes (*freezes*) the agent- and location-specific randomness inherent in the decision (as seen from the perspective of the modeler) such that a realistic representation of choices of secondary locations emerges when allowing agents in the simulation to switch locations while trying to improve their daily activity-travel behavior by the MATSim's coevolutionary algorithm (cf. 1). This *frozen preferences* approach is necessary to prevent that agents optimize their daily schedule in such a way as to simply visit the spatially closest activity facility of a kind. This would be unrealistic as people in reality prefer locations with certain properties (e.g. a tennis court over a swimming pool among leisure facilities). Such distinctions constitute, all in all, constitute details that are beyond the scope of what can be reasonably modeled. The described approach which initially determines preferences and then fixes them by means of frozen error terms, effects

that agents have to optimize between inherent utility of a place (unknown to the modeler) and disutility of reaching that place, such that a realistic spatial distribution can emerge by the iterative demand adaptation process. As such, discretionary activity location choice can be included in the demand adaptation as described in Sec. 1. In the present study where no information of activity facilities is used, activity facilities for discretionary activities are added to the scenario by randomly adding corresponding activity option (with a very low probability) to existing home locations.

3. Results

As a reference, a MATSim simulation scenario for the Zurich metropolitan area, developed at the Institute for Transport Planning and Systems at ETH Zurich, is used (Hörl et al., 2019). The scenario synthesis process is data-driven with inputs from the full Swiss population census (Bundesamt für Statistik, 2016), the general Swiss national household survey pooled for three years, and the Swiss household travel survey, among other data sets (see Hörl, 2020). The process follows closely the one presented for the open scenario of the Île-de-France region (Hörl and Balac, 2020). While the latter can be generated in a reproducible way from open and publicly available data, the Swiss scenario relies heavily on closed, privacy-critical data, which is only available to Swiss research institutions. The synthesis process generates households, persons and activity chains, and it already assigns departure times, primary and secondary locations with good fit to the reference data. Validation was among others performed based on highway counts which for most count stations show a deviation of less than 20%.

Based on this well-validated reference, it seems plausible to do perform a scenario-to-scenario comparison in order to gain an insight into the usefulness of the present approach. Table 1 compares the average duration of activities by type between the reference scenario and the new approach presented in this study. Except education, all average

Table 1: Comparison of activity participation durations [h] by activity type.

Activity type	Reference scenario	New approach
Work	5.45	4.99
Education	4.36	4.95
Shopping	0.77	0.92
Leisure	2.14	1.92
Other	0.64	0.67

activity duration deviate by less than 10% from the reference scenario. Despite the application of actiTopp outside its estimation context (cf. Sec. 4), agents spend roughly correct time shares at available activities.

Fig. 2 shows a comparison of the locations of trip origins throughout the day, with the darker dots denoting a higher number of originating trips. In both plot, the highest number of trips originate from Zurich's center and the overall patterns look very similar.

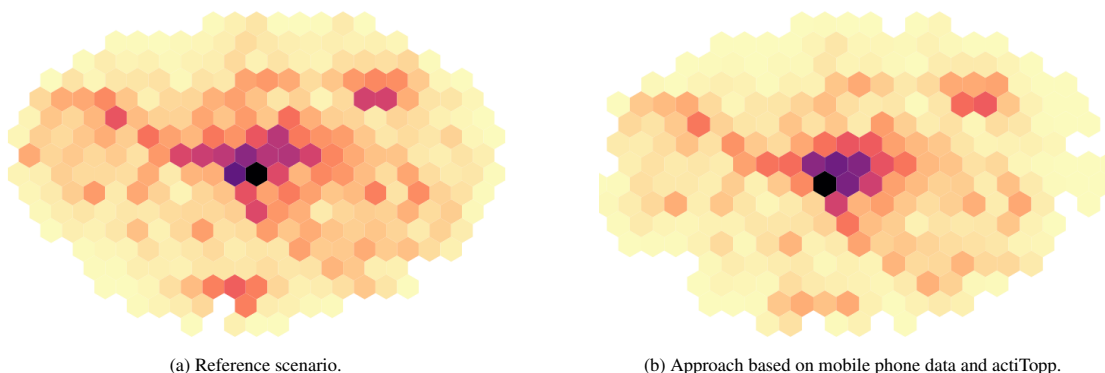


Fig. 2: Comparison of locations of trips starts (Darker colors mean more activity starts).

Fig. 3 depicts daily link volumes. As both pictures look very similar, it is confirmed that the traffic that derives from the modeled activity participation leads to generally usable results. While the distances and duration of trips

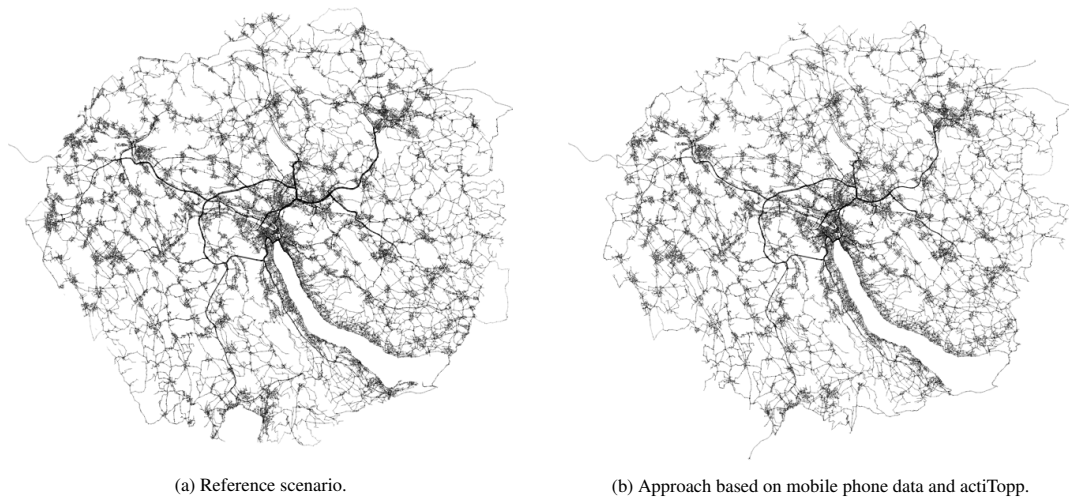


Fig. 3: Comparison of daily link volumes.

to work and education differ between the presented approach and the reference scenario by not more than 10% in each category, deviations for trip to other activities are larger. While the former confirms that the mobile-phone-based approach helps to selected realistic work and education facilities, it seems that the addition of facilities for discretionary described in Sec. 2.4 is over-simplistic and creates too many facilities in too close vicinity of agents, resulting in too short trips. The discussion section presents ideas to overcome this current shortcoming (cf. Sec. 4).

Fig. 4 looks into the simulated traffic as a resulted of the presented tool chain more closely and shows a comparison of simulated traffic in central Zurich on an average day at 7:00:00. While the pattern looks very similar, the overall congestion is somewhat higher in the reference scenario. This is expected as the reference scenario also includes traffic external of the study area, which the approach presented in this paper does not yet contain.

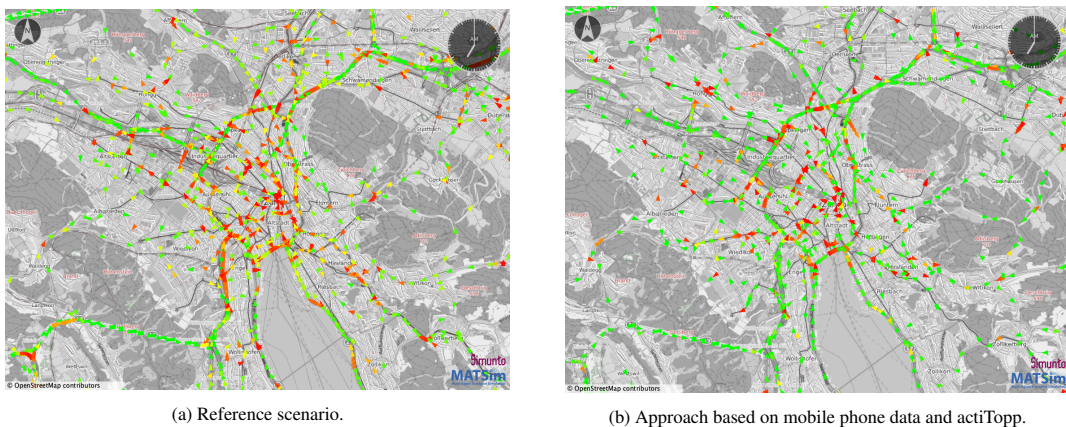


Fig. 4: Comparison of traffic situation in central Zurich at 7:00:00.

4. Discussion

The presented scenario creation approach is transferable to other regions. Censuses (at least in simple versions as required here) are available almost everywhere and mobile-phone data can (at least in technical terms) be derived ubiquitously because mobile phones are used worldwide in similar ways. As no technological break that will fundamentally change the use of mobile phone is foreseeable, approaches using such data should also possess sufficient longevity. The approach is also applicable if OD matrices were provided based on another, potential future data source.

The approach may be extended, e.g. for regions with more limited access to spatial data. In Germany, for instance, all locations in the microcensus ([Stat. Ämter des Bundes und der Länder](#), last accessed 20 June 2020) are (for the application by researchers) coarsened to the level of federal states (Bundesländer). As such, home locations are not sufficiently detailed represented based on this dataset in its available form. While in the current approach the mobile-phone-based OD matrix is only used to derive work locations, it could also be used to extract home locations to refine the spatial information given by a (spatially coarsely resolved) census.

In contrast to deriving activity-travel plans from travel surveys, the presented approach possesses a higher policy sensitivity. Reactions to policies can in principal be represented in terms the choice dimensions that are modeled endogenously by MATSim (i.e. route choice, mode choice, discretionary activity location choice) or actiTopp (activity sequences), while only home and work/education facilities that are fixed based on the corresponding inputs would remain unaffected from policies.

The proposed tool chain can be enhanced in terms of different components. Since actiTopp is estimated based on the MOP (German mobility panel), it is in principle only a valid estimation for Germany. One can argue that for a not highly-specified model like actiTopp whose estimation context is large (whole Germany), it should be possible to extent the application context to a demographically and culturally similar region. Clearly, such a model transfer needs to be confirmed by appropriate validation. If such a direct model transfer should turn out to be invalid, there are model parameter updating procedures that help making models applicable outside their original estimation context ([Rossi and Bhat, 2014](#)). As such, a previous approach to model updating in the context of MATSim ([Ziemke et al., 2015](#)) could be included into the tool chain of the present approach in order to improve model quality if further validation indicates this to be necessary. The latest MATSim model for the Berlin metropolitan region ('Open Berlin Scenario', [Ziemke et al., 2019](#)) is built on such an approach, where full daily plans created partially based on another estimation context are updated for the study region based on local traffic counts. Based on the Cadyts calibration tool, which interacts with MATSim's demand adaption (cf. Sec. 1), the scoring of plans is adjusted to approach real-world observations. Instead of traffic counts (as in the Berlin approach), those time sliced of the OD matrix that have not been used so far (until now only the time slices from 6:00 to 10:00 are used, cf. Sec. 2.2) may serve as real-world measurements in a Cadyts-based travel plan calibration.

As pointed out in the results section (cf. Sec. 3), trips to discretionary activities are too short based on the current development stage. This is very likely because of the somewhat over-simplistic localization of corresponding activity facilities based on home locations (cf. Sec. 2.2). An approach to derive real-world amenities as they are recorded on [OpenStreetMap](#) (last accessed on 20 June 2020) might improve this situation without impairing the goal to only use standardized inputs. In a study on dynamic accessibility computations ([Ziemke et al., 2017](#)), this approach has been established in the MATSim context and shown to work particularly well for amenities (like under consideration here), while it is not so well-suited to derive locations of workplaces (as workplaces do often – in contrast to discretionary activity locations – not constitute a point of interest and are, therefore, not sufficiently contained on OSM).

5. Conclusion and outlook

In this study, a new scenario creation process was shown which intends to make the creation of policy-sensitive agent-based transport simulation scenarios simpler. Only highly standardized input data sources are used and those very sparsely. The created procedure makes use of mobile-phone-sources OD matrices which can be assumed to exist in almost any location world wide. These matrices have been used to derive work location, but could also be used to derive other types of locations. For activity scheduling, the new, aspatial activity-based demand model actiTopp was used, which – based on its input specifications and modeling scopes – complements MATSim's modeling capabilities very well and helps maintaining the tool chain unsophisticated. All other decisions, including the choice of location

for discretionary activities, are modeled within MATSim's coevolutionary approach, which improves the approach's policy sensitivity. A prototype model was set-up for the metropolitan region of Zurich. An existing models of that region was used for comparison and to provide the transport supply system of the region. While the model based on the new approach shows plausible travel patterns and correct average participation rates, the representation of trips to discretionary activities should be improved in future research. The OD matrices can be used more extensively, e.g. for cross-validation against those time slices of the OD matrices, that have not been used in the steps of deriving locations.

Acknowledgements

The authors thank Tim Hilgert for discussions on the actiTopp model. DZ thanks Luk Knapen, Tom Bellemans, and Bruno Kochan for discussion on the interaction of activity scheduling and transport simulations models during repeated research stays at IMOB. The authors acknowledge funding by the National Research Programme "Big Data" (NRP 75) of the Swiss National Science Foundation and thank Swisscom for providing mobile-phone-based OD matrices as well as the Swiss Bundesamt für Statistik for granting access to the STATPOP census dataset.

References

- Bellemans, T., Kochan, B., Janssens, D., Wets, G., Arentze, T., Timmermans, H., 2010. Implementation framework and development trajectory of FEATHERS activity-based simulation platform. *Transportation Research Record* 2175, 111–119.
- Bhat, C., Guo, J., Srinivasan, S., Sivakumar, A., 2004. A comprehensive econometric microsimulator for daily activity-travel patterns. *Transportation Research Record* 1894, 57–66. doi:10.3141/1894-07.
- Bischoff, J., Maciejewski, M., 2014. Agent-based simulation of electric taxicab fleets. *Transportation Research Procedia* 4, 191–198. doi:10.1016/j.trpro.2014.11.015.
- Bundesamt für Statistik, 2016. Population and households statistics (statpop). <https://www.bfs.admin.ch/bfs/en/home/statistics/population/surveys/statpop.assetdetail.8541.html>.
- Hilgert, T., Heilig, M., Kagerbauer, M., Vortisch, P., 2017. Modeling week activity schedules for travel demand models. *Transportation Research Record* 2666, 69–77. doi:10.3141/2666-08.
- Horni, A., Nagel, K., Axhausen, K.W., 2012. High-Resolution Destination Choice in Agent-Based Models. Annual Meeting Preprint 12-1988. Transportation Research Board. Washington, D.C. Also VSP WP 11-17, see <http://www.vsp.tu-berlin.de/publications>.
- Horni, A., Nagel, K., Axhausen, K.W. (Eds.), 2016. The Multi-Agent Transport Simulation MATSim. Ubiquity, London. doi:10.5334/baw.
- Hörl, S., 2020. Dynamic Demand Simulation for Automated Mobility on Demand. Ph.D. thesis. ETH Zurich. Zurich. doi:10.3929/ethz-b-000419837.
- Hörl, S., Balac, M., 2020. Reproducible scenarios for agent-based transport simulation. a case study for paris and Île-de-france. doi:10.3929/ethz-b-000412979.
- Hörl, S., Balać, M., Axhausen, K.W., 2019. Pairing discrete mode choice models and agent-based transport simulation with matsim, in: 2019 TRB Annual Meeting Online, Transportation Research Board, Washington, DC. pp. 19–02409. doi:10.3929/ethz-b-000303667. 98th Annual Meeting of the Transportation Research Board (TRB 2019); Conference Location: Washington, DC, USA; Conference Date: January 13-17, 2019.
- IfV, accessed 27 February 2019. German mobility panel. <http://http://mobilitaetspanel.ifv.kit.edu/english/index.php>. URL: <http://http://mobilitaetspanel.ifv.kit.edu/english/index.php>.
- Kaddoura, I., 2015. Marginal congestion cost pricing in a multi-agent simulation: Investigation of the greater Berlin area. *Journal of Transport Economics and Policy* 49, 560–578.
- Müller, K., Axhausen, K.W., 2011. Hierarchical ipf: Generating a synthetic population for switzerland URL: <http://hdl.handle.net/20.500.11850/39167>, doi:10.3929/ETHZ-A-006620748.
- OpenStreetMap, last accessed on 20 June 2020. <http://www.openstreetmap.org>.
- Rossi, T., Bhat, C., 2014. Guide for Travel Model Transfer. Technical Report. Federal Highway Administration.
- Stat. Ämter des Bundes und der Länder, last accessed 20 June 2020. Mikrozensus. <https://www.forschungsdatenzentrum.de/de/haushalte/mikrozensus>.
- Ziemke, D., Joubert, J.W., Nagel, K., 2017. Accessibility in a post-apartheid city: Comparison of two approaches for accessibility computations. *Networks and Spatial Economics* 18, 241–271. doi:10.1007/s11067-017-9360-3.
- Ziemke, D., Kaddoura, I., Nagel, K., 2019. The MATSim Open Berlin Scenario: A multimodal agent-based transport simulation scenario based on synthetic demand modeling and open data. *Procedia Computer Science* 151, 870–877. doi:10.1016/j.procs.2019.04.120.
- Ziemke, D., Nagel, K., Bhat, C., 2015. Integrating CEMDAP and MATSim to increase the transferability of transport demand models. *Transportation Research Record* 2493, 117–125. doi:10.3141/2493-13.
- Zilske, M., Nagel, K., 2013. Building a minimal traffic model from mobile phone data, in: NetMob 2013, 3rd International Conference on the Analysis of Mobile Phone Datasets, Special session on the D4D challenge, MIT (Cambridge, MA). URL <http://perso.uclouvain.be/vincent.blondel/netmob/2013/NetMob2013-program-v1.pdf>.