

CHAPTER 70

Germany

Johannes Illenberger

The Germany scenario was developed by DB ML AG (DB Mobility Logistics AG), a subgroup of DB AG (Deutsche Bahn AG), the German state-owned railway company, in 2014. To help evaluate MATSim's applicability in the strategic planning process, as well as defining its compatibility in the traditional zone-based four-step process, this scenario has been constructed to establish a Germany-wide O-D matrix for private car travel. A solid understanding of the transport market for private car travel (rail's major competitor) is required for the strategic planning process at the DB ML AG. This scenario intentionally focuses just on road transport since, on one hand, there are already well-established models for rail transport at the DB ML AG and, on the other hand, this scenario is meant to be the first step towards MATSim's application.

Considering this scenario's objective, MATSim is used here as a tool to build a microscopic representation of the current transport market. Unlike the majority of MATSim studies, the focus is not to build and calibrate a behavioral model with forecasting power to answer the "what if" question. Hence, this study emphasizes reproducing empirical measurements, rather than on modeling plausible causalities and behavioral processes.

The final outcome of this exercise is a O-D matrix with average daily trip volumes. Although a higher temporal resolution is possible and also available from travel data, this dimension will not be considered during calibration and validation of the scenario. The matrix is based on a zonal structure with approx. 10 K TAZs, with a granularity comparable to municipalities (LAU 2) and a higher detail in large cities.

How to cite this book chapter:

Illenberger, J. 2016. Germany. In: Horni, A, Nagel, K and Axhausen, K W. (eds.) *The Multi-Agent Transport Simulation MATSim*, Pp. 431–436. London: Ubiquity Press. DOI: <http://dx.doi.org/10.5334/baw.70>. License: CC-BY 4.0

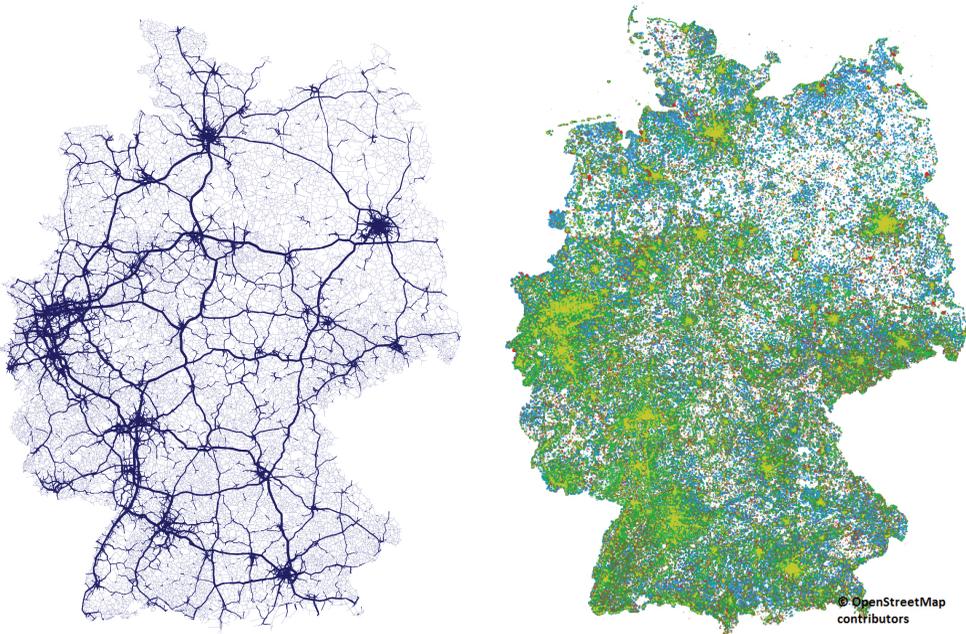


Figure 70.1: Left: Simulation road network including roads down to the level of major arterials. Right: Spatial distribution of activity facilities colored according to activity type.

70.1 Demand and Supply Data

The German national travel survey MiD from 2008 (Follmer et al., 2010) builds the foundation for the synthetic population and its travel plans. The survey features approx. 60 K person records with one-day travel episodes. A travel episode specifies trip sequences with mode, purpose, travel distance and day of reporting. Home locations are known at the level of states and municipality type (urban/rural). From each record, a synthetic person with one travel plan is generated. Initially, activity locations are set to random facilities and each person is cloned multiple times (according to the person’s weight), so that in total a population of 4 M persons results.

The road network is extracted from OSM. The geographical resolution of the O-D matrix allows omission of minor roads (everything below “tertiary” in OSM terminology) and the network is further simplified so that connected nodes with a distance less than 50 meter are merged to one node. The resulting network consists of 126 K nodes and 360 K links (Figure 70.1).

Activity facilities are taken from OSM as well. A facility can be synthesized from a OSM node representing a point-of-interest (shop, restaurant, bar, etc.), a polygon representing a building and a polygon representing a region with specified land use. In the latter case, multiple facilities are generated proportional to the polygon’s area. Activity options are inferred from meta information associated with the node or polygon. Given the huge amount of “home” facilities, only a 20 % subsample is used. This still yields a total of 5.6 M activity facilities.

70.2 Imputation and Calibration

The location of activities—origin and destination of trips—are required for building the O-D matrix. The travel survey does not provide any information on activity locations. Thus, the

study's main task is to impute plausible activity locations, given the sequence of trip distances and underlying geographical distribution of activity facilities. The intermediate solution resulting from this imputation is calibrated against count stations and selected O-D flows from car navigation devices.

The imputation process implementation can be considered as a Monte Carlo Markov Chain simulation converging into a distribution where the activity locations configuration best fits the constraints imposed by trip distances, count stations and selected O-D flows. Solving this task in one simulation process would be congruent with theory. However, considering the scenario size and computational limitations, the process is split into three steps: (i) assigning "home" locations, (ii) generating an initial state with assigned "non-home" activity locations, and (iii) varying a subset of "non-home" activity locations to meet car volumes at count stations and selected O-D pairs' flows. Steps (i) and (ii) can be considered as imputation processes and are realized outside of the MATSim iteration framework. Step (iii) can be considered the calibration step and is realized with a MATSim-Cadyts setup configured as a Monte Carlo engine (see Chapter 48).

70.2.1 Imputation of Home Locations

Home locations are known at state and municipality type levels. The municipality type is divided into six categories by number of inhabitants. Initially, each person is placed on a random home facility, while inhabitants' geographical distribution is given at the TAZs level. A Monte Carlo simulation relocates persons to best meet their specified state and municipality. More formally:

1. Generate an initial configuration \mathcal{P}_k :
 - (a) Randomly assign each person n a home facility.
 - (b) Evaluate the configuration: $H(\mathcal{P}_k) = \sum_n \theta_1 \delta_n + \theta_2 |m_n - m_n^*|$, where δ_n is 0 if the person is located in the correct state and 1 otherwise, m_n denotes the category index of the current person's municipality and m_n^* its target category. Parameters θ_1 and θ_2 control how close the simulation converges to the target values.
2. Generate a new configuration \mathcal{P}_{k+1} by switching the home facilities of two random persons.
3. Accept the new configuration \mathcal{P}_{k+1} with probability $\pi_{k+1} = 1 / (1 + \exp(H(\mathcal{P}_{k+1}) - H(\mathcal{P}_k)))$, otherwise return to \mathcal{P}_k .
4. Repeat step 2 and 3 until the system reaches a steady state distribution.

Switching home facilities, instead of assigning a random facility, ensures that persons' spatial distribution remains constant. Running the simulation for 10^9 iterations results in a configuration where more than 90 % of persons are located in their correct state and, on average, three of four persons are located in their correct municipality; the fourth person is just one category index distant from its target index.

70.2.2 Imputation of Non-Home Locations

Activity locations (non-home) are assigned with an analogous process, like home locations. The simulation relocates activities so that resulting trip distances best meet their empirical target distances. In this imputation step, distance always refers to the beeline distance and thus avoids expensive route search.

About one fourth of all trip chains are composed of more than two trips; that is, trips are not symmetrical. Accordingly, drawing a random activity location on an annulus centered at the origin

activity location, with radius according to the target distance, does not necessarily fulfill the target distances of succeeding trips. The simulation process is specified as follows:

1. Generate an initial state \mathcal{P}_k :
 - (a) Assign a random facility to each non-home activity by considering the activity type and the facility's activity options.
 - (b) Evaluate the configuration: $H(\mathcal{P}_k) = \sum_n \sum_q \theta_3 \left| \frac{d_{nq} - d_{nq}^*}{d_{nq}^*} \right|$, where d_{nq} denotes the realized distance of n 's trip to activity q and d_{nq}^* the target distance.
2. Generate a new configuration \mathcal{P}_{k+1} by assigning a random facility to a random non-home activity (by considering activity type and activity options).
3. Accept the new configuration \mathcal{P}_{k+1} with probability $\pi_{k+1} = 1 / (1 + \exp(H(\mathcal{P}_{k+1}) - H(\mathcal{P}_k)))$, otherwise return to \mathcal{P}_k .
4. Repeat step 2 and 3 until the system reaches a steady state distribution.

A configuration is evaluated based on the relative error or realized distances to target distances, so that short and long trips are treated equally. Parameter θ_3 controls the randomness. That is, if $\theta_3 = \infty$ each trip would exactly (if possible) meet its target distance, which, however, is not the desired solution. Rather, θ_3 is adjusted so that there is some randomness in realized trips distances, but without distorting global target distance distribution.

70.2.3 Calibration

The outcome of step 2 (Section 70.2.1) and 3 (Section 70.2.2) is a valid population with imputed home and activity locations. In the third step, the population is calibrated against measurements of count stations and flows of selected O-D pairs. This step is implemented in a "standard" MATSim-Cadyts combination. The scoring function accounts only for the "linear plan effect", agents are allowed to have only one plan and `SelectExpBetaForRemoval` is used for the `planSelectorForRemoval` parameter. All Cadyts parameters are left to their defaults.

During replanning, non-home activities that are not part of a complex trip chain are relocated. More specifically, an activity is valid for relocation if:

- The facility of the previous and succeeding activity is equal (round trip), or
- the activity is the origin of the first trip, or
- the activity is the destination of the last trip.

The above conditions ensure that complex trip chains that have been adjusted in step 2 (Section 70.2.2) are not distorted. New activity locations are randomly chosen within a distance of $\pm 10\%$ of the trip's target distance, so that global distance distribution is conserved.

70.2.3.1 Counts Calibration

The German Federal Highway Research Institute, BASt (Bundesanstalt für Straßenwesen), provides average daily vehicle volumes (distinguished in car and trucks) yearly, measured at about 1500 count stations on motor- and highways. After separating each station's data into both directions of traffic and validating measured vehicle volumes, about 2 500 link volumes are available for calibration. Empirical car occupancy rates (depending on trip purpose) are taken from the MiD to convert person volumes to car volumes.

70.2.3.2 O-D Calibration

O-D calibration is based on an O-D matrix representing car navigation device flows. Since occupancy rate of devices in cars is unknown, only the distribution of flows is used for calibration.

Further, comparison with other data sources shows that the O-D matrix is only valid for O-D pairs above a distance of 100 kilometers. This appears reasonable, considering that navigation devices are probably not switched on for short (likely commuter) trips. Accordingly, only O-D pairs above 100 kilometers distance and, from those, only the 6 000 pairs with the highest volumes are extracted into a reduced *calibration matrix*. The latter conditions ensures that only pairs with a sufficient sample size are used.

The reduced calibration matrix is normalized to the sum of all trips in a *reference matrix*. The reference matrix contains all trips from the initial population that correspond to all non-null O-D pairs (matrix entries) in the reduced calibration matrix. This yields a calibration matrix with valid absolute trip volumes.

For each O-D pair, a virtual link is inserted into the road network at runtime. A virtual count station with the corresponding count value from the reduced calibration matrix is attached to each virtual link. In the mobility simulation, after a person arrives at its destination, it then travels the virtual link corresponding to the traveled O-D pair. This travel, however, is only communicated towards the Cadyts calibrator by injecting additional `PersonDeparture`, `LinkLeave`, and `PersonArrival` events.

70.3 Simulation Results and Travel Statistics

The synthetic population of 4 M persons corresponds to approx. 8.5 % of the German population that conducts at least one car trip per day. This yields a scaling factor of 11.8 by which all simulation statistics need to be multiplied. The simulation produces an overall trip volume of 57.5 M trips, quite close to 57.2 M trips in the official statistics (DIW, 2014). Passenger mileage of 947 G person-kilometers is slightly overestimated compared to 917 G person-kilometers in the official statistics, yet still reasonable.

Figure 70.2 visualizes the calibration results in a scatter plot. Each dot represents a count station (left) or a O-D pair (right), respectively. On average, absolute value of the relative error yields 0.18, considering vehicle volumes at count stations and 0.16 considering O-D flows. A median relative

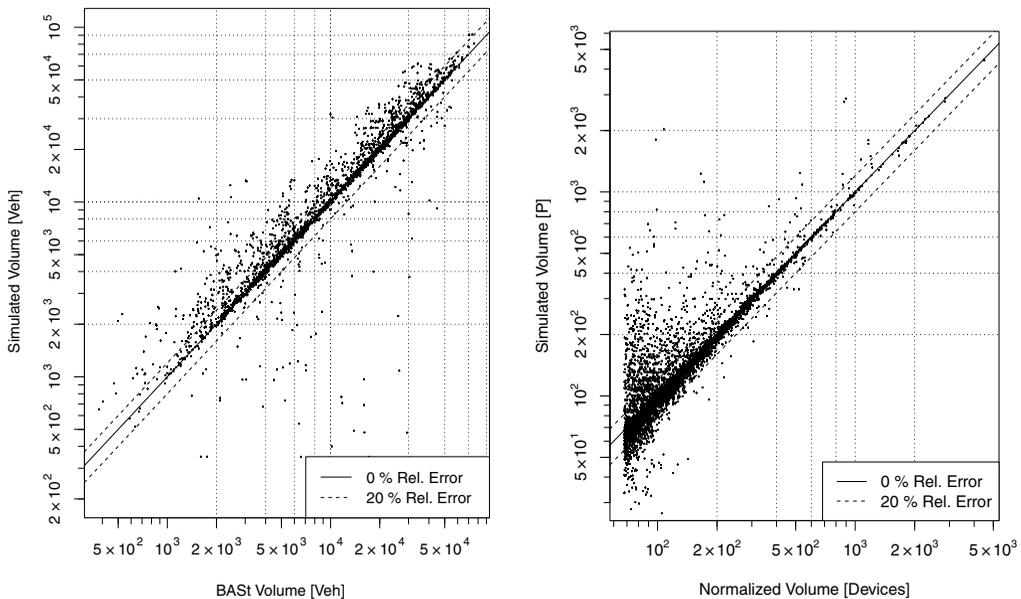


Figure 70.2: Comparison of simulation and empirical measurements. Left: count stations. Right: O-D pairs.

count station error of 0.12 indicates that there are a few count station that are significantly off. These two issues should be noted; first, there is only a description of the count station's location, which is not always unique and may be misinterpreted by the matching algorithm. This can yield a false assignment of count stations to network links. Second, there is no information on the count stations' measuring error, or on temporarily capacity reductions (for instance, caused by construction sites) modeled in the simulation network.

O-D flows error correlates with the O-D pairs' volumes; thus, pairs with low volume show, on average, a higher error. This is related to this scenario and MATSim's characteristics: a population sample size (8.5 %) and the discrete nature of MATSim. For instance, a real-world O-D flow with 118 individuals is represented by ten agents in the simulation. A variation during re-planning to this O-D pairs by, say, one agent already has a significant impact on the scaled real-world value. Averaging over multiple iterations reduced the variance but does not entirely remove the effect.