Berlin II: CEMDAP-MATSim-Cadyts Scenario

Dominik Ziemke

To correctly model initial demand properties not included in MATSim iterations in specific studies (i.e., activity choice), suitable data are needed. Travel diaries containing departure times, mode choice decisions and activity locations are widely used. However, much of this data source content, particularly location information, is considered sensitive in terms of data privacy legislation and thus increasingly difficult to obtain and process in many areas (e.g., in Germany and the United States; Ziemke et al., 2015).

The *Berlin II scenario* (also referred to as the *CEMDAP-MATSim-Cadyts scenario* according to the applied models in its setup), is the outcome of an alternative approach relying exclusively on freely available or easy-to-obtain input data. All of these data do not rely on individual trajectories, but instead on "anonymous" data that is aggregated so much that the data providers are no longer concerned about privacy issues.

The starting point for this scenario is a publicly available commuting matrix containing homes and workplaces of workers with social security on the municipality level. Based on this information, it is possible to model morning and evening commuting peaks.

To obtain a full-population demand representation, two further major modeling steps are required. First, in cases like the Berlin case, see below, where the commuter matrix spatial resolution is quite coarse, higher resolution O-D information is necessary. Second, a procedure is needed to model secondary activities, i.e., all other activities beyond home and work.

The importance of the first step becomes obvious when looking at the German case; here, the whole city of Berlin, with 3.4 million inhabitants, is represented by exactly one zone (Bundesagentur für Arbeit, 2010). In the United States, commuting matrices are typically available only on a county-to-county level. Since such location-aggregation-based matrices may become the rule, rather than the exception, in privacy-sensitive societies, a (generalizable) method to attain O-D information at a higher resolution is needed (Ziemke et al., 2015). The standard solution would be to estimate an activity location choice model. This, however, is difficult if no trip data to estimate

the model is available. O-D matrix estimation studies (van Zuylen and Willumsen, 1980) suggest that traffic counts may be used to make an initially rough O-D matrix more appropriate for a region. As MATSim is not based on O-D flows, but on full daily plans, the issue comes down to whether a procedure exists to update these initial full daily plans using traffic counts. In the approach used to create the Berlin II scenario, a procedure proposed by Flötteröd et al. (2011) and implemented in the software Cadyts—explained in Chapter 32—is applied for this task. Specifically, random draws of possible home and work locations within the home or work municipality given by the commuter matrix are made. Various MATSim plans, each containing one pair of home and work locations, are created for each agent. Then, the Cadyts calibration procedure is applied within the iterative MATSim simulation to select plans and locations more likely to occur with given traffic counts.

As stated above, however, full daily plans (as opposed to mere home-work-home commuting patterns) are needed. Therefore, the second modeling step, the modeling of secondary activities for each individual in the region, needs to be addressed. For the Berlin II scenario, CEMDAP (Comprehensive Econometric Microsimulator for Daily Activity-Travel Patterns (Bhat et al., 2008)) is used to generate initial complete daily plans for each individual. On one hand, however, no CEMDAP parameter set is available for Berlin. On the other hand, and more importantly, one major goal of the study creating the Berlin II scenario was to show its generalizability (Ziemke et al., 2015). So, the model parameters of CEMDAP estimated for the Los Angeles region (the estimation context) are retained and then used to generate initial plans for individuals in Berlin (the application context in the current paper), based on Berlin demographic data.

To sum up, home and work municipalities are taken from the commuter matrix. Within these municipalities, a set of (more precisely spatially defined) potential home and work locations are randomly chosen for each agent. Full daily plans incorporating the various potential locations of each agent are generated with CEMDAP, based on a parameter set from another region.

Then, the Cadyts calibration procedure is used to select those initial full daily plans most consistent with Berlin traffic count data. In other studies, Cadyts has already been applied to update route choice predictions, both for car (Flötteröd et al., 2011a) and for public transit (Moyo Oliveros and Nagel, in press). However, it has not been used to update full daily activity-travel plans, as it was in the procedure that created the Berlin II scenario.

The Berlin II scenario is thus an activity-plan-based MATSim transport model for Berlin based exclusively on freely, or readily, available data. If a commuter matrix, some basic population demographics, and traffic counts (or, theoretically, another suitable data source on which to run the calibration procedure) are available for a particular regional context, the approach used to create the Berlin II scenario can be transferred to that other context. In fact, the Berlin II scenario itself should be seen as a *transferred model*, because initial plans generated by CEMDAP are based on parameter estimates from another geographic region (the Los Angeles area).

Through a validation based on the Berlin 2008 SrV (System repräsentativer Verkehrsbefragungen (Ahrens et al., 2009)), an extensive, regularly-conducted travel survey, the created transport demand representation quality has been successfully tested. So far, the Berlin II scenario exists for a 1% and a 10% population sample of all persons, i.e., including workers without social security, as well as non-working people, aged 18 and above, for the study region. Currently, only motorized traffic is considered. Stability tests, showing that plausible agents' daily plans continue to be chosen when Cadyts calibration functionality is switched off, have been successfully carried out. This is a clear indication that the scenario is applicable and meaningful for policy studies.

Further improvements, like the addition of public transport and a more realistic representation of the population, are planned. Moreover, similar approaches to integrating activity-travel pattern generators (e.g., the FEATHERS model) with MATSim in transport simulation are planned.