How to Improve MATSim Destination Choice For Discretionary Activities?

Andreas Horni
Kay W. Axhausen
How to Improve MATSim Destination Choice For Discretionary Activities?

Andreas Horni
Institute for Transport Planning and Systems (IVT)
ETH Zürich
CH-8093 Zürich
phone: +41-44-633 31 51
fax: +41-44-633 10 57
horni@ivt.baug.ethz.ch

Kay W. Axhausen
Institute for Transport Planning and Systems (IVT)
ETH Zürich
CH-8093 Zürich
phone: +41-44-633 39 43
fax: +41-44-633 10 57
axhausen@ivt.baug.ethz.ch

May 2012

Abstract

This paper proposes ways to improve MATSim destination choice for discretionary activities. Naturally, small-scale destination choice models can feature a much higher level of detail than large-scale microsimulation scenarios. This upscaling problem mainly concerns data availability and computational limitations, but, due to the relative novelty of microsimulation destination choice models, also a certain implementation lag exists. This paper discusses on the example of MATSim and in a preliminary sense, where upscaling of small-scale findings for application in large-scale microsimulation scenarios is possible and potentially beneficial.
1 Introduction and Problem

Modeling destination choice for discretionary activities (e.g., shopping and leisure) involves a broad range of disciplines having produced an ample body of research (see Section 2.1). Naturally, a relatively large gap exists between this broad basis of knowledge and its application in large-scale transport microsimulations. Main barriers for application of state-of-art complex destination choice models in microsimulations, are lack of required large-scale data, computational issues and, due to the relative novelty of microsimulations and possibly catalyzed by the fist two barriers, a certain lack of recent implementation progress.

This paper discusses, in a preliminary sense, where upscaling of small-scale findings for application in large-scale microsimulation scenarios is possible and potentially beneficial.

The significance of the upscaling problem is often overlooked, when taking a strictly theoretical focus as its sources are rooted in practice. Consequently, to not miss the crux of the matter, the investigation needs to be grounded on a specific model and specific data. Thus, MATSim, in particular its recently implemented destination choice module, and the Swiss data readily available to modelers is focused.

2 Review

2.1 Small-Scale Destination Choice Models

Prominent choice determinants are travel time, travel distance and price and quality level. The influence of person attributes on store choice is studied in e.g., Uncles (1996), Shim and Eastlick (1998). Many researchers additionally incorporate constraints given by e.g., store opening hours or by the travel time budgets defined within the individuals activity chains (e.g., Arentze and Timmermans 2005, Delleart et al. 1998, Kitamura 1984). The moderating influence of the costumers’ store loyalty on the choice process is examined in East et al. (1998), Knox and Denison (2000). Studies considering region of Zurich are Carrasco (2008), Kawasaki and Axhausen (2009), Horni et al. (2012).

2.2 Large-Scale Transport Microsimulation Destination Choice Models

In the microsimulation community the term microsimulation is ambiguously used, sometimes denoting movement simulation of persons and vehicles (as a replacement of volume-delay functions in aggregate models) and sometimes additionally including preceding choice processes. As we focus on destination choice, we are interested in comprehensive simulators only and, thus, we use the term microsimulation in this sense throughout the paper. Numerous transport microsimulations are available, amongst others ALBATROSS, AMOS, OpenAmos, CEMDAP, MERLIN, MOBITOPP, New Yorks Best Practice Model, PCATS, Ramblas, San Francisco microsimulation model, TASHA, TRANSIMS, VISEM & Vissim.

However, literature reviews such as Algers et al. (1998), Henson and Goulias (2006) are rare and only cover the very basics of each model. An overview of the destination choice models applied in large-scale microsimulation is planned as a continuation of this paper.

MATSim destination choice is discussed in Section 2.4.

2.3 Data Availability

According to the authors’ experience and personal communication, Swiss data base mainly provided by BfS (2011) is comfortable; it can thus be interpreted as an upper bound for governmental data availability.

The main Swiss data sets used in MATSim are:

- Census of Population (Swiss Federal Statistical Office, 2000), a full survey, applied to create the MATSim population, including their home and work locations on hectare and
municipality level respectively,

- National Travel Survey (Swiss Federal Statistical Office 2006), an approx. 30,000 person sample, used for MATSim demand creation (activity chains and times)
- Business Census (Swiss Federal Statistical Office 2001), identifying enterprises at hectare level, utilized for creation of activity locations,
- Network Data, (TomTom MultiNet 2011, NAVTEQ 2011, Vrtic et al., 2005), spanning navigation and planning networks,
- Road Counts, (e.g., ASTRA 2006), specifying hourly traffic volumes per lane, mainly applied for MATSim validation.

Data covering only Zurich region concern parking supply data, green times (Balmer et al., 2009), store service hours (Meister 2008) and public transport lines and schedules (e.g., Rieser 2010, p.70ff).

### 2.4 MATSim Destination Choice

#### 2.4.1 MATSim—In Brief

MATSim is an activity-based, extendable, open source, multi-agent simulation toolkit implemented in JAVA and designed for large-scale scenarios and is a co-evolutionary model. A good overview of MATSim is given in Balmer et al. (2006). In competition for space-time slots on transportation infrastructure with all other agents, every agent iteratively optimizes its daily activity chain by *trial and error*. Every agent possesses a fixed amount of day plans memory, where each plan is composed of a daily activity chain and an associated utility value (in MATSim, called *plan score*).

Before plans are executed on the infrastructure in the network loading simulation (e.g., Cetin 2005), a certain share of agents (usually around 10%) is allowed to select and clone a plan and to subsequently modify this cloned plan.

If an agent ends up with too many plans (usually set to “4-5 plans per agent”), the plan with the lowest score (configurable) is removed from the agent’s memory. One iteration is completed by evaluating the agent’s day described by the *selected* plan.

If an agent has obtained a new plan, as described above, then that plan is selected for execution in the subsequent network loading. If the agent has *not* obtained a new plan, then the agent selects from existing plans. The selection model is configurable. In many MATSim investigations, a model generating a logit distribution is used.
Computation of plan score is compatible with micro-economic foundations. The basic MATSim utility function was formulated in [Charypar and Nagel (2005)] from the Vickrey model for road congestion as described in [Vickrey (1969)] and [Arnott et al. (1993)]. Plan utility described in detail in [Charypar and Nagel (2005)] is computed as the sum of all activity utilities plus the sum of all travel (dis)utilities.

### 2.4.2 State of MATSim Destination Choice

The Swiss Census of Population 2000 ([Swiss Federal Statistical Office, 2000](#)) can identify home and work locations for every Swiss resident at hectare and municipality level respectively. Clearly, such information cannot be logged for discretionary activities. However, to run an activity-based simulation, reasonable destinations for these activities must be assigned. First, a simple neighborhood search, as described in [Balmer et al. (2009)](#), was employed in a pre-processing step. That approach is not part of the optimization process and does not accurately model destination choice.

A first improvement in destination choice—including it in the optimization process—was introduced by [Horni et al. (2009)](#), based on Hägerstrand’s time geography. However, unobserved heterogeneity was not taken into account explicitly in that module or in MATSim. Thus, a significant underestimation of travel demand resulted and the module could not be productively employed. Furthermore, that module is based on local search. Local search applicability, however, is questionable on discontinuous destination choice utility space.

MATSim has been made fully compatible with discrete choice methodology with the integration of the second destination choice module described in [Horni et al. (2012)](#).

MATSim and its destination choice module takes into account travel time and distance, travel costs, store size and competition effects at activity locations ([Horni et al. (2009)](#)). Agents’ attributes taken into account in MATSim are age, mobility tools, occupancy, home and work location. Destinations are characterized by location, service hours and one of the types home, work, shop, leisure and education. For validation MATSim mainly uses road count data.

### 2.5 Upscaling Small- to Large-Scale Models

Concluding the Sections 2.1-2.4, the main barriers for upscaling small-scale findings for large-scale application in microsimulations are sketched in this section.
2.5.1 Limited Data Availability

Clearly, models of higher resolution, such as microsimulations, require more and different data than traditional methods (for activity-based models discussions see e.g.,[Axhausen](1997), [Kitamura](1996)). Fortunately, data collection techniques have evolved in the same vein as the models; [Nagel and Axhausen](2001, p.6) expect a "virtual explosion of data availability" besides others due to the spread of novel communication devices, usually equipped with GPS units. However, data collection is generally associated with privacy, cost and technical issues (for example, precision of GPS and GSM data) generating hard limits of data availability, such that some needed data possibly will always be missing.

This is a serious problem as besides higher resolution and thus potentially higher accuracy, sensitivity to parameters and input data and thus to policy measures is often an argument for disaggregate rather than aggregate methods ([Sbayti and Roden](2010) p.4), ([Lemp et al.](2007). Consequently, a weak data base might annihilate the conceptual advantages of microsimulation models. Furthermore, validation as a crucial methodological modeling step, is limited by lack of data ([McNally and Rindt](2008) p.7). Question is here, to which extent local small-scale validation is applicable.

Reliable imputation methods are thus of great importance. Synthetic data can, although a priori only matching marginal distributions, generate value by capturing correlations correctly.

On the other hand, speaking against an extensive data base, models should be general, i.e., flexible and transferable ([Patriksson](1994) p.5), meaning that they are applicable without extensive data collection and model estimation efforts (see also Figure 1). In other words, data should be condensed in general model concepts.

2.5.2 Computational Difficulties

Another source, limiting fast progress in destination choice modeling, are computational issues. For example, destination choice is associated with large numbers of destinations potentially leading to very large choice sets. In large-scale scenarios this often represents a limiting factor. For a more elaborate discussion and approaches to deal with these problems see [Horni et al.](2012). Another source for computational problems is the stochasticity of microsimulation runs ([Horni et al.](2011a). Dependent on the resolution levels, many simulations runs, i.e., samples are required, which limits level of sophistication for a single run.
2.5.3 Implementation Backlog

Limited data availability and computational issues slow down implementation of destination choice models in microsimulations. In addition to these factors, the effort for model application in complex large-scale frameworks generates a certain implementation backlog. Clearly, while identifying efficient microsimulation improvements, emphasis should be put on implementation backlog as it is expected that it can be resolved comparatively easy.

3 Research Avenues

3.1 Further Heterogeneity of Agents and Alternatives: Prices and Income

Compared to the models mentioned in Section 2.1 MATSim attribute range and thus agents’ and alternatives’ heterogeneity is very limited. Prices and income and derived measures such as value-of-time (VOT) are essential and central in any econometric model, but they are not yet included in MATSim. Conceptually, prices and income are easy to survey as they are not subject to latency like other choice determinants. Practically, privacy issues and the large number of suppliers render data collection nevertheless difficult and costly.
To generate income models for Switzerland valuable sources are provided such as Swiss Federal Statistical Office (2008b, 2007, 2006). To get a more comprehensive model, usage of rents as a proxy for income could be inspected. At the price frontier very few data is available.

Important to test is spatial price and income heterogeneity, in other words, scale of wealth separation. Smaller heterogeneity means that model resolution must be higher to capture income and price differences. In conclusion, extensive spatial aggregation of incomes and prices thus harbors the risk of not exploiting microsimulation power. This might be true for other choice determinants as well.

### 3.2 Destination Choice Equilibration

The central concept in transport modeling is equilibrium (e.g., Wardrop 1952, Beckmann et al. 1956). Traditionally, assignment procedures, modeling route choice, equilibrate static network flows given a fixed demand. While for route choice a strong influence of attributes governed by competition (e.g., travel time) can be assumed beyond doubt, this is less evident for destination choice. It is thus not clear to which extent destination choice actually needs to be subject of the iterative equilibration process of MATSim. Relaxing the strict equilibrium assumption has potential to strongly reduce the computational burden. In terms of behavioral base, papers, investigating the empirical basis of the convenient assumption of equilibrium (e.g., Mahmassani and Chang 1986) should be consulted and their applicability to destination choice should be assessed.

### 3.3 Finer Activity Classification

The most frequently used Switzerland and Zurich scenario contain only 5 activity types (home, work, education, leisure and shop). Improved versions differentiating shopping and leisure activities (Horni et al. 2011b) are available, but only in an experimental manner. The National Travel Surveys (Swiss Federal Statistical Office 2006) and Swiss Federal Enterprise Census 2001 (Swiss Federal Statistical Office 2008a) provide a relatively detailed classification of activities for demand and supply side respectively. Most activities are in principle performed by anybody, e.g., grocery and non-grocery shopping does not distinguish person groups. However, the exploitation of the data promises higher quality in particular in connection with agglomeration effects and multi-purpose shopping activities.
3.4 Spatial Correlations in Destination Choice

On both supply and demand side, interactions between actors, i.e., customers and firms (see Figure 3) exist, which are observable in spatially correlated destination and location choices and which materialize in agglomerations, a very important topic in economics. In this paper, spatial correlations in customers destination choices are focused.

3.4.1 Customer Interaction Effects at Activity Locations

The central influence of interaction in transport infrastructure for people’s route and departure time choice has been recognized early (e.g., Pigou [1920], Knight [1924], Wardrop [1952]).

Similarly, agent interaction in activities infrastructure might affect destination choice (Axhausen [2006]), as it generally changes activity utility. Person interactions at activity locations can have positive or negative influence on persons’ gained activity utility. Presence of other people at...
recreational places such as bars, discos or party locations usually increases utility, whereas competition for parking lots (see also Section 3.7) or crowdedness in shops clearly reduce utility (Albrecht, 2009, p.119 ff). Marketing science literature provides ample evidence of significance of these effects (Baker et al., 1994, p.331), (Eroglu and Harrell, 1986, Eroglu and Machleit 1990, Eroglu et al. 2005, Harrell et al., 1980, Hui and Bateson, 1991, Pons et al., 2006).

In modeling, supply side capacities, such as store sizes, parking facility sizes, number of cash points are common. However, demand and supply are only very seldom equilibrated. Rare examples are de Palma et al. (2007) for residential location choice and Vrtic (2005) for route and mode choice. Microsimulation models factoring in interaction effects in the activity infrastructure are Vovsha et al. (2002), Horni et al. (2009), Waraich and Axhausen (2011).

While the inclusion of destination interaction effects in microsimulations is conceptually expedient, before continuing this line of research, the magnitude of interaction effects, i.e., their significance, needs to be quantitatively researched (see also Section 3.2). Furthermore, capacity data is required but difficult to collect. For Switzerland, this might be solved by deriving capacities from the available disaggregate employment information (given by the number of full time equivalents) (Swiss Federal Statistical Office, 2008a). Future analyses must also answer the
question if activity infrastructure load should be microsimulated (analog to the network loading simulation) or if the typical aggregate cost-load-curves can be applied approximately.

Different utility-load relationships must be assumed. In relatively static contexts with sharp capacity limits also utility-load function shows a sharp decrease at capacity limit. In more variable contexts, for example, inside stores or on very large parking sites, a softer form of utility-load function can be expected. An example, lying in-between, might are large restaurants.

Models taking into account spatial correlations can be efficiently estimated using copulas (Bhat and Sener 2009).

3.4.2 Spatial Distribution of Destinations

Clustered destinations, i.e., agglomerations, help minimizing travel effort between the shopping activities of multi-stop shopping trips (see e.g., Bernardin et al. (2009), p.144), Arentze et al. (1994), p.89, Arentze et al. (2005), Popkowski Leszczyck et al. (2004), Messinger and Narasimhan (1997), Oppewal and Hoyoake (2004). Even for single-purpose shopping, agglomerations might be beneficial due to a risk reduction of not finding specific products at current location or if shopping is done as a leisure activity.

In conclusion, agglomerations generate utility beyond sum of single opportunities (see also Teller and Reutterer (2008), Teller (2008)). Including these effects in the models, and thus capturing frequencies at large shopping malls and nightlife areas better, is particularly important for weekend scenarios currently under development for MATSim. Models considering spatial distribution of destinations are Fotheringham (1985, 1983a,b), Fotheringham et al. (2001), Timmermans et al. (1992), Berry et al. (1962)

3.5 Choice Dimension Dependencies

In MATSim to date, essentially all combinations of choices of different dimensions (i.e., time-route-mode-destination choices) (see e.g., Hannes et al. (2008), Timmermans (1996), Cadwallader (1995)) are evaluated over the course of the iterations. However, replanning of agents’ day plans is not done under consideration of choice dimensions’ dependencies. A priori using knowledge about choice combinations, e.g., its likelihood, might generate a substantial speed-up, also relaxing the computational problems mentioned above.
3.6 Cognitive Spatial Models

Cognitive models of persons’ spatial mental map are promising in destination choice context, in particular to get under control its large choice sets. Papers to be regarded are Axhausen (2006), Chorus and Timmermans (2009), Hannes et al. (2008), Mondschein et al. (2008), Arentze and Timmermans (2004), Golledge and Timmermans (1990), Bettman (1979), Timmermans (2008), Cadwallader (1975). Apart from Dobler et al. (2009) cognitive models have not been applied yet in MATSim.

3.7 Parking

According to literature, parking search induced traffic is substantial (Shoup, 2005) and consequently an ample body of parking literature (for a review see e.g., Young et al. (1991)) exists, spanning a huge number of empirical studies investigating parking itself but also its interaction with other travel choices (such as destination choice) exist (van der Waerden et al., 2009, 2006; Marsden, 2006; Widmer and Vrtic, 2004; Anderson and de Palma, 2004; Golias et al., 2002; Hensher and King, 2001; Gerrard et al., 2001; Baier et al., 2000; Albrecht et al., 1998; van der Waerden et al., 1998; Axhausen et al., 1994; van der Waerden et al., 1993; Glazer and Niskanen, 1992; Topp, 1991; Axhausen and Polak, 1991; Arnott et al., 1991; Polak and Axhausen, 1990; Feeney, 1989; Miller and Everett, 1982; Gillen, 1978; 1977; Maley and Weinberger, 2011), but also numerous simulations such as Benenson et al. (2008), Gallo et al. (2011), Thompson and Richardson (1998), Dieussaert et al. (2009), Young (1986), Young and Thompson (1987) have been developed.

An experimental parking search model has been implemented in MATSim (Waraich and Axhausen, 2011) and further consideration of parking for destination choice modeling seems reasonable, although, preliminary results of a current GPS data analysis at the authors’ institute show that parking search is might basically overestimated in Switzerland.
4 References


