Coupling customers’ destination choice and retailers’ location choice in MATSim

Working Paper

Author(s):
Horni, Andreas; Ciari, Francesco; Axhausen, Kay W.

Publication date:
2012-09

Permanent link:
https://doi.org/10.3929/ethz-b-000053546

Rights / license:
In Copyright - Non-Commercial Use Permitted

Originally published in:
Arbeitsberichte Verkehrs- und Raumplanung 808
Coupling Customers’ Destination Choice and Retailers’ Location Choice in MATSim

Andreas Horni
Francesco Ciari
Kay W. Axhausen

Working paper 808
Transport and Spatial Planning September 2012
**Abstract**

The multi-agent transport simulation MATSim (MATSim, 2012) contains modules for customers’ destination choice (Horni et al., 2012) and retailers’ location choice (Ciari and Axhausen, 2011; Horni and Ciari, 2009, 2011). This paper’s goal is combining these two models into a dynamic and micro-simulated demand-supply model for shopping activities. Both models are methodologically improved by analyzing retailer clustering and its effects on customer destination choices and by discussing extension of the normative retailer location choice model toward a behavioral model.

**Keywords**

Shopping destination choice, retailing location choice, microsimulation

**Preferred citation style**

1 Introduction

1.1 General Context

Transport microsimulations are an example of disaggregate activity-based transport models being state-of-art in research but not yet state-of-practice. Modeling unit is the individual with its activities and trips chain defined over the whole modeling period. Microsimulations are conceptually appealing and thus, associated with enormous expectations. However, to date, these models do not yet outclass the aggregate models and superiority is difficult to show quantitatively (e.g., [Lemp et al., 2007]). Reasons are, on the one hand, yet limited calibration and validation data and on the other hand simply the newness of the approach not yet providing a broad base of sophisticated models ([Horni and Axhausen, 2012]). In conclusion, microsimulations are not only a research tool but likewise a research topic; this paper focuses on the later, i.e., methodological, side.

1.2 Problem Description / Research Goal

This paper’s goal is is combining the MATSim retailer location choice and customer destination choice module and to discuss retailer clustering in agglomerations and its effects on customer destination choices for future application in a MATSim model. The paper represents a follow-up work of both [Ciari and Axhausen, 2011] and (Horni et al., 2012). Methodological progress in supply side simulation is increasingly important as the coupling of MATSim with land-use models (e.g., Urbansim ([UrbanSim, 2011])) will be intensified in the near future ([Nicolai et al., 2011]).

The paper starts with a short characterization of real-world customer destination choice (Section 1.3) and retailer location choice (Section 1.4) and their modeling. Based on that, the proposed model is discussed in Section 2.

1.3 Customer Destination Choice

Consumer shopping destination choice involves a broad range of disciplines, such as transport and urban planning, marketing and retailing science, economics, geography, psychology (e.g., environmental psychology) and sociology. This has produced an ample body of research. A plethora of choice determinants are identified where central ones are travel time and distance, price level, product quality and diversity, parking situation and store size. Clearly, strong
interdependence with other travel choice dimensions exist, with e. g., mode choice (e. g., Hannes et al., 2008; Timmermans, 1996; Cadwallader, 1995).

Consumer choices are often (and also in this paper) modeled within the discrete choice framework (McFadden, 1978), where determinants are usually attributed to the persons (e. g., age), to the alternatives (e. g., store size) or the choice situation (e. g., season, time of day, weather etc.).

While the discrete choice standard multinomial logit (MNL) model (containing gravity models as special cases) does not account for correlated alternatives as present for example in agglomerations, improved models have been developed which explicitly consider positive and negative agglomeration effects such as Fotheringham (1985, 1983a,b); Bernardin et al. (2009).

1.4 Retailer Location Choice

Location choice is a very important business decision (Fox et al., 2007; Rogers, 2007). It greatly influences business success. At the same time the decision is overwhelmingly complex and associated with high costs for corrections. An overview of the numerous retailer location choice determinants and location search strategies can, for example, be found in Ciari et al. (2008); Bodenmann (2005); Brown (1994); Hale and Moberg (2003); Hernández and Bennison (2000); Hernández et al. (1998). For model construction it is important to note, that location choice strategies are, due to their high complexity, driven by the whole range from hard (or rational) to soft (or emotional) criteria (see e.g., Scherer and Derungs (2008, p.12)). In other words, location choice is essentially characterized by heuristics (Maier and Toedtling, 2006, p.23/26/41), (Schmidt, 1980, p.60), which usually lead to approximate, suboptimal, but efficiently computable solutions. Furthermore, relocation of stores is associated with very high costs, thus, search is basically non-iterative, meaning that ex-post optimization through relocation is relatively weak. However, a certain global optimization exists due to business failure (Maier and Toedtling, 2006, p.33).

Complexity of the choice process and thus, of the choice model is, inter alia, generated by latency of choice determinants and temporal variability. A normative model might assume instantaneous profit as the only optimization objective. However, profit is—although some indicators might are known—essentially latent and varies with time (see also Maier and Toedtling, 2006, p.23). Factors, which promote high profits over longer time periods (such as political stability or potential for expansion, i.e., adaptability (Scherer and Derungs, 2008, p.9)) like-wise may shall be included as optimization objectives in their own right. Furthermore, as internal profit-generating mechanisms strongly differ between firms, optimization objectives are also dependent

Innes et al., 1990; Recker and Kostyniak, 1978; Timmermans, 2008; Timmermans and van der Waerden, 1992; Timmermans et al., 1982; van der Waerden et al., 1998; Yang et al., 2009; O’Kelly, 1983; Kitamura et al., 1998; Barnard, 1987; Bekhor and Prashker, 2008)
The driving sources of retail clustering and dispersion are manifold, where spatial distribution of residents and their buying behavior, infrastructure and relations between firms, and, of course, location management play an important role (see also Maier and Toedtling (2006, p.96ff)).

In this paper, clustering, induced by customers and infrastructure are explicitly considered: Consumers may have additional benefit from visiting retail clusters due to shorter ways for multi-purpose shopping, often larger range of choices and additional services such as good parking facilities, restaurants or day nurseries in shopping malls (see also Maier and Toedtling (2006, p.99)). Infrastructure can contribute to clustering by influencing consumers’ destination choices. Not considered here are retailer input side positive external agglomeration effects due to spatial proximity between firms dealing with each other (e.g., sub-supplier and supplier).

2 Method

2.1 MATSim in Brief

MATSim (MATSim-T, 2011) is an activity-based, extendable, open source, multi-agent simulation toolkit implemented in JAVA and designed for large-scale scenarios. It is a co-evolutionary model; 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 or by directed search within one iteration. 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. Computation of plan utility is compatible with micro-economic foundations and given as the sum of all activity utilities plus the sum of all travel (dis)utilities. In other words, agents perform optimization with an objective function (agent’s utility function) and a search strategy (trial-and-error or directed search).

Agents’ day plan optimization is usually assumed to lead to an equilibrium state, where no agent can unilaterally improve. This equilibrium—solution of a non-cooperative game—is interpreted as a Nash equilibrium (Nash, 1951) or in transport assignment context as a Wardrop equilibrium (Wardrop, 1952). Finding this equilibrium represents a fixed point problem (e.g., Bierlaire and Crittin, 2006): choices of agents affect infrastructure conditions and vice versa or in more general demand and supply is interdependent. In numerics a plethora of methods to solve such problems exists, which dependent on initial state may actually find a fixed point, oscillate or even show chaotic behavior. In MATSim, search for this equilibrium is based on a co-evolutionary algorithm, where in principle, this equilibrium can be searched by random search or by exhaustive search, i.e, every agent randomly changes its plan or simply enumerates through...
all plans and memorizes the best one. However, the problem is characterized by combinatorial complexity and, thus, more efficient searches must be applied for most choice dimensions. E.g., for routing Dijkstra algorithm is applied, instead of randomly selecting one of the infinitely many routes between origin and destination. Characteristics of the co-evolutionary algorithm are only researched scarcely, nevertheless, in practice MATSim equilibrium searching has been shown to be productive.

2.2 Customers’ Destination Choice in MATSim

As detailed in Horni et al. (2012), destination choice is implemented in the discrete choice framework. Decisions are modeled as a utility maximizing choice from a finite set of alternatives. Utility is composed of observed factors and random error terms. Inclusion of random error terms allows to model choices which are—at least in the modeler’s eyes—irrational, in other words, sub-optimal i.e., pareto-dominated in terms of observable choice determinants. This is relevant for destination choice modeling but even more important for retailer location choice modeling as shown later. Due to stochasticity of this approach, clearly, application must be principally based on ensembles, i.e., on multiple Monte-Carlo runs with different random seeds (Horni et al., 2011).

In MATSim at the moment the only observed destination choice determinant is travel time, where choice, is embedded in the agent’s daily activity chain. In other words, succeeding and preceding activity and its location influence choice. Agent i’s destination choice for a specific shopping activity is modeled as:

\[ dc_{\text{customer}}^i = \arg \max_{j \in \text{choice set}} f(t_{\text{travel},ij}, \epsilon^\text{customer}_{ij}) \]

where the travel time to/from destination j from previous/to succeeding location in activity chain \( t_{\text{travel},ij} \) is dependent on infrastructure conditions and the agent’s complete day plan. Through minimization of \( t_{\text{travel},ij} \) subject to constraints agents’ destination choices are, in more general terms, dependent (amongst other factors) on accessibility.

To complete the gravity model approach, it is planned—to begin with—to include store size as an attracting factor as investigated in Horni et al. (2009).

2.2.1 Destination Choice and Correlations

MATSim destination choice is based on a simple MNL model, meaning that correlations of alternatives as present in retailer agglomerations are not taken into account explicitly. In other words, an isolated store is, a priori, equally attractive to an otherwise identical store being part
of an agglomeration.

These correlations of potential destinations are considered in this paper as follows. Expectedly, through minimization of travel times between the shopping activities of multi-stop shopping trips, retail agglomerations are, implicitly, preferred to a certain extent (see [Bernardin et al. (2009, /p.144), Arentze et al. (1994, /p.89), Arentze et al. (2005); Popkowski Leszczyc et al. (2004); Messinger and Narasimhan (1997); Oppewal and Hoyoke (2004)]. Customer preference for agglomerations, not explained by travel time reductions (examples are given in Teller and Reutterer (2008); Teller (2008)) are modeled with an additional agglomeration term \( \tau^{+}_{agglo} \). Value of this parameter, as well as, definition of functional form needs empirical investigation. Here, we assume an S-shaped function

\[
\tau^{+}_{agglo,j} = f \left( \text{sizeof agglomeration}_j \right)
\]

where further attributes, such as variability of store types, will be included in the future. Agent \( i \)'s choice thus becomes:

\[
dc_{i}^{customer} = \arg \max_{j \in \text{choice set}} f \left( t_{travel,ij}, \epsilon_{ij}^{customer}, \tau^{+}_{agglo,ij} \right)
\]

Adding \( \tau^{+}_{agglo,ij} \) to MATSim utility function is straight-forward. Modeling multi-stop shopping trips is more complex. Daily activity chains are derived from National Travel Survey (Swiss Federal Statistical Office, 2006), which reports these multi-shopping activities as one single shopping activity. As information is provided about the number of visited stores, the existing MATSim scenario can be enhanced by splitting these activities, which is planned for the future. With implementation of multi-stop shopping also multi-purpose shopping should be implemented which means that shopping activities need to be differentiated, for example, into grocery and non-grocery shopping. Thereby, part of the unobserved heterogeneity which is now contained in \( \tau_{agglo} \), can then be explained: For stores of the same category, probably competition or substitution effects between firms are dominant, whereas the consumers’ preference to do multi-purpose shopping at one location with complementary goods stores generates agglomeration effects.

Competition effects between consumers (or negative agglomeration effects)—captured with an additional parameter \( \tau_{comp,ij} \)—are generated through limited parking supply (van der Waerden et al. [1998]; Weis et al. [2011]), crowdedness in shops (e.g., Albrecht (2009); Baker et al. (1994); Eroglu and Harrell (1986); Eroglu and Machleit (1990); Eroglu et al. (2005); Harrell et al. (1980); Hui and Bateson (1991)) or overfull restaurants to name a few.

Agent \( i \)'s choice finally is:

\[
dc_{i}^{customer} = \arg \max_{j \in \text{choice set}} f \left( t_{travel,ij}, \epsilon_{ij}^{customer}, \tau^{+}_{agglo,ij}, \tau_{comp,ij} \right)
\]

This utility function’s coefficients can be estimated similar to Datta and Sudhir (2011, /p.8)
reporting results for US grocery shopping.

2.3 MATSim Retailer Location Choice in MATSim

Location choice models can be roughly divided into normative, behavioral and structural approaches (Maier and Toedtling, 2006, /p.24ff). In this section, we describe a normative approach (loosely based on Ciari and Axhausen (2011)). The coupling of this normative model with destination choice allows to research retailer clustering. We also discuss conversion of this normative model into a behavioral model.

MATSim model creation always means formulation of a utility function and design of an efficient search strategy being productive for large-scale scenarios (recent projects simulate metropolitan area Zurich (Balmer et al., 2009) or also complete Switzerland).

2.3.1 Normative Model

Coupling of above destination choice model with a normative location choice model is comparatively straight forward. Retailer location choice should only be done every $n^{th}$ iteration when consumer decisions have stabilized after last supply change.

Utility Function: Following neo-classical location choice theory (Maier and Toedtling, 2006 /p.24f,44), the utility function can be set up as

$$l_{c_{retailer}} = \arg\max_{j \in \text{choice set}} \text{profit}_j,$$

where profit at location $j$ is:

$$\text{profit}_j = \text{in}_j - \text{out}_j.$$

$out$ are costs for e.g., wages and rents. $in$ is revenue, which is here calculated as $\text{revenue} := \bar{e} \times n_{\text{customers}}$, i.e., as the average expenses per person $\bar{e}$ times the number of actual customers in the store. Optimally, actual rents or real estate prices should be generated endogenously based on modeled demand.

Similarly as, for example, in Huang and Levinson (2011) this normative location choice model can be coupled with a behavioral destination choice model as formulated above, or else it can be coupled with normative consumer decisions, i.e., with a consumer utility function that does not contain random error terms (but agglomeration and competition terms $\tau_{agglo}$ and $\tau_{comp}$).
**Search Space Handling:** Exhaustively searching potential locations is computationally infeasible as the huge number of potential locations and agents defines a gigantic problem characterized by combinatorial complexity; all combinations of locations and agents need to be evaluated with exhaustive search. Only solution is, also for normative model, application of search heuristics.

A natural heuristic approach would be to move retailers toward network links, with a high load of agents doing a shopping activity. Remember that this information can be extracted from the previous MATSim iteration. Obviously, this system is prone to get stuck in oscillations. Thus, damping has to be incorporated, which can be done by introducing probabilistic choice weighted by load as described above. Additional damping can be introduced by also taking into account agents’ primary activities (work and home activities); these activities are spatially fixed in MATSim. In [Ciari and Axhausen (2011)] only primary activities are considered.

**Limitations and Productiveness of Normative Model:** MATSim and the above two modules have the following current limitations. While households are a basic modeling unit in economics, transport models have started with individuals (individual trips to be more specific) and are now on their way toward household units. MATSim is based on individuals not households (see also [Maier and Toedtling (2006)](/p.10)), although there is some recent progress in this direction. The proposed model focuses on firms output side (see also [Maier and Toedtling (2006)](/p.19)), i.e., influence of sub-suppliers is neglected. This also means that only agglomeration effects due to customers’ preferences and infrastructure can be modeled ([Maier and Toedtling, 2006](/p.99)). Firm networks and potential cannibalism effects are neglected in the first instance ([Maier and Toedtling, 2006](/p.67)). Furthermore, zoning regulations are not yet taken into account. Finally, for calculation of revenue, agents are not differentiated, for example, with respect to purchasing power.

Value of this model depends on level of optimization achieved in reality. As shown in Section 1.4, a high optimization level in location choice must be strongly questioned, i.e., retailers’ location choice is strongly driven by heuristics.

A normative model, thus, may produce value for individual location choice consulting but not for large-scale location choice prediction and modeling of agglomeration effects.
2.3.2 Behavioral Model

MATSim can be adapted to reproduce heuristic choice outcomes relatively easily by application of random error terms. Retailers’ location choices can be modeled as:

$$\text{arg max}_{l \in \text{choice set}} f(\text{profit}_l, \epsilon_{kl})$$

for retailer $k$ and location $l$, where $\epsilon_{kl}$ is the error term. This is identical with

$$l_{ck}^{retailer} = \text{arg satisfy}_{l \in \text{choice set}} f(\text{profit}_l).$$

Location choices differ significantly for large and small firms [Maier and Toedtling (2006), p.92], where large firms tend to decide more optimal (due to usually more planning resources). This effect can be modeled in the future by associating smaller random error terms to the decisions of large firms.

This first and natural approach, although perfectly consistent with discrete choice theory, suffers from following problems: Application of above utility function statistically fits heuristic choices and thus is rather a regression model than actually a behavioral model. First, random error terms $\epsilon_{retailer}$ by definition have no explanatory power. Second, profit as a location choice determinant is questionable, as it is usually only known exactly ex-post, and re-location is associated with high costs. This would bring us back to the normative model.

Consequently, a more sophisticated approach than application of error terms is required.

A closer reproduction of actual decision processes both in terms of choice determinants and search strategy means that agents must not have information, which is actually unavailable in real choice process. For example, exact information on profit, catchment areas and customers is not known exactly. Some indicators or proxies for these choice key determinants, however, derived from retailers already located there, are usually available and, of course, taken into account for the location choice. Thus, completely falling back on static observable proxies such as residential density or accessibility does not produce a consistent solution.

In other words, location choice is a long-term sequential and heuristic process, meaning that retailers usually know about and consider resident competitors for location choices. Thus, clearly, perfect model would be a long-term, rule-based land-use and transport model. MATSim, however, is an equilibrium model, computing a typical work-day. In other words, MATSim’s initial state and trajectory are meaningless and only final equilibrium is relevant, which also means that essentially only choices simultaneous in nature can be modeled. Long-term sequential and heuristic processes are usually modeled with rule-based approaches which is, besides equilibrium models the second large approach in transport microsimulations.
Modeling agglomeration emergence is also much more problematic in the behavioral model than in the normative model. In the normative model agglomerations are generated endogenously through retailers’ profit maximization; \( \tau_{agglo} \) in customer decisions will drive more people to retail clusters and generate more revenue and profit for the clustered firms. In the behavioral equilibrium-based model, however, relocation dependent on customer frequency information is problematic as argued above.

3 Conclusions and Outlook

In this paper a coupled MATSim model of customers’ shopping destination choice and retailers’ location choice is discussed, which extends Ciari and Axhausen (2011) and Horni et al. (2012) by correlations. To begin with cluster effects are considered explicitly on demand side. In future, further agent types as, e.g., proposed in Arentze and Timmermans (2003) may be included.

A normative location choice model, fitting well in MATSim, is described. Preliminary conclusion is that extension of this normative to a behavioral model is methodologically problematic, and instead, application of traditional land-use models (such as UrbanSim (2011)), usually based a sequential, long-term and rule-based approach should be considered.

Location choice is, in the aggregate, an very complex process. Accordingly, every model, inevitably, has numerous limitations/approximations (see also Section 2.3), where it is difficult to estimate the effects of such approximations. In other words, mathematical modeling of location choice is an enterprising undertaking and should be constantly critically reflected.

The normative model has following potential for investigations. The model allows to assess cluster building as result of the interplay of (i) accessibility, (ii) consumer benefit and loss due to agglomeration-specific clustering effects (\( \tau_{agglo,ij} \) and \( \tau_{comp,ij} \), respectively) and (iii) rents. As these can be controlled to a certain extent in reality by location management, the model may can be a comparative basis for location management and planning. With respect to cluster effects, the model allows to quantitatively analyze the antagonism between agglomeration and dispersion, or, in other words the relation between agglomeration benefits and monopoles given by space respectively accessibility (see e.g., Maier and Toedtling (2006, /p.55, 63)).

In a setting, where location choice is only performed for a few retailers (such as in Ciari and Axhausen (2011)), the normative framework may be also applicable for location choice consultancy.

Due to MATSim’s high-resolution Swiss data base, especially in terms of (navigation) networks (e.g., Tele Atlas MultiNet (2010)) and business building data (Swiss Federal Statistical Office 2008) it may be possible to thoroughly analyze accessibility configurations as for example
described in Maier and Toedtling (2006, p.63).

A main point for future work is empirical specification of $\tau_{agglo}$ and $\tau_{comp}$. These parameters can either be derived through calibration or through estimation based on surveys.
References


Fotheringham, A. S. (1983b) Some theoretical aspects of destination choice and their relevance


Fox, E. J., S. Postrel and A. McLaughlin (2007) The impact of retail location on retailer revenues:


