Agent-based simulation of travel demand: Structure and computational performance of MATSim-T

M. Balmer, K. Meister, M. Rieser, K. Nagel and K.W. Axhausen

Arbeitsbericht Verkehrs- und Raumplanung 504

June 2008
Kurzfassung


Schlagworte

MATSim-T, Verkehrsnachfrage, Rechenleistung, Agenten-basiertes Model,
Abstract

The model toolkit MATSim-T provides a variety of tools and resulting approaches to model travel demand and traffic flow and their interactions. The currently preferred configuration is presented in this paper together with detailed information about its computational performance. The application is relatively small, but as computing times scale linearly for the system it gives an idea of how the system can be used for practical planning studies. The outlook highlights the next steps of the development.

Keywords

MATSim-T, Travel demand, agent-based micro-simulation, computational performance

Preferred citation style

1 Introduction

The activity approach was first suggested as an alternative to the trip-based four-stage travel demand and assignment model nearly thirty years ago (Jones, Dix, Clarke and Heggie, 1983) building on path braking work on choice (Chapin, 1974; Domencich and McFadden; 1975) and constraints (Hägerstrand, 1970; Lenntorp, 1978) in travel demand. The research since then has addressed an enormous range of issues under this heading (see Goodwin, Kitamura and Meurs, 1990; Recker and Kitamura, 1985; Jones, Koppelman and Orfeuil, 1990; Damm, 1983 for reviews and the books of the conference series of the International Association for Travel Behaviour Research for more recent overviews, e.g. Axhausen, 2006 or Hensher, 2001). In recent years one stream of this work has focused on – finally - replacing the four-stage model with advanced, but robust application-ready models. While the model systems are far from replacing the trip-based commercial software now reflecting 50 years of development, they have made rapid advances in the last years. The conference in Austin documented the state-of-the-art two years ago (Bradley and Bowman, 2006) and this conference will provide an update, which will reflect the new concern with the implementation and speed of the model systems. This and the other invited papers will explain how they generate the initial agent population and its travel demand, how the systems allow the agents to adapt their demand into a coherent whole or how they capture the relaxation (learning) process, and how

the interactions on the networks are modelled. In addition, they will provide a breakdown of the computing effort for a typical application or sets of application of the relevant system.

The MATSim-T (Multi-Agent Transport Simulation Toolkit) presented in this paper reflects a long development history. German approaches to the micro-simulation of travel demand have existed since the mid-70’s (Poeck and Zumkeller, 1976; Zumkeller, 1989; Axhausen and Herz, 1989), and an early integration of travel demand and traffic flow simulation (Axhausen, 1988; Axhausen, 1990 were combined with ideas developed as part of the TRANSIMS project (TRANSIMS, 2006; Nagel, Beckman and Barrett, 1998) and fast traffic flow simulations (Nagel and Schreckenberg, 1992; Gawron, 1998). This basic set of concepts and ideas have been transformed and replaced in most part since the MATSim-T development started at ETH Zürich in 1998. The development is now undertaken jointly at TU Berlin, ETH Zürich and CNRS Lyon. Still, the basic evolutionary (relaxation) strategy has been retained throughout: an initial travel demand is executed with the traffic flow simulation, which returns improved and more consistent generalised cost estimates (travel time estimates, at minimum). These are used to update the previous schedules (activity sequences with their timing, locations, modes and routes; fellow travellers and activity participants and expenditures). This is repeated until there are no chances for further unilateral improvement anymore. Applications of the toolkit to path-dependent problems are possible, but will not be discussed here (Illenberger, Flötteröd and Nagel, 2007; Rommel, 2007)

The structure of the paper will follow the task list set out above plus one: How to generate agents; How to produce and update the schedules of the agents; How to execute the schedules; How to perform the iterative relaxation and finally report the performance for a specific, small application: route, time and mode choice relaxation of the greater Zurich area containing a 10% population sample of ca. 0.19 million agents / daily schedules (further details below). The paper concludes with a sketch of the further development planned for MATSim-T
and identifies chances for collaborative work among the various model systems presented here.

2 Conceptual overview

MATSim’s approach is iterative: Starting from an initial condition, the system is run over and over again while agents adapt. At the same time, boundary conditions are kept constant. Iterations have always been used in computational implementations of the Nash/Wardrop equilibrium assignment procedure (e.g. Sheffi, 1985); it has been clear for many years that feedback should consider all other choice dimensions besides route choice as well (e.g. Loudon, Parrameswaran and Gardner, 1997); and both evolutionary game theory (e.g. Hofbauer and Sigmund, 1988 and 2003; Hofbauer, Sigmund and Hofbauer, 1998; Yang, 2005) and the science of complex adaptive systems (e.g. Holland, 1998; Arthur, 1994) have shown the way to move forward. MATSim develops its iterative approach directly as an extension of the assignment procedure: The route adaptation process is extended towards other choice dimensions, such as time choice, mode choice, location choice, etc.

This means that, quite similar to a standard numerical simulation problem, a MATSim run can be characterised by the following elements:

- Boundary/initial conditions (land use, transport network, demographics, etc.)
- List of choice dimensions that are adapted

A problem with this characterisation is that many choice modules end up being listed twice: A route finding algorithm, for example, is needed both for the initial conditions and during the adaptation. This justifies the conceptual approach that is taken in the following, which looks at agent generation and scheduling separately, and takes for granted that decision modules
can either be run once, during the initial demand generation, or repeatedly, during the iterations.

### 3 Generating agents

Agent-based models require agents, preferably grouped into households, and even better grouped into social networks (Hackney and Axhausen, 2006; Arentze and Timmermans, 2006; Axhausen, 2005). The literature reports a large number of systems, often based on iterative proportional fitting (e.g. Beckman, Baggerly and McKay, 1996), which draw/generate agents from the fitted multi-dimensional table; the dimensions being the variables of interests with the required attribute values. The crucial issue is the construction of the tables with all required dimensions from a number of tables with fewer, but overlapping dimensions. This is required, as confidentiality requirements normally preclude the publication of high dimensional tables, especially for fine grained geographies.

The Swiss case has allowed us to ignore this issue, as we have access to household and person data from both a private “census”, as well as from the census (BfS, 2000) at a level of spatial resolution, which matches the finest network descriptions we are likely to use with reasonable accuracy. We generate the agents and their households by drawing from these populations.

While economic, this is not a long term solution, as the data ages, especially through residential moves, and more importantly it is not transferable to application outside Switzerland. See in the conclusions for the further work planned.

The place of work and education is drawn conditional on the home location from the census derived commuter matrix of Switzerland, which has been published every decade in electronic form since 1970.
Table 1 summarizes the approaches taken for MATSim-T’s preferred configuration.

### Table 1 Generating agents: Current preferred configuration of MATSim-T

<table>
<thead>
<tr>
<th>Task</th>
<th>Frequency per run</th>
<th>Model type</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Home location</td>
<td>Once</td>
<td>Conditional probability</td>
<td></td>
</tr>
<tr>
<td>Household characteristics</td>
<td>Once</td>
<td>Conditional probability</td>
<td></td>
</tr>
<tr>
<td>Person characteristics</td>
<td>Once</td>
<td>Conditional probability</td>
<td></td>
</tr>
<tr>
<td>Work/education location</td>
<td>Once</td>
<td>Conditional probability</td>
<td></td>
</tr>
<tr>
<td>Mobility tool ownership (car and season tickets)</td>
<td>Once</td>
<td>Imputed (MNL)</td>
<td>Ciari et al., 2008</td>
</tr>
</tbody>
</table>

### 4 Scheduling

The modelling of the schedule is the central task of an activity-based modelling approach, realising its vision of human behaviour as a coherent (daily) whole. The dimensions of the problem are:

- Number and type of activities
- Sequence of activities
- Starting time and duration of the activities
- Composition of the group undertaking the activity
- Expenditure and its allocation among the participants
- Secondary location choice (non-work and non-education)
- Mode/vehicle choice
- Route choice
- Point of egress from the vehicle
- Composition of the group travelling together
• Travel expenditure and its allocation among the travellers

The ideal model would address the joint processes by which this schedule is created (Axhausen, 2005). Travel behaviour models have so far not achieved this level of complexity, although there has been substantial progress from the initial treatment of isolated individual trips (Willumsen and Ortuzar, 2001). The models of the activity approach distinguish themselves by the selection of the dimensions which they treat, and by the grouping of the treated dimensions into joint sub-models. The approaches used to determine the attribute value (option) of the dimension are either best-response models or (stochastic) imputation models based on model derived probabilities or on conditional (observed) probabilities (Ben-Akiva and Lerman, 1985; Train, 2003). The best-response models choose the alternative with the highest score or utility, which is found by the appropriate mechanism for the dimension at hand. It could be a regression equation, a discrete choice model, an optimisation algorithm or ranking system.

Table 2 summarises the current preferred configuration of MATSim-T. The toolkit allows a wide variety of other configurations, some of which have been reported in previous papers. See below for a list of papers reporting MATSim-T tools and configurations. It is obvious that MATSim-T engages only with some of the dimensions, but those which transport modelling has traditionally addressed. It leaves out the social aspects of travel, i.e. the group with which the person travels or undertakes the activity, and it ignores how the expenditure is divided among those benefiting from it. Finally, it avoids the challenges of multi-modal connection/route choice (See Bovy and Hoogedoorn, 2005 for a possible approach), but especially the challenges inherent in modelling the choice of parking (See Axhausen, 1989 and 1990 for an early attempt).

The activity chains (number, type and sequence) are drawn conditional on the socio-demographics of the agents from the observed distribution in the population. This limits the ability of the configuration to capture changes in these aspects, which was the special target of...
the work of Bowman and the work in his wake (Bowman, 2005; Shiftan and Suhrbier, 2005). An IPF-based approach (Hettinger, 2007) allows us to change the frequencies of the chains given new exogenously given marginal distributions of chain lengths and activity purposes.

Classifying the observed chains not only by length and type of activities, but also by their duration provides an initial set of activity starting times (and durations). The adaption of the starting times and durations in the iterative evolutionary scheme is guided by a scoring function (utility function), which is used across all modules of the system (See below). The planomat (Meister, Balmer, Axhausen and Nagel, 2006; Meister, Frick and Axhausen, 2005) module implements a genetic-algorithm based search for the optimal set of starting times and durations for all activities in the chain simultaneously, which is globally aware of the travel times between the selected destinations of the chain. This best-response model has been shown to speed up the evolutionary search by reducing the number of iterations. An alternative completely random perturbation of the starting times and duration is slower overall and is unable to find better solutions that are radically different from the agent’s current solutions.

The choice of the non-work and non-education locations is based on a simplified approach. Based on the home location or the home and work location the choice set is selected as all relevant facilities within a radius around home and work. The radius depends on the population size of the municipality to which home and work belong. The random selection among the facilities is proportional to their capacity and inversely proportional to the distance from home or work. If no facility can be found within the radius, it is iteratively increased until one is found. The size and types of the facilities are obtained from the census of workplaces (BfS, 2001).

Mode choice between car driver, car passenger, walking, cycling and public transport is currently undertaken across the full range of options only once. The model randomly chooses the mode for the chain based on a suitably estimated multinomial logit model (MNL) (Ciari, Balmer and Axhausen, 2007 report the chain based model). A subtour is any sequence of ac-
activities which starts and ends at the same location. For example, the chain home – work – shop – work – leisure – home (where both work activities are performed at the same location) contains two subtours: home-work-leisure-home and work-shop-work. Ciari, Balmer and Axhausen, 2008 report models at this level of resolution to be implemented in the near future. Next to the socio-demographics and the ownership of the mobility tools the MNL is based on the distances involved. The parameters were estimated with the data of the Swiss national travel diary survey 2005 (BfS, 2006).

Depending on the traffic flow simulation chosen (see below), the learning process allows the choice between car travel and public transport travel by including a plan with only public transport used for the relevant trips. The public transport travel times can be either a simplified estimate based on the car travel times, or derived from a stop-to-stop travel time matrix obtained from an available public transport assignment model (or time-table information system).

Route choice is currently only performed for car-based travel. It is a best-response model returning the time-of-day dependent (in 15 min-intervals) least-cost-path between the starting link and the destination link, as origin and destination are coded at this level. The spatial resolution, as discussed above, is determined by the level of network detail. The algorithm is a landmark A* implementation of the Dijkstra algorithm, which performed best in a series of tests with the high resolution networks currently used (Lefebvre and Balmer, 2007). Public transport users are moved in one step from their origin to their destination.

The results of calculations are stored in one incrementally expanding plans – file, which is defined by an XML schema (W3C, 2006) (See www.matsim.org for format specifications), which in turn defines the range of issues which can be addressed. The plans file contains the information about the person and its activities, respectively about the routes chosen between the activities. The file and the underlying schema impose the consistency of the results.
Table 2  Scheduling: Current preferred configuration in MATSim-T

<table>
<thead>
<tr>
<th>Task</th>
<th>Frequency per run</th>
<th>Model type</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number, sequence and type of activities</td>
<td>Once</td>
<td>Conditional probability</td>
<td>Hettinger, 2007</td>
</tr>
<tr>
<td>Start and duration of activities</td>
<td>Per iteration</td>
<td>Best response model (GA-based optimizer)</td>
<td>Meister et al., 2006</td>
</tr>
<tr>
<td>Composition of the group undertaking the activity</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Expenditure and its allocation among the participants</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Secondary location choice</td>
<td>Once</td>
<td>Imputed (Proportional to size and distance)</td>
<td></td>
</tr>
<tr>
<td>Mode/vehicle choice</td>
<td>Initial iteration</td>
<td>Imputed (Chain based MNL)</td>
<td>Ciari et al., 2007</td>
</tr>
<tr>
<td>Mode choice (car, public transport)¹</td>
<td>Per iteration</td>
<td>Imputed (MNL based on plan score)</td>
<td></td>
</tr>
<tr>
<td>Route choice</td>
<td>Per iteration</td>
<td>Best response model (A* landmark shortest path)</td>
<td>Lefebvre and Balmer, 2007</td>
</tr>
<tr>
<td>Point of egress from the vehicle</td>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>Composition of the group travelling together</td>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>Travel expenditure and its allocation among the travellers</td>
<td>---</td>
<td>---</td>
<td></td>
</tr>
</tbody>
</table>

¹ The public transport candidate plan is generated as part of the initial demand generation for all agents which have a car available for travel.

It should be stressed that the discussion so far has described the preferred configuration at this point. Past implementations employed other modules with different emphases. It is obviously possible and the intent of the design of the toolkit that further modules can be implemented and used to enrich the capabilities of the toolkit.
5 Traffic flow simulation

A variety of traffic flow simulations are available for MATSim-T, reflecting its long development path. Currently in use are two version of the queue-based simulation idea. The first version (a re-implemented version of Cetin, 2005) is time-step based and is implemented in Java, as the rest of the toolkit, and requires therefore no file I/O of the plans and events file (See below). The second, substantially faster alternative is an event-oriented queue-based simulation (Charypar, Nagel and Axhausen, 2007), but being implemented in C++ it requires file I/O, which nearly cancels out its speed advantage. The event-oriented version has been expanded to model the capacity effects of the effective green times of the approaches of signal controlled junctions.

The events-file produced by both implementations for analysis purposes contains the link-by-link travel times of all travellers. It is the basis, on which the best response modules (router, planomat) calculate the time-of-day travel times. It is also the main basis of the various visualisation tools for on-line and off-line tracing of the agents through the net (See examples on www.matsim.org).

6 Relaxation scheme

MATSim-T is currently primarily used to search for a steady-state approximating a dynamic Nash-equilibrium (Nash, 1951; Wardrop, 1958 for the restricted case of route choice; Peeta and Ziliaskopoulos, 2001 for a review of dynamic traffic assignment). The performance of an agent’s plan is scored at the end of each iteration and retained (Raney and Nagel, 2006). For a pre-determined share of the agents, normally 10% (Rickert and Nagel, 1999), new plans are generated by searching for new shortest-path or by optimizing the starting times and durations (Nagel and Barrett, 1997). For the remaining agents the next plan to be executed is selected
randomly proportional to the logit-transformed scores (Raney and Nagel, 2006). The number of plans stored by the agent is defined by the user, but at least 2 plans per agents are necessary for the relaxation process (3-6 plans per agent are useful). The number is limited by the core memory of the machine employed. The number of iterations is set a priori, but a deterministic stopping criterion could be used as well. It should be noted, that this scheme allows for random break-downs, when gridlock situations occur due to a high resolution network with many short (low storage) capacity links. The probability of such events is drastically reduced when mode choice is part of the learning process. The system recovers quickly from these events, but they slow down the overall search.

The scoring function (Charypar and Nagel, 2005) is currently based on two ideas: logarithmically decreasing marginal utility of the activity duration and a Vickrey (1969) inspired valuation of the timing of the activities\(^2\). The activity types have given optimal operating points, which ensure that in the optimum all types have the same marginal utility. The activities and the facilities at which they can be executed have given time-windows (see below) which allows the formulation of late-arrival penalties. As the whole day is scored, no special penalties for early arrival are required, as early arrivals, i.e. before the facility is open, is penalized by the opportunity cost of foregoing an activity elsewhere. The parameter set is therefore very parsimonious (travel time by mode, late arrival, operating points and logarithmically increasing activity utilities by type). The activity utilities are constrained to be larger then zero. In addition there is penalty, if the time-of-day of the first departure is before the time-of.-day of the last arrival back home. The parameters are informed by a review of the literature, but have not yet been empirically estimated. The scoring is undertaken in monetary units for simplicity and clarity. In spite of its simplicity, the realism of the results is surprising and the dynamics are consistent with expectations.

\(^2\) See Chaumet, Locher, Bruns, Imhof, Bernard and Axhausen (2007) for the review of the relevant models.
7 Computational performance

The computational performance is tested with a small scenario. It is relevant, as the processes in MATSim-T scale linearly with the number of agents or the number of out-of-home activities\textsuperscript{3}. The system is built to deal with $10^7$ agents, $10^6$ facilities, links and nodes each. The Greater Zürich scenario, as well as the Switzerland scenarios are built with geocoded data from the year 2000 census of population (agents, households, commuting matrices), the year 2000 census of workplaces (facilities by type and capacity) and the national travel survey for the years 2000 and 2005 (477 types of activity chains / 9429 types of activity chains with duration classes; we distinguish eight classes of agents by age and work status).

For the Greater Zürich scenario (Figure 1) we draw a 10\% random sample from those agents in the Switzerland scenario, whose routes cross the area delineated by a 30km radius circle around the centre of the Zürich at Bellevue Platz. This includes transit traffic through the country, which was generated with the relevant border survey data. Those crossing the study area or entering/leaving it are represented by agents with a plan including either the single observed trip (for transit traffic) or two trips (for e.g. commuters from outside Switzerland).

The network is an updated and corrected version of the Swiss National Transport model (Vrtic, Fröhlich and Axhausen, 2003). The available NavTeq or Teleatlas data were not employed for this application.

\textsuperscript{3} Charypar, Axhausen and Nagel, 2006 show that the parallel version of the event-oriented traffic flow simulation scales linearly up to 64 CPUs. The remaining steps scale linearly with the number of agents (initial demand, mode choice) or the number of activities (location choice, timing and durations, route choice). The route choice scales with the average number of links in a route and responds strongly to network resolution.
Figure 1  Example scenario: Greater Zurich area (30km circle around “Bellevue”).

The different routes show examples of agents included in the scenario.

The number of the agents and of the elements describing the networks and activity system are shown in Table 3.

The computing times for the Greater Zürich scenario using the preferred MATSim configuration described above addressing the start times and durations, the routes and car/public transport mode choice during the iterations are shown in Figure 2 and Table 4. The calculations were performed on a dual-core AMD Opteron machine with 2.6 GHz CPU and 8 GB RAM; the time-step traffic flow simulation was run on one CPU, while replanning (router, planomat) was spread of two threads. The scheduling modules are fast: the planomat (timings) because of its efficiency in finding the optimal solution within a few iterations, which balances the relative low speed per agent. The router is a fast implementation. The time-step based traffic flow simulation needs the largest share of the computing time here, but this can be reduced.
through parallelisation. In addition, for larger scenarios we can switch to the event-oriented traffic flow simulation, which is substantially faster. It is currently being reimplemented in Java to avoid costly file I/O.

Table 3  Example scenario: Size

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Size</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Links</td>
<td>60'492</td>
<td>Directed links</td>
</tr>
<tr>
<td>Nodes</td>
<td>24’180</td>
<td></td>
</tr>
<tr>
<td>Agents within the study area</td>
<td>181’693</td>
<td>Containing 10% of all agents performing at least one activity in the area</td>
</tr>
<tr>
<td>Households within the study area</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average number of trips</td>
<td>3.1</td>
<td>Generated by agents residing with the study area</td>
</tr>
<tr>
<td>Trips (agents) crossing the study area</td>
<td>5’791</td>
<td>Agent = trip; linked to 880 home facilities outside Switzerland</td>
</tr>
<tr>
<td>Number of modes/activity types</td>
<td>5/17</td>
<td>Car driver, car passenger, bicycle, walk, public transport</td>
</tr>
<tr>
<td></td>
<td></td>
<td>work_sector2, work_sector3; kindergarten;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>primary school; secondary school; higher education; other education;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>retail_lt100sqm; retail_get100sqm; retail_get400sqm, retail_get1000sqm,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>retail_gt2500sqm, other retails; culture; restaurant et al.; sports.</td>
</tr>
<tr>
<td>Number of homes (facilities)</td>
<td>1’313’337</td>
<td></td>
</tr>
<tr>
<td>Number of out-of-home activity facilities</td>
<td>382’979</td>
<td></td>
</tr>
</tbody>
</table>

For a 100% Switzerland scenario we expect a factor 50 larger computing times, as the Greater Zürich scenario covers one fourth of the population. Assuming that a conservatively estimated speed up of 20 is possible with a 32 core machine the 100 iterations could be run within 2 ½ days. This leaves space for an increasing complexity of the behavioural models or network resolution, as the number of iterations can be further reduced through these more sophisti-
icated models (See the dramatic gain in the score after the first run of the planomat – GA tool optimising the starting times and durations of the activities).

Table 4  Example scenario: Computing times

<table>
<thead>
<tr>
<th>Processing step</th>
<th>Total for scenario</th>
<th>Unit Speed</th>
<th>Frequency per run</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>[h]</td>
<td>[Units/sec]</td>
<td></td>
</tr>
<tr>
<td>Generating agents (CH)</td>
<td>0.12</td>
<td>Agent 20’000</td>
<td>Once</td>
</tr>
<tr>
<td>Scheduling (Fixed components) (CH)</td>
<td>14.40</td>
<td>Agent 140</td>
<td>Once</td>
</tr>
<tr>
<td>Filter 1.9 million agents (Zurich area)</td>
<td>0.30</td>
<td></td>
<td>Once</td>
</tr>
<tr>
<td>Filter 10% agents of Zurich area</td>
<td>0.10</td>
<td></td>
<td>Once</td>
</tr>
<tr>
<td>Scheduling (without routing)</td>
<td>0.04</td>
<td>Agent 100</td>
<td>Per iteration$^1$</td>
</tr>
<tr>
<td>Scheduling (routing)</td>
<td>0.00</td>
<td>Agent 1’000</td>
<td>Per iteration$^1$</td>
</tr>
<tr>
<td>Time-step traffic flow simulation</td>
<td>0.15</td>
<td>Agent 300</td>
<td>Per iteration</td>
</tr>
<tr>
<td>Learning/plan selection</td>
<td>0.00</td>
<td>Agent 250’000</td>
<td>Per iteration</td>
</tr>
<tr>
<td>Total for one iteration$^1$</td>
<td>0.22</td>
<td></td>
<td>Per iteration</td>
</tr>
<tr>
<td>Total for run (for 100 iterations)$^2$</td>
<td>23.00</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$^1$ Including various I/O operations; $^2$ without the initial demand generation

Configuration: Time-step traffic flow simulation (mobsim) runs on a single CPU, replanning (router, planomat – time optimisation) runs in two threads; Dual-core AMD Opteron 2.6 GHz, 8 GB RAM
Figure 2  Computing times by step for the Greater Zürich scenario with 188’000 agents

Time-step traffic flow simulation (mobsim) runs on a single CPU, replanning (router, planomat – timing optimisation) runs in two threads; Dual-core AMD Opteron 2.6 GHz, 8 GB RAM
Figure 3  Development of the average score for the Greater Zürich scenario

See Table 2 for the configuration of the MATSim-T with a time-step based traffic flow simulation

8  Validation tools

The question of the validity of the model for this application is not the issue of this paper. The tools shown below are shown to illustrate the range of possibilities available to MATSim-T and by extension to any agent-based model. The outputs of the model are the final plans file (i.e. the final schedule including the routes between the activity locations) and the final events file tracing all trips from node to node by time-of-day.

For the study area above and for Switzerland in general the traffic counts of recent years are available (by hour-of-day and day-of-the-year) (160 counters in the study area and 578 in
Switzerland have been attached to a link in the network; more will be added in the near future). Using the possibilities of the KML language for Google Earth an interface has been written to allow queries, such as those shown in Figure 4.

**Figure 4** Example comparison between simulated and counted traffic flows

![Graph showing traffic volume comparison](image)

**Link:** Zürich, Rosengartenstrasse, traffic flowing south.

The availability of home locations, origins and destinations and the routes for each agent chosen allows further analyses. Figure 5 and Figure 6 make use of this data. The first shows the origins and destinations of the trips crossing a particular link (by time of departure and arrival), while the second shows the home locations of these users. The second analysis would not be feasible with a traditional trip-based model, since for non-home-based trips the home location of the travellers cannot be known from the trip alone.
Figure 5  Origins and destinations of the agents crossing one link between 16:00 and 17:00.

Link: Zürich, Rosengartenstrasse, traffic flowing south.
Figure 6  Home locations of the agents crossing one link between 16:00 and 17:00

Link: Zürich, Rosengartenstrasse, traffic flowing south.

9  Outlook and further work

The computing times documented above show that MATSim-T is becoming a reasonable alternative to traditional aggregate four-step-approaches, especially if these implement a dynamic traffic assignment. But it is also clear, that for even larger scale scenarios the computational performance has to be improved further. Our emphasis will be on the reduction of the number of iterations required, as well as on the further improvement of the route choice calculations.
The current set of modules is not very behaviourally sophisticated in their approach to scheduling. We are treating most dimensions in isolation. While the relaxation system assures that the final solution incorporates the interactions in the proper way, it is behaviourally and conceptually unsatisfactory. The easy next steps are currently in the implementation phase: the sub-tour based mode choice will improve the realism of the chains. The current secondary location choice is surprisingly robust in spite of the primitive utility function employed. The availability of the detailed facility information (i.e. individual stores, restaurants, work places, etc.) asks for a more complete description of the utility derived from a visit. This shift to a fuller description of the utility of the activity itself comes at the price of a higher computational cost and higher data requirements. The data requirements fortunately can be satisfied through the ready availability of commercial points-of-interest data bases.

Location choice, but also other planning questions require the full description of the monetary costs of the activity chain and its movements: tolls, parking, within-household car sharing, shared/rental car use, and electricity supply to plug-in hybrids. The provision of a coherent tool to describe the various price schedules for these costs is a high priority for MATSim-T and the other agent-based models. The current ad-hoc solutions are sub-optimal.

The integration of location and a behaviourally more sophisticated modal choice into the learning (relaxation) scheme is a second obvious step given the strong interaction between these dimensions and the timing choice already integrated. While in the interim it might be enough to include single-dimension tools into the learning loop, previous experience has shown that it is superior conceptually and computationally to integrate steps into an appropriate best-response (imputation) model. The extension of planomat with mode choice is currently under way. In this context, we will also estimate the parameters of the utility function fully; again the robustness of the system against true counts is surprising given the current derivation of the utility function.
The traffic flow simulation will need to be expanded into a multi-modal tool, allowing the agents to move between different vehicles/modes on a consistent network. This is of special interest for capturing the dynamics of public transport services.

Finally, the agent-based systems and MATSim-T will have to address the issue of agents for the supply-side, i.e. agents which provide or remove the capacities of the networks or of the facilities. While for computational reasons it is improbable that agent-based models of travel behaviour are a good basis for a complete land-use model system, they could address individual issues, which are especially linked to travel: network expansion, public transport networks and services, parking provision, pricing choices. Initial work along these lines is on-going in the MATSim-T context and we hope to be able to present results in the near future.

The system’s strict modularity is an open invitation to other research groups to integrate their approaches into the framework. The strong consistency checks imposed through the XML-file formats allows for an improved semantic consistency, e.g. for agent generation or the activity schedule formation and adaption.

10 Acknowledgements

The work reported here was made possible by the generous and gratefully acknowledged support of ETH Zürich Research Fund, Kanton Zürich, KTI Fonds, German Ministry for Education and Research and the VW Foundation.

We would like to thank all of colleagues and students, who have contributed to the development and testing of the system of the years.
11 Literature


### 12 MATSim papers and reports


