

## **An Agent Based Demand Modeling Framework for Large Scale Micro-Simulations**

2005-10-31

Michael Balmer  
Institute for Transport Planning and Systems (IVT)  
ETH Zurich  
Switzerland  
Phone: +41-1-633 27 80  
Fax: +41-1-633 10 57  
Email: balmer@ivt.baug.ethz.ch

Kay W. Axhausen  
Institute for Transport Planning and Systems (IVT)  
ETH Zurich  
Switzerland  
Phone: +41-1-633 39 43  
Fax: +41-1-633 10 57  
Email: axhausen@ivt.baug.ethz.ch

Kai Nagel  
Transport Systems Planning and Transport Telematics (VSP)  
TU-Berlin  
Germany  
Phone: +49-30-314-23308  
Fax: +49-30-314-26269  
Email: nagel@vsp.tu-berlin.de

Words: 7113  
Figures: 4 (= 1000 words)  
Total: 8113

Note: Version for CD-ROM. A reduced version (less than 7500 words) will be delivered if this paper is accepted for publication.

**ABSTRACT**

Micro-simulation tools are becoming increasingly important in traffic demand modeling. The major advantage over traditional 4-step models lies in the fact that each traveler is simulated individually. This means, for example, that decision making processes can be included for each individual. The traffic demand is a result of different decisions made by each individual. Those decisions lead to plans which the individuals try to optimize. Therefore, such micro-simulation models need appropriate initial demand patterns for all given individuals.

The challenge is now to create individual demand patterns out of general input data. In practice there is a large variety of input data. They can differ in quality, spatial resolution, purpose, and so on. The challenge for a flexible demand modeling framework is to combine this variety to produce individual demand patterns. Even more, it has to define precise interfaces to provide portability to other models, programs and frameworks and it should be suitable for large-scale applications of many million individuals. Since the model has to be adaptable to the given input data, the framework needs to be easily extensible with new algorithms and models.

The demand modeling framework for large scale scenarios presented in this paper fulfils all these requirements. By modeling the demand for two different scenarios (Zurich—Switzerland and the Länder Berlin and Brandenburg—Germany) we show the flexibility of the framework in aspects of diverse input data, interfaces to third party products, spatial resolution and—last but not least—the modeling process itself.

## MOTIVATION

Micro-simulation is becoming increasingly important in traffic simulation, traffic analysis, and traffic forecasting (1, 2, 3). Some advantages over conventional models are:

- Computational savings in the calculation and storage of large multidimensional probability arrays.
- Larger range of output options, from overall statistics to information about each synthetic traveler in the simulation.
- Explicit modeling of the decision making processes of the individuals.

The last point is important since it is not a vehicle which produces traffic; it is the person who drives it. Persons do not just produce traffic; instead each of them tries to manage his day (week, life) in a profitable way. They go to work to gain money, they go hiking for their health and pleasure, they visit their relatives for pleasure or because they feel obliged to do so, they shop to cook a nice dinner at home, and so on. Since not all of this can be done at the same location they have to travel, which produces traffic. To plan an efficient day, many decisions are made by each person:

- Which route should I take to get to work? - *Route choice decision*
- Which mode should I use to go to the lake? - *Mode choice decision*
- Should I drink another beer before going home? - *Activity duration choice decision*
- Should I go shopping near my home or at the mall? - *Location choice decision*
- When should I do sports today? - *Activity starting time choice decision*
- Should I go to visit my friend? - *Activity type choice decision*
- Whom should I take along? - *Group composition decision*
- Should I go swimming before or after work? - *Activity sequence decision*

There are more decisions to make; some of them are made hours (days, months) in advance while others are made as spontaneous reactions to specific circumstances. Many decisions induce other decisions. For example, if I am late for work, I am supposed to work longer, so there is no time left to go shopping today, so I need some time tomorrow to do the shopping. This example shows the importance of describing schedules for each individual in a simulation model, because it is the schedule and the decisions made by the person who adhere to this schedule that produces traffic.

To simulate a typical day in an urban area, micro-simulation tools need information about the schedule of each individual and some knowledge about people's decision making process. The challenge is to create this individual demand out of general input data. In practice there is a large variety of input data. They can differ in quality, spatial resolution, purpose, and so on. The challenge for a flexible demand modeling framework is to combine this variety to produce individual schedules. Even more, it has to define precise interfaces to provide portability to other models, programs and frameworks and it should be suitable for large scale scenarios including many million individuals. Since the model should be adaptable to the given input data, the framework needs to be easily extensible with new packages, algorithms and models.

This paper presents such a modeling framework for large scale scenarios. After a summary about other work and describing the program structure, the framework is used to model daily demand for two different scenarios: One—medium resolution scenario—takes place in the greater Zurich area (Kanton Zurich—Switzerland) and consists of about 1.3 million individuals. The second—large scale scenario—is defined for Berlin and Brandenburg of Germany for about 7 million inhabitants. The two scenarios differ in the amount of available information, in spatial resolution, in size and—last but not least—also in the demand modeling process itself. We then give a short overview about the further use of the generated daily demand. For that, an iterative, large scale micro-simulation model is used (4). The paper finishes with a summary and outlook to future work.

## OTHER WORK

The work presented in this paper falls into the area of activity-based demand generation (ABDG). There exist a fair number of ABDG packages, e.g. (1, 2, 3, 5, 6, 7). Internally, those packages are structurally comparable to what is described here, and in terms of methods they are in fact considerably more sophisticated. Yet, their output is typically in terms of (time-dependent) origin-destination matrices, to be fed into dynamic traffic assignment (DTA) models, e.g. (8, 9, 10).

An important exception is TRANSIMS (11), which generates individual activity plans as input to the DTA. TRANSIMS was difficult to obtain outside the U.S. for a couple of years, thus spawning MATSIM (4) as an alternative. In the meantime, TRANSIMS has become open source (tmip.fhwa.dot.gov/transims), but MATSIM now goes beyond TRANSIMS in the following aspects:

- TRANSIMS uses “flat” file formats between modules, while MATSIM uses more powerful hierarchical XML formats (12).

- With the XML format, it is possible to do all information exchange between modules always with the same file format and the same DTD, varying only the level of detail of the included information (12). This means that arbitrary combinations of partial ABDG modules can be used.
- MATSIM generically keeps track of multiple plans per agent (12); for TRANSIMS, this would need considerable changes in the implementation.
- The MATSIM traffic flow simulation, while simplified when compared to the TRANSIMS traffic flow simulation, runs considerably faster, thus allowing meaningful runs within days instead of weeks (13).
- In contrast to TRANSIMS, the MATSIMS DTA process keeps track of the consistency of activity chains along the time axis – travelers have to spend minimum amounts of time at activities before they can proceed.

The present paper follows TRANSIMS by being fully traveler (agent) oriented on all levels, and it follows the MATSIM standards in terms of the structuring and formatting of the information. In contrast to TRANSIMS, the current paper is flexible and universal in terms of the data input requirements, and tests this with data available in Switzerland in and Germany. All other ABDG packages besides TRANSIMS would not only have to overcome the data requirements problem, but to our knowledge would also need source code change to allow for traveler-based output.

In this situation, it was decided to proceed by designing and implementing placeholder modules that do something plausible based on the available input data, until better modules with flexible enough data input and traveler-oriented data output become available. Those better approaches could be in the form of modified versions of the above-mentioned or other packages, or in the form of enhancements of our own code.

Other approaches are the agent-based land-use models, for example URBANSIM (14), ILUTE (15), or the models by Abraham/Hunt (16). These models face similar difficulties as the approach described in our paper, in particular the need to assemble a consistent agent-based view of the world from diverse data sources (17). In the longer run, it would be useful to have a “plug-and-play” approach between these different models, i.e. where modules between different modeling systems can be interchanged and coupled in arbitrary ways. In fact, at least one such effort is under way (18). However, the software engineering challenges for such a project, despite considerable progress in the last decade, are still enormous, in particular for large scale scenarios (19). Therefore, for the time being it seems more productive to make progress in smaller steps, which in our case means the implementation of a general framework, demonstrating the overall feasibility and highlighting potential problems.

## FRAMEWORK

The demand modeling process for each individual is highly dependent on the available input data. The more precise the data is, the fewer methods or algorithms are needed to synthesize parts of the demand of an individual. For example, if we had the complete information about what people are doing, it would be a simple conversion to describe it in an appropriate data format. Yet in practice, we do not have access to all information because of data privacy, imprecise or aggregated data, costs of the required surveys, and limitation of census data.

Therefore, a demand modeling framework (see FIGURE 1) has to be flexible enough to handle this variety. In other words, it has to provide interfaces for the different data types in appropriate data formats; it has to allocate an appropriate data structure in memory such that each data point can be accessed in a feasible amount of time and it needs to ensure the uniqueness of data points while creating, manipulating and deleting them.

### General Framework Structure

The framework consists of necessary modules, optional modules, and a special *world* package (see FIGURE 1). The necessary packages provide general XML parsing and XML writing interfaces, global constants and a global and unique data structure for accessing input parameters. Optional packages are based on a specified input written in XML format (4). The base functionality of each optional package is to read the defined XML data, store them in an appropriate data structure and write it again enriched, reduced or even unchanged in the XML data format.

### The Framework Packages in Details

#### *The parser, writer and gbl Package*

The *parser* package provides a general base class for parsing XML data via the SAX parser (20). The *writer* package is the base for writing XML files. The purpose of the *gbl* package is to hold global constants and globally accessible functions.

### *The config Package*

In the whole framework exactly one configuration data structure exists which holds all required input parameter for a specific demand modeling process. Typical parameters are locations of input data, different file formats, special function parameters, etc. The information stored in this package can be accessed from every part of the framework (global singleton design pattern) (21).

### *The world Package*

The *world* package has a special functionality. It describes the region for which the demand is to be modeled. Therefore, it is required to guarantee that only one world exists during the demand modeling process. This package holds all information about cells, blocks, zones, municipalities, etc. which are modeled as primitive shapes. If the other optional packages refer to such shapes, they have to point to them in the *world*. With this, uniqueness of each shape object is guaranteed. Even more, it is also a control mechanism to verify if other input data is consistent with the given scenario.

Another, quite important functionality of the *world* package has to be mentioned. During the demand modeling process, the resolution of the world can differ. For example, the land-use information might be based on a raster of 100x100 meter cells, while the commuter matrix is based on traffic analysis zones. Therefore, it is necessary to (dis)aggregate one data set into another world resolution. The world holds this mapping for an arbitrary amount of resolutions of the same region. With it, on the fly (dis)aggregation is provided. The mapping between two resolution layers can be either generated by the framework according to a specified mapping rule or it can be given as predefined input data.

While this package plays a central role in the framework, it is not a required package. The reason for that is that there can be demand modeling steps for which spatial information is not needed (i.e. generating departure times of activities, assigning activity chains to individuals, etc.).

### *The landuse and commuters Package*

Those optional packages store information for the demand modeling process. Dependent on the input data those packages are used i.e. to find locations of work activities, to create sets of leisure facilities, and so on. The framework allows one to add other optional packages in the same manner as the *landuse* and *commuters* package to add new data sources.

### *The plans Package*

This is—in principle—the core of the demand modeling framework. The data structure is based on the MATSIM schedule DTD (in MATSIM called a plans-DTD) which also can be accessed via the MATSIM homepage (4) and it is used as a “working file”. The reason for this choice lies in the high flexibility of the MATSIM schedule file. In the minimum version the file only holds the identity number (id) of all simulated persons. On the other hand, it is possible to add a large amount of information about each person, such as age, sex, car ownership, home and work location (at different world resolutions), etc.

Additionally each person can access one or many different individual schedules, describing when a person wants to start an activity, where it will be performed, which route and mode to take to go from one location to the next, and so on.

The internal data structure of the *plans* package provides exactly the same flexibility as the XML file format. Therefore, it is possible to add sequentially more and more schedule details to a given incomplete MATSIM plans file.

## **Demand Algorithms**

Algorithms can be added to each package to verify, manipulate, add or delete data items according to the purpose of the algorithm. Since for each new scenario different algorithms have to be used or implemented, it is of great interest that algorithms should be clearly separated from the data structure. They also should be easily exchangeable by others. The order in which algorithms are called should be flexible as well.

The algorithms are collected into a sub package of that data structure which they manipulate. Therefore, an algorithm at *org.matsimJ.commuters.algorithms* manipulates data in the *commuters* data structure (see below for examples).

## **External Models**

At any point in time during the demand modeling process, the framework allows one to store all data into well defined XML data files. They do—per definition—respect the format as described in the underlying DTDs. Therefore, a clean and stable interface to third party programs and models is available even when those models are not part of the demand modeling framework described here.

For the two case studies described in the following two sections, we will use the external Secondary Location Choice Module (22) in its adapted version described in (23), which chooses locations for secondary activities in such a way that each agent improves its daily activity chain. Primary and secondary activities are defined such that in each given activity chain, there are well defined primary activities while the others are defined as secondary activities (23). The module uses the same DTD, allowing simple information interchange.

### Computational Issues

One important issue for demand modeling is the amount of input information needed. Because of the variety of possible demand modeling algorithms and input sources it is necessary to have fast access to that data. One simple and very fast solution is to load all information into memory and to provide a hierarchical data structure (tree structure) to access any item from any other location in  $O(2 * \log(n))$ , where  $n$  is the depth of the data tree.

The hierarchical data structure is already provided by the input data (XML format), but the available space in the memory might not be sufficient. While the description of the world, the land-use data information, commuter matrices, etc. typically hold a relatively small amount of data, the amount of information for the individuals goes far beyond the size of a typical memory capacity (on average around 1-2 GB of memory).

The demand modeling framework therefore uses the idea of sequential individual demand generation (streaming of individuals). In other words, the framework reads one individual at a time, runs the defined algorithms for it, writes the results to file and frees the memory. In this way, the number of individuals of the given scenario is unlimited. This idea will still work if—instead of single persons—we will model demand at the household level for a small number of persons. But the limit of this approach will be reached if we also want to add the concept of social networks (24). In this case demand modeling of one individual can—in principle—depend on all other individuals in the scenario and therefore the whole population must be stored in memory.

Nevertheless, the *plans* package still allows one to store all individuals in memory, if the amount of data is not too large or if one has access to machines with sufficient amount of memory. The user of the framework can switch between “streaming” and “no-streaming” by setting a defined parameter flag.

The concept of “streaming individuals” has one conceptual disadvantage: Some demand modeling processes need summary information about a given population, which can not be calculated before the last individual is parsed. If the whole population does not fit simultaneously into computer memory, there is no other choice than to parse the population twice; first, to obtain aggregated information, second, to distribute it again to each individual in the desired way. If a problem is “doubly constrained” and an iterative procedure is necessary, this process may have to be repeated several times until convergence is reached. This is a typical example of a “trade-off” between computational speed and amount of memory.

### DEMAND MODELING FOR THE KANTON ZURICH, SWITZERLAND

Using the framework described in the previous section we present the steps taken to model daily demand for the Kanton Zurich, Switzerland. The following subsections describe briefly which input data were used and which algorithms were employed to model the daily demand.

Note that the algorithms used are indeed not very sophisticated. They are implemented to present the use of the framework rather than to deliver “state-of-the-art” demand modeling processes.

#### Data Resources

The *world* describes the region (Kanton Zurich of Switzerland) at two different resolutions:

- 170 municipalities and 12 additional districts inside the city of Zurich (municipality level denoted as  $M$ )
- Raster of 100x100 meter cell resolution (raster level denoted as  $R$ ). In total 167881 cells are given.

The mapping between those two levels is also available. The mapping rule is non-ambiguous. Each raster cell belongs to none or one municipality while each municipality holds at least one raster cell. Even more, the municipalities hold only those cells which include urban areas, e.g. “lake cells” are not part of the mapping. Information about population distribution in the municipalities is generated by (25). It holds about 1.3 million inhabitants with the following data:

- Home location ( $R$ )
- Population group (children, worker, non-worker, senior)
- Mobility (car, season ticket ownership, bike, walking)
- Age
- Sex

The Swiss Federal Statistical Office (26) provides land-use data holding information about capacities of different activity types like “work”, “shopping”, “education”, etc. That information is also based on raster level ( $R$ ).

The “Pendlermatrix 2000” (27) holds the information about work and education commuters at municipality level ( $M$ ).

The microcensus (28) is a periodic survey of the travel behavior of the Swiss population. It has been run every five years since 1974. The microcensus is carried out by the Federal Office for Spatial Development, ARE, in cooperation with the Swiss Federal Statistical Office, BFS, and contains detailed information about the mobility behavior of 30000 persons from all over Switzerland (29). About 1670 different activity chains can be found in Micocensus2000. Most of them appear very rarely, so only the 100 most frequently occurring activity chains are used later, which still cover more than 90% of all days. By reducing the number of activity types to “home”, “work”, “education”, “shop” and “leisure” some activity types like “service” were recoded to “work”. We end up with 21 different activity chains (23) and their distribution.

### Demand Modeling Process Steps

The demand modeling process is split into six sequential steps (see FIGURE 2 for an overview). Each process step (except the first one) uses one specific data resource to add details to each individual schedule:

1. *Conversion* This process step converts the input population file into the XML person description file. None of the person tags holds a plan yet. But additional attributes like age, sex, etc are included. It also holds the information in which land-use raster element ( $R$ ) this person lives.
2. *Mapping* Each raster element of the land-use data belongs exactly to one municipality. By using this mapping, each agent can be assigned to the municipality of his home location.
3. *Distribution* Given the distribution of the activity chains described above, we assign one of the chains to each person according to the given distribution. We also respect the fact that children do not go to work, therefore persons of young age are not allowed to hold an activity chain including a “work” activity. Note that this independent random sampling from aggregate distributions can cause a lack of consistency with the given distribution. To avoid this, it is needed to parse the population twice to obtain aggregate information (see above).
4. *Distribution* The “Pendlermatrix 2000” (see above) holds the information about work and education commuters at municipality level ( $M$ ). With the assumption about primary activities given in (23) we are able to add the locations of the primary activities “work” and “education”. Unfortunately we do not have similar data for the primary activities “shop” and “leisure”. As long as no better data or no appropriate algorithm is available we assume that those activities are undertaken in the same municipality (district) where the person lives.
5. *Disaggregation* The process step 6 uses the location choice (see FIGURE 2) which is working at the raster level of detail ( $R$ ). Therefore, we need to disaggregate the locations of the secondary activities to that level. It is done by uniformly picking one of the raster elements of the given municipality (district).
6. *Secondary activity location choice (extern)* The module of Marchal and Nagel (22) is used as an example of an external program to add the missing locations of the secondary activities.

The final schedule of each person in the scenario now describes the following:

- Person attributes
- Which activities will be performed
- In which order the activities will be performed
- Where the activities will be performed

There are some final algorithmic steps missing in that modeling process: There is no information about the times when an activity should be performed and how long it should take. We also still do not know which mode of transportation will be used and which route to take to go from one location to the next. These steps are left out by design because we will use the MATSIM iterative optimization process (12) to create that information. A short overview about this process will be given below.

### DEMAND MODELING FOR THE LÄNDER BERLIN AND BRANDENBURG, GERMANY

The scenario Berlin/Brandenburg does have some differences compared with the Kanton Zurich scenario, due to the fact that only other, less precise information is available.

Note that we again do not use sophisticated algorithms. It is more of interest to show that some algorithms of the Zurich scenario can be reused while others have to be added.

#### Data Resources

The *world* describes the region of Berlin and Brandenburg (30). In this scenario tree different spatial resolutions are used:

- The region is split up into 1008 traffic analysis zones TAZ (zone level denoted as  $Z$ )
- Each zone consists of several Blocks. 12260 blocks are available for this scenario (block level denoted as  $B$ )

- A raster of 169420 cells is also available. Each cell has an extent of 500x500 meters (raster level denoted as *R*)

The mapping rules are the following: Each traffic analysis zone consists of at least one block. Each block belongs exactly to one traffic analysis zone. A cell refers to none, one or many blocks while each block refers to at least one cell.

The raster is synthetically generated and therefore, no additional information is encoded. The creation of the raster and the mapping is also done by the above described framework using a special *world* algorithm in the *org.matsimJ.world.algorithms* package.

Only the population of Berlin itself was available at that time consisting of about 3.3 million individuals (30). The details were not as extensive as for the population of Kanton Zurich. The only available items are:

- Home locations (*B*)
- Age

The land-use of Berlin and Brandenburg is based on traffic analysis zone level *Z*. It consists of capacities for “home”, “work”, “education”, “shopping” and “leisure” activities.

For this scenario, there is no commuter matrix (and also no production-attraction matrix) available; only a 24 hour origin-destination matrix is given (31). We use this matrix in two different ways. First, for all trips starting in Berlin we use this information to guess where individuals of the corresponding zone are performing their primary activity. In other words, we use it in a similar way as the commuter matrix of the Kanton Zurich scenario. The remaining trips starting in the traffic analysis zones of Brandenburg are used to create the missing part of the population and also to define the location of their primary activity. This is—of course—just a very approximate way to interpret an origin-destination matrix, but still a step to close the information gap. It will be one of the highest priorities of our future work to design a better primary activity location choice method for the Berlin/Brandenburg scenario.

We use the same activity pattern distribution as in the Kanton Zurich scenario. Those patterns reflect average Swiss behavior and thus not necessarily those of Berlin/Brandenburg, but it is a feasible start as long as no other data is available.

### Demand Modeling Process Steps

In comparison with the process of the Kanton Zurich scenario there is one step more because three different levels of spatial resolution are used instead of two. Further more, three additional steps are included (see FIGURE 3):

- A. *Creation* As already mentioned above, there is no information about the population of Brandenburg. Because of the fact that the number of trips of the origin-destination matrix starting in zones of Brandenburg fits with its total population, we assigned for each of those trips one individual to that start zone. This is—of course—a rough approximation of the distribution of the population of Brandenburg.
  - B. *Disaggregation* Since the individuals are placed at traffic analysis zone level (*Z*) we need to map them to block level (*B*) we assume a uniform distribution over all blocks of a given zone.
  - C. *Land-use disaggregation* To use the secondary activity location choice module (22) the land-use information has to be given at the raster level. Therefore, the TAZ based land-use data also has to be disaggregated into raster level (*R*). It is done in a first step by uniform distribution over all blocks (*B*) of each TAZ. As a second step another uniform distribution is used to propagate the land-use data from block level (*B*) to raster level (*R*).
1. *Conversion* This is the same conversion as already described above. Together with the two special steps A and B the whole population is now given in the appropriate MATSIM XML plans format (4).
  2. *Mapping* To add the corresponding TAZ level to the population of Berlin a mapping from block level to TAZ level has to be done. This mapping is non-ambiguous.
  3. *Disaggregation* The last remaining spatial level (raster level *R*) is mapped in this step. The method is again an additive uniform distribution over each cell of the given block.
  4. *Distribution* Same as for the Zurich scenario
  5. *Distribution* In contrast to the same step of the Kanton Zurich process we assign all different primary activity types according to the given matrix. Because we do not have any information about the purpose of a given trip of the matrix we cannot distinguish it from others. The location of the primary activity is chosen according to the distribution of the destinations of those trips starting at the home location of the current individual.
  6. *Disaggregation* In this step we again need to disaggregate the location of the secondary activity to raster level (*R*). It is done in the same way as described for the land-use (see special process steps).

7. *Secondary activity location choice (extern)* The secondary location choice module by Marchal and Nagel (22) is again used to find the missing locations in the same way as described in process step 6 of the Kanton Zurich scenario.

The final demand of that scenario again holds the same data items as described in the Zurich scenario.

### MATSIM ITERATIVE OPTIMIZATION PROCESS

As already mentioned, we use the generated demand as input for the MATSIM iterative optimization process. In the work of Raney and Nagel (12) is shown that feasible timing of given schedules can be generated by starting with random initial departure times and durations of activities. Therefore, it is not necessary to include an appropriate algorithm into the demand modeling framework but we can use MATSIM as a final external module for that.

In contrast to (12) we use an enhanced time allocation module, called *planomat* (32). Using the concept of genetic algorithms this module can be used for time scheduling as well as for location or mode choice. Since activity chain choice and location choice is already done by the demand modeling framework and the traffic flow simulation of MATSIM (13) handles only the individual transport mode, the functionalities of the *planomat* module is reduced to time choice only.

MATSIM optimizes (daily) schedules and not single trips. Therefore, the consistency at the individual level is guaranteed (an agent cannot leave a location before he arrives at it). This is—in fact—one of the most important issues of describing demand on the basis of individuals instead of losing important information by using origin-destination matrices.

### Time Scheduling Results

Using the same MATSIM setup as described in (23) we run a one percent sample of the population of the Kanton Zurich scenario. By an artificial reduction of the capacities of each link of the given network (13) by a factor of 50 we can produce similar congestion patterns as we would by using all persons. Because of the small sample size, the simulations display considerable traffic pattern fluctuations from one iteration to the next; using those simulations to identify bottlenecks would be difficult. Nevertheless, as will be explained in the following, the aggregated time structure of the results is rather plausible.

FIGURE 4 shows the differences between the initial departure time schedule (iteration 0) and the departure time schedule after 400 iterations. The figure highlights (bold lines) two groups of individuals:

- Bold grey line: Individuals with schedules containing “work” as the second and second to last activity in the chain
- Bold black line: Individuals with schedules containing “leisure” as the second and second to last activity in the chain

Since the “work” activities should be performed between 7am and 6pm and the duration of performing “work” should be around 8 hours (in Charypar and Nagel (33) described as the “operating point”  $t^*$ ), they produce the expected morning and evening peak (rush hour) after 400 iterations.

The leisure activities do not have such a hard constraint (the setup defines it between 6am and midnight). The “operating point”  $t^*$  of performing one leisure activity is set to 2 hours. Therefore, there is much more flexibility of time scheduling for leisure activities. FIGURE 4 reflects that fact. While the time choice at iteration 0 is unsatisfying, the improvement after 400 iterations is immense. Only a few persons choose the morning rush hour to travel while the major part performs their leisure activity in the evening (finishing their last leisure activity almost before midnight, when almost no traffic exists).

### DISCUSSION AND OUTLOOK

A 1% population sample was used to reduce needed computational resources. Running 10% samples, however, is not a major challenge, and even running full populations is possible – the necessary resources grow considerably more slowly than the sample size. We will report on such results in the future. Even with full populations, the effect of fluctuations from one iteration to the next is considerable. However, so far our average per-link error has always been considerably larger than iteration-to-iteration fluctuations – indicating that at this point fluctuations are not the most important problem. More discussion of these issues unfortunately goes beyond the scope of the current paper.

The *planomat* MATSIM module mentioned above provides an attractive amount of functionality. Since it provides—besides the already used time choice—also a location choice and a mode choice module it is of interest to actually use these functionalities. For that we need to provide this module with additional information. As a first step, we need to provide an alternative to individual transport mode in MATSIM. Therefore, the demand modeling framework will be extended such that it delivers information about travel times from one zone (block, district, TAZ or municipality) to another. With it, the *planomat* module decides which mode will be used

for each person in the population. In simple words, we expect the modal split to be generated as part of the process.

As a second step, we will add another package to the framework which contains data about activity spaces, catchment areas and commuter sheds (34). With it, we can replace primary activity location choice (commuter package) and also secondary activity location choice (external module) by adding a set of possible locations for all available activity types to each person of the population. This set will then be used by the *planomat* module to find iteratively—in the MATSIM framework—an appropriate location for the given activities of each individual. As already mentioned above, it is a quite simple task to extend the demand modeling framework with the idea of intra-household interaction. Externally, the main change would be to group household members into corresponding xml brackets in the plans file. All modules would function as before at the person level, and household interactions could be introduced module by module. One of the first changes would be to ensure that the plans selection method (in the MATSIM-DTA) keeps the household plans consistent, for example regarding arrangements about time-shared car use. This would entail to make changes from one plan to another during iterations no longer on the person but on the household level.

## SUMMARY

This paper presented an agent based demand modeling framework for large scale micro-simulations. Using the framework for the Zurich and the Berlin/Brandenburg scenario it is shown that

- it is flexible enough to handle a variety of input data (Zurich data differ from Berlin/Brandenburg data),
- it is flexible enough to extend or replace algorithms (different algorithms for Berlin/Brandenburg commuter data than for Zurich commuters),
- it provides (dis)aggregation to different spatial resolutions (two levels of resolution for Zurich scenario, three levels for Berlin/Brandenburg scenario),
- it provides a robust interface to third party models, programs and frameworks (the external modules: secondary activity location choice, and MATSIM dynamic traffic assignment),
- it is suitable for unlimited number of individuals (about 7 million persons for the Berlin/Brandenburg scenario), and
- it is easy to extend, to replace algorithms with more enhanced ones, to add new algorithms for existing packages and to add new packages to handle new input data.

Nevertheless, it is only a framework. The algorithms presented for modeling are very simple. Resources are needed to enhance those algorithms and to validate the resulting demand against behavioral issues.

This work also shows the importance of interaction between the transportation community and computer scientists. To satisfy the requirements concerning data management, data processing, computational design and implementation, runtime issues, etc., it is necessary to include computer knowledge into the transportation research process.

Last but not least we hope that this work encourages researchers with some basic programming skills to use, enhance and extend the framework according to their own desires and wishes.

## REFERENCES

1. Vovsha, P., E. Petersen, and R. Donnelly. Micro-simulation in Travel Demand Modeling: Lessons Learned from the New York Best Practice Model. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1805*, TRB, National Research Council, Washington, D.C., 2002, pp. 68-77.
2. Bowman, J. L., M. Bradley, Y. Shiftan, T. K. Lawton, and M. Ben-Akiva. Demonstration of an activity-based model for Portland. In *World Transport Research: Selected Proceedings of the 8th World Conference on Transport Research 1998, No. 3*, Elsevier, Oxford, 1999, pp. 171-184.
3. Bhat, C. R., J. Y. Guo, S. Srinivasan, and A. Sivakumar. A Comprehensive Econometric Microsimulator for Daily Activity-Travel Patterns. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1894*, TRB, National Research Council, Washington, D.C., 2004, pp. 57-66.
4. MATSIM, Multi Agent Traffic SIMulation. <http://www.matsim.org>. Accessed October 2005.
5. VISEM, Planung Transport und Verkehr, PTV. [http://www.ptv.de/cgi-bin/traffic/traf\\_visem.pl](http://www.ptv.de/cgi-bin/traffic/traf_visem.pl). Accessed October 2005.
6. Pendyala, R. M. Phased Implementation of a Multimodal Activity-Based Travel Demand Modeling System in Florida, Final Report. Research Center, Florida Department of Transportation, Tallahassee, 2004. [http://wwwprod.dot.state.fl.us/research-center/Completed\\_Proj/Summary\\_PTO/FDOT\\_BA496rpt.pdf](http://wwwprod.dot.state.fl.us/research-center/Completed_Proj/Summary_PTO/FDOT_BA496rpt.pdf). Accessed October 2005.
7. Arentze, T., F. Hofman, H. van Mourik, and H. Timmermans. ALBATROSS: Multiagent Rule-Based Model of Activity Pattern Decisions. In *Transportation Research Record, No.1706*, TRB, National Research Council, Washington, D.C., 2000, Paper No. 00-0022, pp. 136-144.
8. DynaMIT. <http://mit.edu/its/dynamit.html>. Accessed October 2005.
9. DYNASMART. <http://www.dynasmart.umd.edu>. Accessed October 2005.
10. Friedrich, M., I. Hofstätter, K. Nökel, and P. Vortisch. A dynamic traffic assignment method for planning and telematic applications. In *Proceedings of Seminar K, European Transport Conference*, Cambridge, GB, 2000.
11. TRANSIMS, TRansportation ANalysis and SIMulation System. Los Alamos National Laboratory, Los Alamos, NM. <http://transims.tsasa.lanl.gov>. Accessed October 2005.
12. Raney, B. and K. Nagel. An improved framework for large-scale multi-agent simulations of travel behavior. In *P. Rietveld, B. Jourquin and K. Westin: Towards better performing European Transportation Systems*, in press, 2005. <http://www.vsp.tu-berlin.de/publications>. Accessed October 2005.
13. Cetin, N. and K. Nagel. A large-scale agent-based traffic micro-simulation based on queue model. Conference paper, Swiss Transport Research Conference (STRC), Monte Verita, Ascona, Switzerland, 2005.
14. Waddell, P., A. Borning, M. Noth, N. Freier, M. Becke, and G. Ulfarsson. Microsimulation of Urban Development and Location Choices: Design and Implementation of UrbanSim. *Networks and Spatial Economics*, Vol. 3, No. 1, 2003, pp. 43-67.
15. Salvini, P. A., and E. J. Miller. ILUTE: An Operational Prototype of a Comprehensive Microsimulation Model of Urban Systems. *Networks and Spatial Economics*, Vol. 5, No. 2, 2005, pp. 217-234.
16. Hunt, J. D., R. Johnston, J. E. Abraham, C. J. Rodier, G. R. Garry, S. H. Putman, and T. de la Barra. Comparisons from Sacramento Model Test Bed. In *Transportation Research Record, No. 1780*, TRB, National Research Council, Washington, D.C., 2000, Paper No. 01-3000, pp. 53-63.
17. Abraham, J. E., T. Weidner, J. Gliebe, C. Willison and J.D. Hunt. Three Methods for Synthesizing Baseyear Built Form for use in Integrated Land Use-Transport Models. In *Transportation Research Record, No. TBD*, TRB, National Research Council, Washington, D.C., 2005.
18. Waddell, P., H. S.Úevčíková, D. Socha, E. Miller, and K. Nagel. OPUS: An International Collaboration to Develop an Open Platform for Urban Simulation. Presented at the 9th Conference on Computers in Urban Planning and Urban Management (CUPUM), University College London, UK, 2005.
19. Nagel, K. and F. Marchal. Computational Methods for Multi-agent Simulations of Travel Behavior. In *Proceedings of the meeting of the International Association for Travel Behavior Research (IATBR)*, Lucerne, Switzerland, 2003. <http://www.ivt.baug.ethz.ch/allgemein/iatbr2003.html>. Accessed October 2005.
20. SAX, Simple API for XML. <http://www.saxproject.org>. Accessed October 2005.
21. Gamma, E., R. Helm, R. Johnson, and J. Vlissides. *Design patterns: elements of reusable object-oriented software*. Addison-Wesley Professional Computing Series, 1995.
22. Marchal, M. and K. Nagel. Modeling location choice of secondary activities with a social network of cooperative agents. In *Transportation Research Record, No. TBD*, TRB, National Research Council, Washington, D.C., 2004.

23. Balmer, M., M. Rieser, A. Vogel, K.W. Axhausen and K.Nagel. Generating day plans using hourly origin-destination matrices, in *T. Bieger, C. Laesser and R. Maggi (eds.) Jahrbuch 2004/05 Schweizerische Verkehrswirtschaft*, SVWG, St. Gallen, Switzerland, 2005, pp5-36,.
24. Axhausen, K.W. Activity spaces, biographies, social networks and their welfare gains and externalities: Some hypotheses and empirical results. Paper for the PROCESSUS Colloquium, Toronto, 2005. [http://www.civ.utoronto.ca/sect/traeng/ilute/processus2005/PaperSession/Paper18\\_Axhausen\\_ActivitySpacesBiographies\\_CD\\_.pdf](http://www.civ.utoronto.ca/sect/traeng/ilute/processus2005/PaperSession/Paper18_Axhausen_ActivitySpacesBiographies_CD_.pdf). Accessed October 2005.
25. Frick, M. A. Generating Synthetic Populations using IPF and Monte Carlo Techniques: Some New Results. Conference Paper, 4th Swiss Transport Research Conference, Monte Verita, Ascona, Switzerland, 2004. <http://www.ivt.ethz.ch/vpl/publications/reports>. Accessed October 2005.
26. BFS, Swiss Federal Statistical Office. <http://www.bfs.admin.ch/bfs/portal/en/index.html>. Accessed October 2005.
27. Vrtic, M. and K.W. Axhausen. Experiment mit einem dynamischen Umlegungsverfahren. *Strassenverkehrstechnik*, Vol. 47, No. 3, 2003, pp 121-126.
28. ARE, Bundesamt für Raumentwicklung und Bundesamt für Statistik. *Mobilität in der Schweiz, Ergebnisse des Mikrozensus 2000 zum Verkehrsverhalten*, Bern und Neuchâtel, Switzerland, 2001. <http://www.are.admin.ch/are/de>. Accessed October 2005.
29. Axhausen, K. W. and M. Frick. Nutzungen, Strukturen, Verkehr. In *G. Steierwald, H.-D. Künne and W. Vogt (eds.) Stadtverkehrsplanung: Grundlagen, Methoden, Ziele*, Springer, Heidelberg, Germany, pp. 61-79.
30. RBS, Raumbezugssystem Berlin. *Bevoelkerungsverteilung, Teil-Verkehrszellen und Block Geometrie / Hierarchie*, Statistisches Landesamt Berlin, 2000. <http://www.statistik-berlin.de>. Accessed October 2005.
31. SfsB, Senatsverwaltung für Stadtentwicklung Berlin. *Datengrundlagen Stadtentwicklungsplan Verkehr*, 1998. <http://www.stadtentwicklung.berlin.de>. Accessed October 2005.
32. Meister, K. and M. Balmer. An improved replanning module for agent-based micro-simulations of travel behavior, Working Paper, Institute for Transport Planning and Systems (IVT), ETH Zurich, Zurich, Switzerland, 2005. <http://www.ivt.ethz.ch/vpl/publications/reports>. Accessed October 2005.
33. Charypar, D. and K. Nagel. Generating complete all-day activity plans with genetic algorithms. *Transportation*, Vol. 32, No. 4, 2005, pp. 369-397.
34. Axhausen, K.W., M. Botte and S. Schönfelder. *Systematic Measurement of Catchment Areas. CTPP 2000 Status Report*, U.S. Department of Transportation in cooperation with the TRB Census Subcommittee, 2004. <http://www.fhwa.dot.gov/ctpp/sr0804.pdf>. Accessed October 2005.

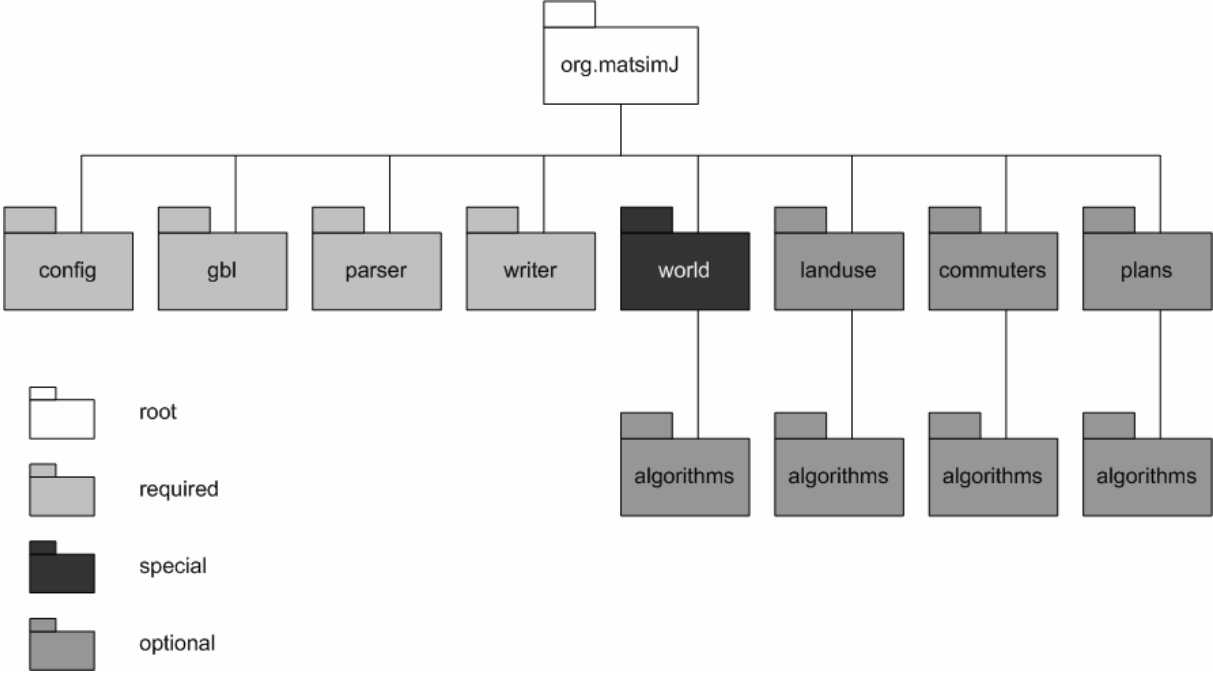
**LIST OF FIGURES**

FIGURE 1 Package structure of the demand modeling framework.

FIGURE 2 Process plan for modeling demand of Kanton Zurich.

FIGURE 3 Process plan for modeling demand of Berlin and Brandenburg.

FIGURE 4 Departure time histograms of a 1% sample population of Kanton Zurich.



**FIGURE 1** Package structure of the demand modeling framework.

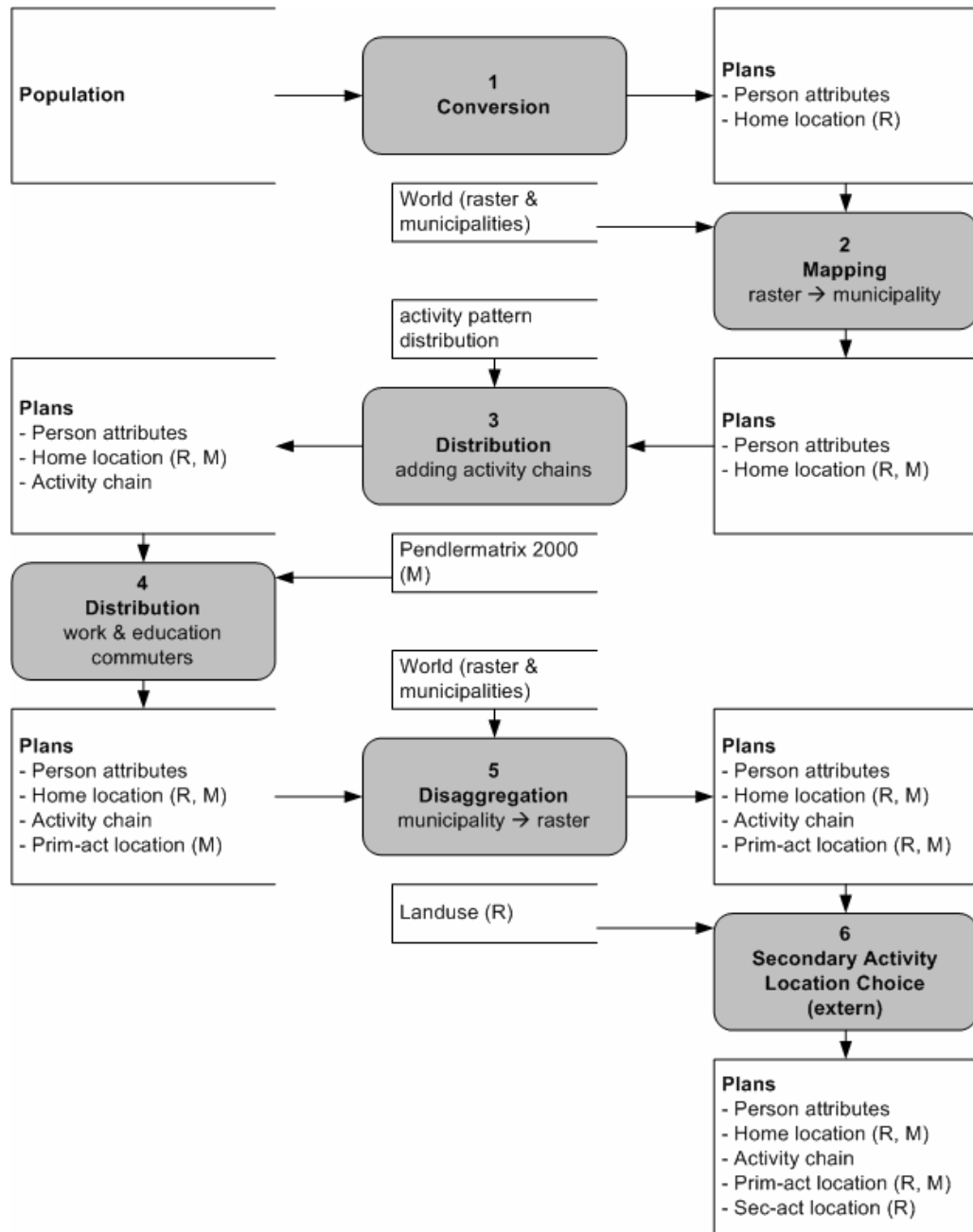


FIGURE 2 Process plan for modeling demand of Kanton Zurich.

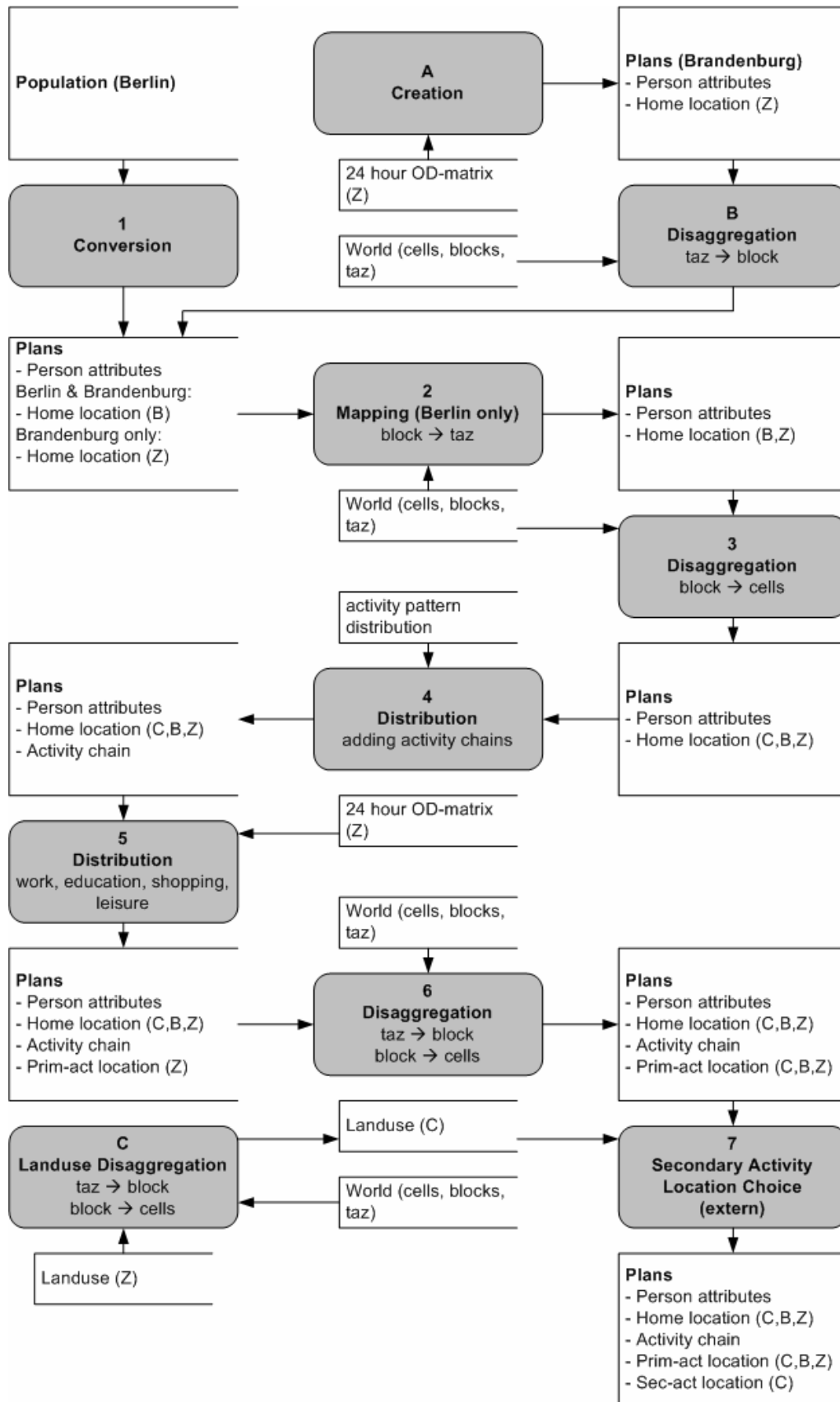


FIGURE 3 Process plan for modeling demand of Berlin and Brandenburg.

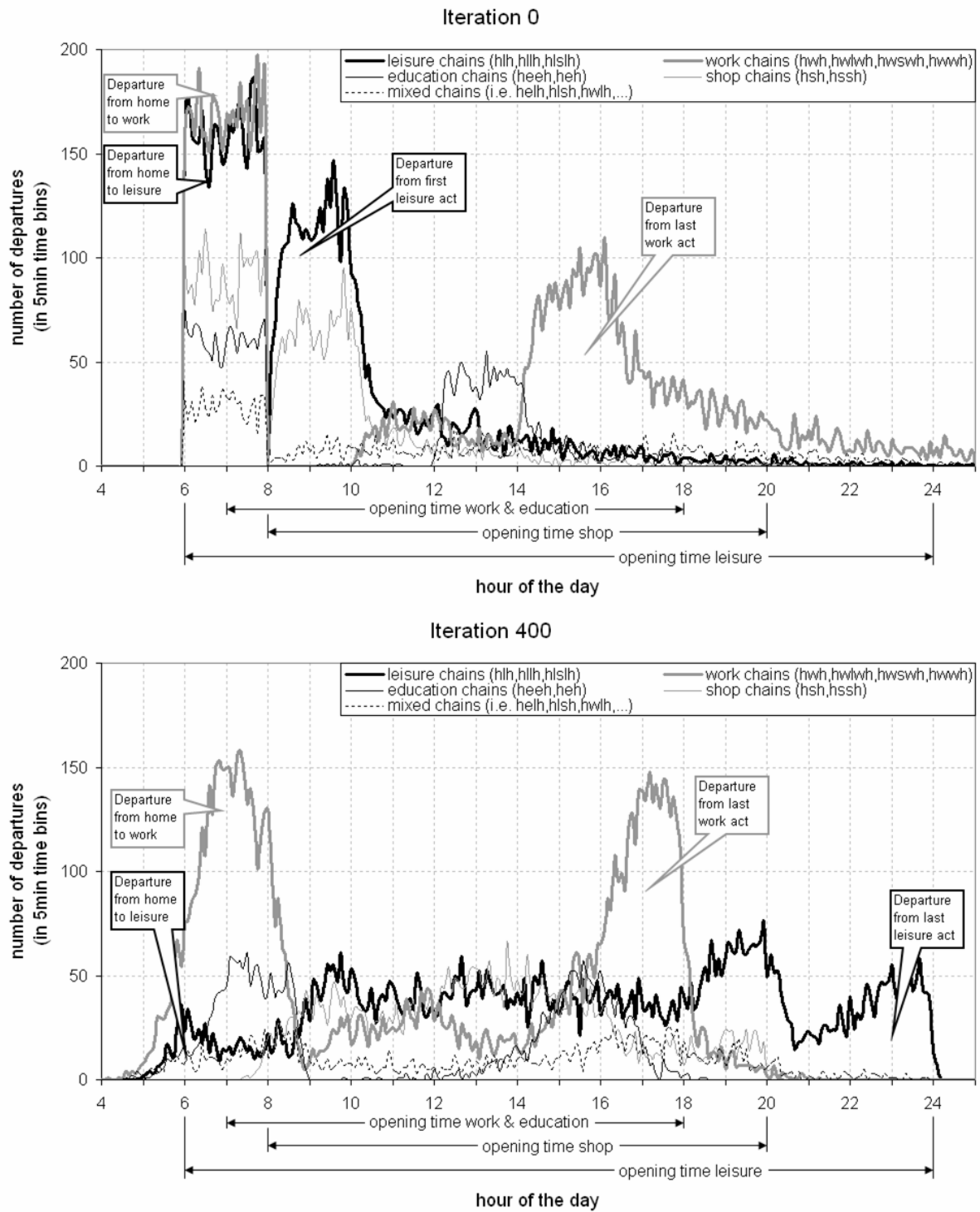


FIGURE 4 Departure time histograms of a 1% sample population of Kanton Zurich.