Agent-Based Demand-Modeling Framework for Large-Scale Microsimulations

Michael Balmer, Kay W. Axhausen, and Kai Nagel

Microsimulation is becoming increasingly important in traffic demand modeling. The major advantage over traditional four-step models is the ability to simulate each traveler individually. Decision-making processes can be included for each individual. Traffic demand is the result of the different decisions made by individuals; these decisions lead to plans that the individuals then try to optimize. Therefore, such microsimulation models need appropriate initial demand patterns for all given individuals. The challenge is to create individual demand patterns out of general input data. In practice, there is a large variety of input data, which can differ in quality, spatial resolution, purpose, and other characteristics. The challenge for a flexible demand-modeling framework is to combine the various data types to produce individual demand patterns. In addition, the modeling framework has to define precise interfaces to provide portability to other models, programs, and frameworks, and it should be suitable for large-scale applications that use many millions of individuals. Because the model has to be adaptable to the given input data, the framework needs to be easily extensible with new algorithms and models. The presented demand-modeling framework for large-scale scenarios fulfils all these requirements. By modeling the demand for two different scenarios (Zurich, Switzerland, and the German states of Berlin and Brandenburg), the framework shows its flexibility in aspects of diverse input data, interfaces to third-party products, spatial resolution, and last but not least, the modeling process itself.

Microsimulation is becoming increasingly important in traffic simulation, traffic analysis, and traffic forecasting (*1; 2,* pp. 171–184; *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; and
- Explicit modeling of the decision-making processes of individuals.

The last point is important since it is not the vehicle that produces traffic, it is the person who drives the vehicle. People do not just produce traffic; instead, each of them tries to manage his or her day (week, life) in a profitable way. They go to work to earn money, they go hiking for health and pleasure, they visit relatives for pleasure or because they feel obliged to do so, they shop to cook dinner at home,

and so on. Since not all this activity can be done at the same location, they have to travel, thus producing traffic. To plan an efficient day, each person makes many decisions:

- 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 play tennis today? (activity starting time choice decision)
- Should I go to visit a friend? (activity type choice decision)
- Who 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, whereas others are made spontaneously as a reaction to specific circumstances. Many decisions induce other decisions. For example, if one is late for work, one must work longer, leaving no time to go shopping that day; therefore, one will need 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 adheres to this schedule that produce traffic.

To simulate a typical day in an urban area, microsimulation tools require information about the schedule of each individual and some knowledge about people's decision-making processes. The challenge is to create this individual demand out of general input data. In practice, there is a large variety of input data. These data 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. In addition, the data have to define precise interfaces to provide portability to other models and programs, and they should be suitable for large-scale scenarios that include many millions of 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.

Such a modeling framework is presented for large-scale scenarios. After a summary about related research and a description of the program structure, the framework is used to model daily demand for two different scenarios. One is a medium-resolution scenario that takes place in the canton of Zurich and consists of about 1.3 million individuals. The second large-scale scenario is defined for Berlin and Brandenburg for about 7 million inhabitants. The two scenarios differ in the amount of available information, spatial resolution, and size, as well as in the demand-modeling process itself.

M. Balmer and K. W. Axhausen, Institute for Transport Planning and Systems (IVT), ETH Zurich, Switzerland. K. Nagel, Transport Systems Planning and Transport Telematics (VSP), Technischen Universität Berlin, Berlin, Germany.

A short overview is then given about further use of the generated daily demand. For that, an iterative, large-scale microsimulation model is used (*4*).

RELATED RESEARCH

The work presented here falls into the area of activity-based demand generation (ABDG) and encompasses a fair number of existing ABDG packages (*1; 2,* pp. 171–184; *3; 5–7*). Internally, those packages are structurally comparable with what is described here, and in terms of methods, they are considerably more sophisticated. Yet their output is typically expressed in terms of (time-dependent) origin–destination (O-D) matrices to be fed into dynamic traffic assignment (DTA) models (*8–10*).

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

• TRANSIMS uses "flat" file formats between modules, whereas MATSIM uses more powerful hierarchical XML formats (*12*).
• With the XML format, it is always possible to do all information exchanges between modules with the same file format and the same document type definition (DTD), varying only the detail level of the included information (*12*). This development means that arbitrary combinations of partial ABDG modules can be used.
• MATSIM keeps track of multiple plans per agent (*12*) generically; for TRANSIMS, this function would require considerable changes in the implementation.
• The MATSIM traffic flow simulation (although simplified when compared with TRANSIMS) runs considerably faster, thus allowing meaningful runs in days instead of weeks (*13*).

• In contrast to TRANSIMS, the MATSIM DTA process keeps track of activity chain consistency along the time axis; travelers have to spend a minimum amount of time at an activity before they can proceed.

The approach described here is similar to TRANSIMS in being fully traveler-oriented (agent-oriented) on all levels, and it is similar to the MATSIM standards in terms of structuring and formatting information. In contrast to TRANSIMS, the approach described here is flexible and universal in terms of the data input requirements and this aspect is tested with data available in Switzerland and Germany.

Other approaches are the agent-based land use models, for example, UrbanSim (*14*), ILUTE (*15*), or the models by Hunt et al. (*16*). These models face difficulties similar to the experiences described here, particularly the need to assemble a consistent agent-based view of the world from diverse data sources (*17*).

In the long run, it would be useful to have a plug-and-play approach between these different models, that is, modules could be interchanged between different modeling systems and coupled in various ways.

FRAMEWORK

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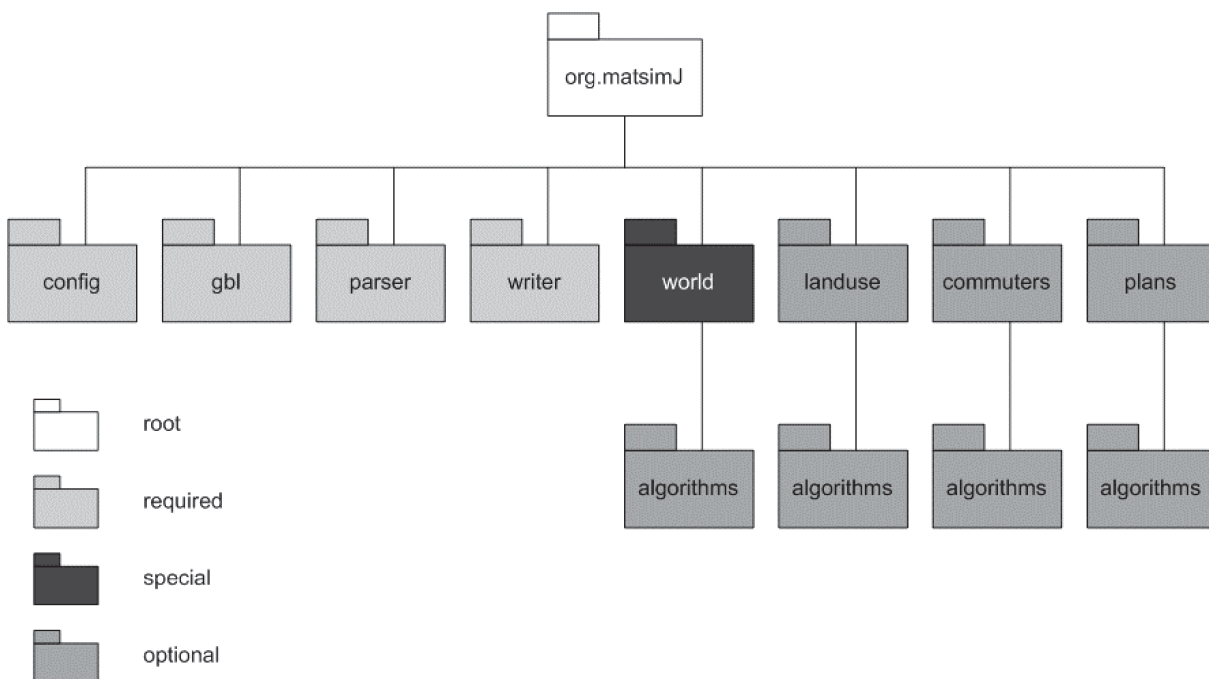

FIGURE 1 Package structure of demand-modeling framework.

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 required 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 them again in the XML data format in an enriched, reduced, or unchanged form.

Framework Packages

Parser, Writer, and Gbl Packages

The parser package provides a general base class for parsing XML data via the SAX parser (*18*). The writer package provides the base for writing XML files. The purpose of the gbl package is to hold global constants and globally accessible functions.

Config Package

In the entire framework, just one configuration data structure exists, which holds all required input parameters for a specific demand-modeling process. Typical parameters are locations of input data, different file formats, special function parameters, and so on. The information stored in this package can be accessed from every part of the framework (global singleton design pattern) (*19*).

World Package

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

Another important functionality of the world package has to be mentioned. During the demand-modeling process, the resolution of the world can differ. For example, land use information might be based on a raster of 100- by 100-m cells, whereas a commuter matrix is based on traffic analysis zones. Therefore, it is necessary to disaggregate one data set into another world resolution. The world package will hold this mapping for an arbitrary number of resolutions of the same region. With it, on-the-fly disaggregation is provided. The mapping between two resolution layers can either be generated by the framework according to a specified mapping rule or given as predefined input data.

Although this package plays a central role in the framework, it is not required because certain demand-modeling steps do not require spatial information (e.g., generating departure times of activities).

Land Use and Commuters Packages

The land use and commuters packages are optional packages that store information for the demand-modeling process. Dependent on the input data, these packages are used, for example, to find locations of work activities or to create sets of leisure facilities. The framework allows the addition of other optional packages in the same manner as the land use and commuters packages for new data sources.

Plans Package

The plans package 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 can also be accessed via the MATSIM website (*4*), and it is used as a working file, utilizing the extreme flexibility of the MATSIM schedule file. In the minimum version, the file holds only the identity numbers (IDs) of all simulated persons. However, it is possible to add a large amount of information about each person to the file, such as age, sex, car ownership, home and work location (at different world resolutions), and more.

In addition, 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 forth.

The internal data structure of the plans package provides exactly the same flexibility as the XML file format. Therefore, it is possible to sequentially add additional schedule details to an 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 different algorithms have to be used or implemented for each new scenario, it is critical that algorithms be clearly separated from the data structure. They also should be easily exchangeable. The order in which algorithms are called should be flexible as well.

The algorithms are collected into a subpackage of the data structure that they manipulate. Therefore, an algorithm at org.matsimJ. commuters.algorithms manipulates data in the commuters data structure (see examples later in this discussion).

External Models

At any point during the demand-modeling process, the framework allows the storage of all data into well-defined XML data files. By definition, these files respect the format as described in the underlying DTDs. Therefore, a clean 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 sections, the external secondary location choice module (*20*) in its adapted version as described by Balmer et al. (*21*, pp. 5–36) was used, which chooses locations for secondary activities in such a way that each

agent improves its daily activity chain. In each given activity chain, there are clearly specified primary activities, and others are defined as secondary activities (*21,* pp. 5–36). 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 those data. One simple and rapid solution is to load all information into memory and 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. Although the description of the world, the land use data, commuter matrices, and others typically holds a relatively small amount of data, the amount of information for individuals goes far beyond the size of typical memory capacity (on average, around 1 to 2 GB).

To handle this problem, the demand-modeling framework 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 in a given scenario is unlimited. This idea will still work if, instead of single persons, demand at the household level for a small number of persons is modeled. But the limit of this approach will be reached if it is also desired to add the concept of social networks (*22*). In that 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 a sufficient amount of memory. The user of the framework can switch between streaming and no streaming by setting a defined parameter flag.

DEMAND MODELING FOR CANTON OF ZURICH, SWITZERLAND

With the framework described in the previous section, the steps taken to model daily demand for the canton of Zurich, Switzerland, are presented. The following subsections describe briefly which input data were used and which algorithms were employed to model the daily demand.

It should be noted that the algorithms used are not very sophisticated. They demonstrate the use of the framework rather than delivering state-of-the-art demand-modeling processes.

Data Resources

The world package describes the region (canton of Zurich, Switzerland) at two different resolutions:

• Municipality level *M:* 170 municipalities and 12 additional districts inside the city of Zurich and
• Raster of 100- by 100-m cell resolution (raster level denoted *R*); in total, 167,881 cells are given.

Mapping between those two levels was also available. The mapping rule is nonambiguous. Each raster cell belongs to no or one municipality, whereas each municipality holds at least one raster cell. Further, the municipalities hold only those cells that include urban areas; that is, the "lake cells" are not part of the mapping.

Information about population distribution in the municipalities was generated by Frick (*23*). It holds about 1.3 million inhabitants with the following data:

• Home location (*R*),
• Population group (children, workers, nonworkers, seniors),
• Mobility (car, season ticket ownership, bike, walking),
• Age, and
• Sex.

The Swiss Federal Office for Statistics (BFS) (*24*) provides land use data that contain information about capacities of different activity types like work, shopping, education, and so on. That information is based on raster level (*R*).

The Pendlermatrix 2000 (commuter matrix) (*25*) contains information about work and school commuters at the municipal level (*M*).

The microcensus (*26*) is a periodic travel behavior survey of the Swiss population. It has been run every 5 years since 1974. The microcensus is carried out by the Federal Office for Spatial Development (ARE) in cooperation with the BFS and contains detailed information about the mobility behavior of 30,000 persons from all over Switzerland (*27,* pp. 61–79). About 1,670 different activity chains can be found in Microcensus 2000. Most of them appear rarely, so only the 100 most frequently occurring activity chains were used, which still covers more than 90% of all chains. By reducing the number of activity types to home, work, education, shop, and leisure, some activity types like service were recoded to work. The result was 21 different activity chains (*21,* pp. 5–36) and their distribution.

Steps in Demand-Modeling Process

The demand-modeling process is split into six sequential steps (see Figure 2 for an overview). Each process step (except the first) 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, and so on are included. It also defines in which land use raster element (*R*) this person lives.

2. Mapping. Each raster element of the land use data belongs to just one municipality. By using this kind of mapping, each agent can be assigned to the municipality of his or her home location.

3. Distribution. Given the distribution of the activity chains described earlier, one of the chains was assigned to each person according to the given distribution. The fact that children do not go to work was also taken into account, and therefore persons of young age are not allowed to hold an activity chain including a work activity. Independent random sampling from aggregate distributions can cause a lack of consistency with the given distribution. To avoid this problem, it was necessary to parse the population twice to obtain aggregate information (see Computational Issues).

4. Distribution. The Pendlermatrix 2000 (*25*) contains information about work and school commuters at the municipality level (*M*). With the assumption about primary activities (*21,* pp. 5–36), it was possible to add locations for the primary activities of work and education. Unfortunately, similar data for the primary activi-
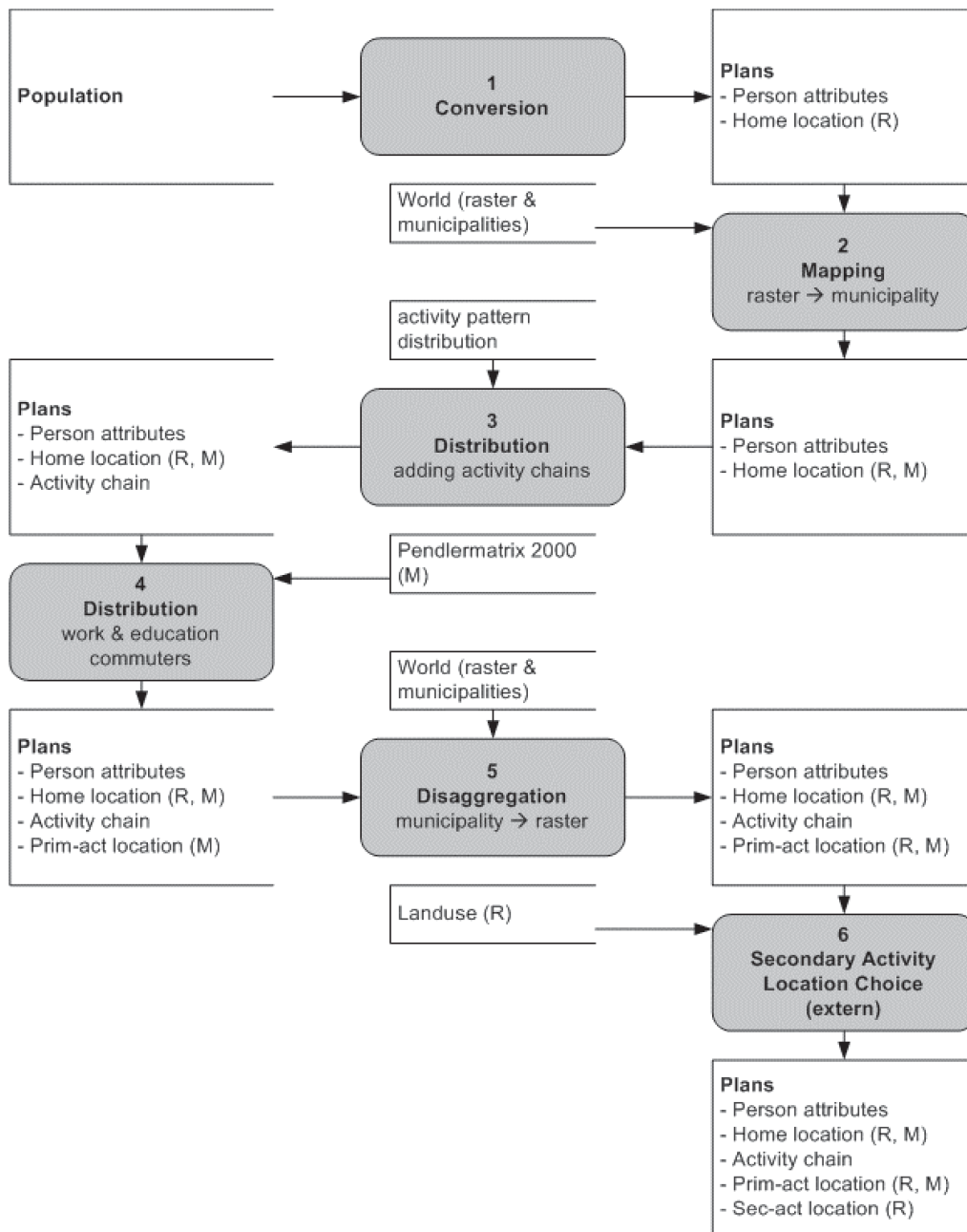
FIGURE 2   Process plan for modeling demand for canton of Zurich.

ties of shopping and leisure were not available. Until better data (or an appropriate algorithm) are available, it will be assumed that those activities are undertaken in the same municipality (or district) in which the person lives.

5. Disaggregation. Step 6 uses location choice (see Figure 2) working at the raster level of detail (*R*). Therefore, it was necessary to disaggregate the locations of the secondary activities to that level. This step was accomplished by uniformly selecting one of the raster elements of the given municipality (or district).

6. Secondary activity location choice (external). The module of Marchal and Nagel (*20*) was used as an external program example to add missing locations of secondary activities.

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

- Person attributes,
- Which activities will be performed,
- In what order the activities will be performed, and
- Where the activities will be performed

Some final algorithmic steps are missing in this modeling process: there is no information about the times when an activity should be performed or how long it should take. It also is not known which mode of transportation will be used and which route to take to go

from one location to the next. These steps were left out purposely because it was desired to use the MATSIM iterative optimization process (*12*) to create that information. A short description of this process follows later.

DEMAND MODELING FOR GERMAN STATES OF BERLIN AND BRANDENBURG

The scenario for Berlin and Brandenburg, Germany, is somewhat different from that for the canton of Zurich because only less precise information is available. Again, sophisticated algorithms are not used. It is more interesting to show that some algorithms from the Zurich scenario (tree structure) could be reused whereas others had to be added.

Data Resources

The world here describes the states of Berlin and Brandenburg (*28*). In this scenario, different spatial resolutions are used:

- The region is split into 1,008 traffic analysis zones (TAZs) (zone level denoted *Z*);
- Each zone consists of several blocks; 12,260 blocks are available for this scenario (block level denoted *B*); and
- A raster of 169,420 cells is also available; each cell has an extent of 500 m by 500 m (raster level denoted *R*).

The following mapping rules apply. Each TAZ consists of at least one block. Each block belongs to only one TAZ. A cell refers to none, one, or many blocks, and each block refers to at least one cell.

The raster is synthetically generated; therefore no additional information is encoded. The raster and mapping here were also created by the framework described earlier using a special world algorithm in the org.matsimJ.world.algorithms package.

Only the population of Berlin itself (about 3.3 million individuals) was available at that time (*28*) and less detail was obtainable than for the population of the canton of Zurich. The only available items were home locations (*B*) and age.

Land use data for Berlin and Brandenburg are based on TAZ level *Z*. The data consist of capacities for home, work, education, shopping, and leisure activities.

For this scenario, no commuter matrix was available (and no production–attraction matrix), so only a 24-h O-D matrix is given (*29*). This matrix was used in two different ways. First, for all trips starting in Berlin, the information was used to guess where individuals of the corresponding zone were performing their primary activity. In other words, it was used in a way similar to the commuter matrix of the canton of Zurich scenario. The remaining trips starting in the TAZs of Brandenburg were used to create the missing part of the population and to define the location of their primary activity. This is, of course, a very approximate way to interpret an O-D matrix, but a step was needed to close the information gap. One of the greatest priorities of future efforts is to design a better primary activity location choice method for the Berlin–Brandenburg scenario.

The same activity pattern distribution was used as that for the canton of Zurich scenario. Those patterns reflect average Swiss behavior and thus do not necessarily reflect those of Berlin and Brandenburg, but it is a feasible start as long as no other data are available.

Demand-Modeling Process Steps

Three different levels of spatial resolution were used here compared with two for the Zurich scenario, resulting in seven steps. Three additional steps were also needed to fill in the information gap and are designated A, B, and C (see Figure 3):

A. Creation. As mentioned earlier, no information is available about the population of Brandenburg. Because the number of trips in the O-D matrix starting in the zones of Brandenburg fits with its total population, one individual was assigned for each of those trips to that start zone. This approximation is, of course, a rough estimate of the distribution of the population of Brandenburg.

B. Disaggregation. Since the individuals are placed at the TAZ level (*Z*), it was necessary to map them to block level (*B*), so a uniform distribution over all the blocks of a given zone was assumed.

C. Land use disaggregation. To use the secondary activity location choice module (*20*), the land use information has to be given at the raster level. Therefore, the TAZ-based land use data also have to be disaggregated into a raster level (*R*). This separation was made in a first step by using uniform distribution over all the blocks (*B*) of each TAZ. In a second step, another uniform distribution was used to propagate the land use data from block level (*B*) to raster level (*R*).

1. Conversion. The conversion is the same as that described earlier. 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 is necessary. This mapping is nonambiguous.

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. The distribution is the same as that for the Zurich scenario.

5. Distribution. In contrast to the same step for the canton of Zurich process, all the different primary activity types were assigned according to the given matrix. Because there was no information about the purpose of a given trip for the matrix, it cannot be distinguished from the others. The location of the primary activity is chosen according to the destination distribution of those trips starting at the home location of the current individual.

6. Disaggregation. Again it was necessary to disaggregate the location of the secondary activity to the raster level (*R*), which was done in the same way as that described for land use (see Step C).

7. Secondary activity location choice (external). The secondary location choice module by Marchal and Nagel (*20*) was again used to find the missing locations as described in Step 6 for the canton of Zurich scenario.

The final demand of the Berlin–Brandenburg scenario holds the same data items as those described for the Zurich scenario.

MATSIM Iterative Optimization Process

As already mentioned, the generated demand as input for the MATSIM iterative optimization process was used. The work of Raney and Nagel (*12*) shows 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
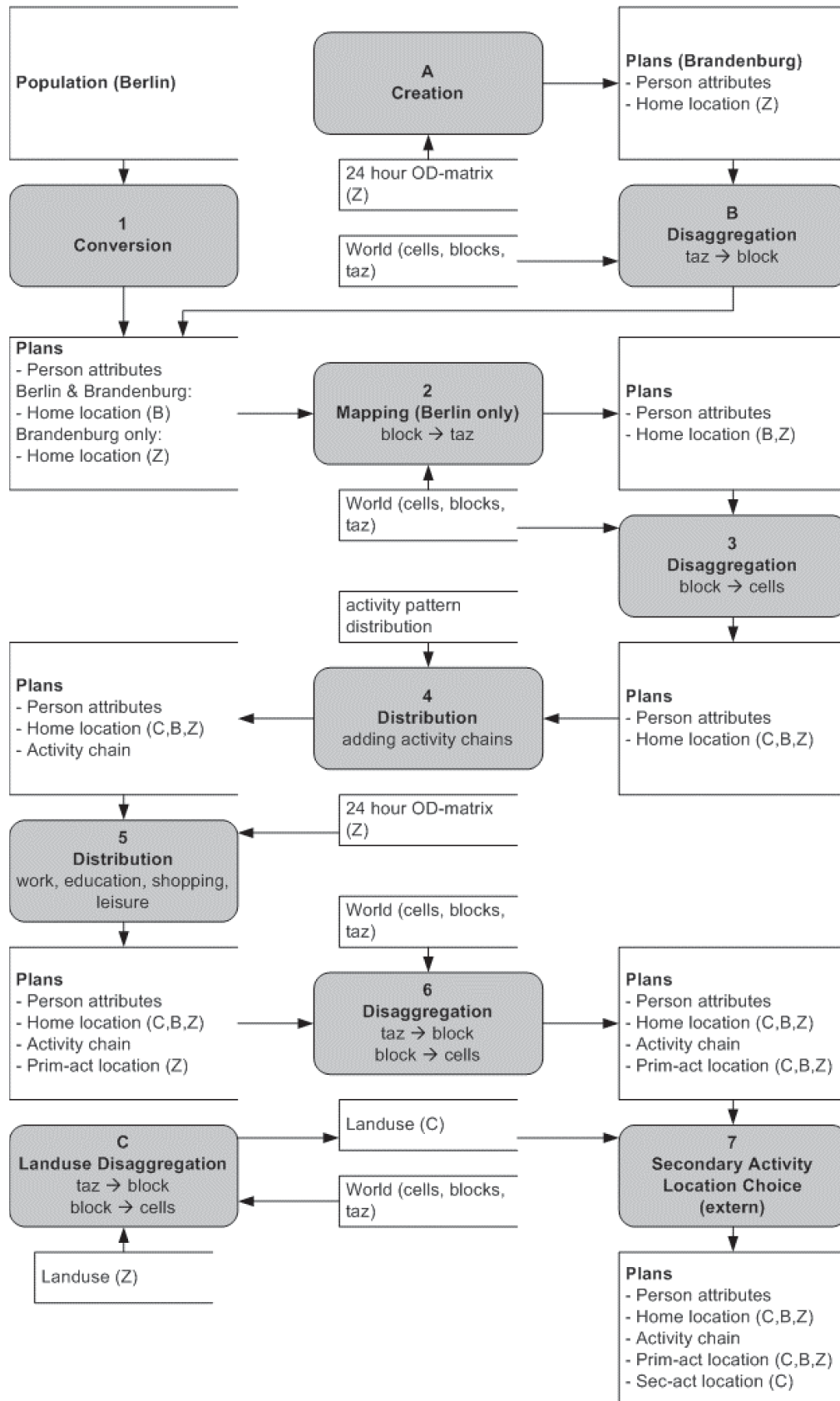
FIGURE 3   Process plan for modeling demand for Berlin and Brandenburg.

appropriate algorithm in the demand-modeling framework since MATSIM can be used as a final external module for that purpose.

In contrast to the work by Raney and Nagel (*12*), an enhanced time allocation module called planomat was used (*30*). With 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 are already done by the demand-modeling framework and the traffic flow simulation of MATSIM (*13*) handles the individual transport mode, the functionality of the planomat module is reduced to time choice.

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

Time Scheduling Results

With the same MATSIM setup as that described by Balmer et al. (*21,* pp. 5–36), a 1% sample of the canton of Zurich scenario population was run. By artificially reducing each link's capacities within the given network (*13*) by a factor of 50, congestion patterns similar to those obtained by using all persons were produced. Because of the small sample size, the simulations displayed 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 quite plausible.

Figure 4 shows the differences between the initial departure time schedule (Iteration 0) and the departure time schedule after 400 iterations. Two groups of individuals are highlighted:

• Heavy gray line. Individuals with schedules containing work as the second and second-to-last activity in the chain.
• Heavy 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 7:00 a.m. and 6:00 p.m. and the duration of performing work should be around 8 h [in the work by Charypar and Nagel (*31*) 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 rigid constraint (the setup defines it between 6:00 a.m. and midnight). The operating point *t** of performing one leisure activity is set to 2 h. Therefore, there is much more time-scheduling flexibility for leisure activities and Figure 4 reflects that fact. Although the time choice at Iteration 0 is unsatisfying, the improvement after 400 iterations is immense. Only a few persons choose the morning peak hour to travel, whereas the majority conduct their leisure activities in the evening (finishing their last leisure activity just before midnight, when almost no traffic exists).

DISCUSSION AND OUTLOOK

A 1% population sample was used to reduce the required computational resources. Running 10% samples, however, is not a major challenge, and even running full populations is possible—the nec-

essary resources grow considerably more slowly than the sample size. Such results will be reported on in the future.

The planomat MATSIM module mentioned earlier provides an attractive amount of functionality. Since it also supplies, besides the already-used time choice, a location choice and a mode choice module, it is interesting to actually use these functionalities. To do that, the module requires additional information.

As a first step, it is necessary to provide an alternative to the individual transport mode in MATSIM. Therefore, the demand-modeling framework will be extended so 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.

As a second step, another package will be added to the framework containing data about activity spaces, catchment areas, and commuter sheds (*32*). With those data, the primary activity location choice (commuter package) and secondary activity location choice (external module) can be replaced by adding a set of possible locations for all available activity types to each person in 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, it is a relatively simple task to extend the demand-modeling framework via the idea of intrahousehold 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.

SUMMARY

An agent-based demand-modeling framework for large-scale microsimulations is presented. With the framework for the Zurich and the Berlin–Brandenburg scenarios, it is shown that the framework

• Is flexible enough to handle a variety of input data (Zurich data differ from Berlin–Brandenburg data);
• Is flexible enough to extend or replace algorithms (different algorithms for Berlin–Brandenburg commuter data than for Zurich commuters);
• Provides disaggregation to different spatial resolutions (two levels of resolution for the Zurich scenario and three levels for the Berlin–Brandenburg scenario);
• Provides a robust interface to third-party models, programs, and frameworks (the external modules: secondary activity location choice and MATSIM dynamic traffic assignment);
• Is suitable for an unlimited number of individuals (about 7 million people for the Berlin–Brandenburg scenario); and
• 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 data
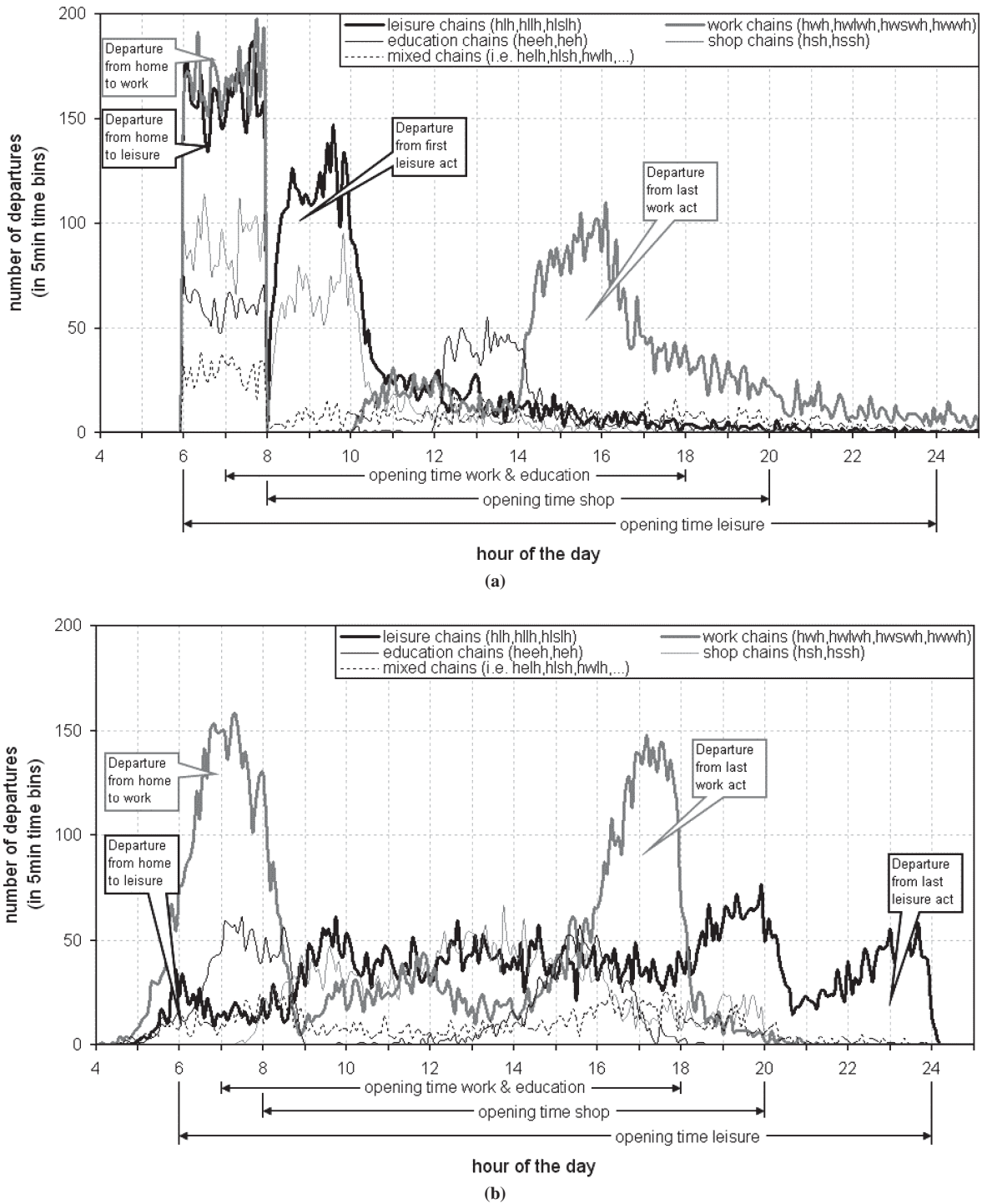
FIGURE 4 Departure time histograms of 1% sample population of canton of Zurich: (*a*) initial schedule and (*b*) schedule after 400 iterations.

management requirements, data processing, computational design, implementation, and run time issues, it is necessary to integrate computer knowledge into the transportation research process.

Last, but not least, it is hoped that this work will encourage researchers with some basic programming skills to use, enhance, and extend the framework according to their own desires and needs.

REFERENCES

1. Vovsha, P., E. Petersen, and R. Donnelly. Microsimulation 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,* Transportation Research Board of the National Academies, 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,* Vol. 3, Elsevier, Oxford, United Kingdom, 1999.

3. Bhat, C. R., J. Y. Guo, S. Srinivasan, and A. Sivakumar. Comprehensive Econometric Microsimulator for Daily Activity-Travel Patterns. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1894,* Transportation Research Board of the National Academies, Washington, D.C., 2004, pp. 57–66.

4. *MATSIM, Multi Agent Traffic SIMulation.* www.matsim.org. Accessed Oct. 2005.

5. *VISEM.* Planung Transport und Verkehr (PTV), Karlsruhe, Germany. www.ptv.de/cgi-bin/traffic/traf_visem.pl. Accessed Oct. 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. www.dot.state.fl.us/research-center/Completed_Proj/Summary_PTO/FDOT_BA496.pdf. Accessed Oct. 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: Journal of the Transportation Research Board, No. 1706,* TRB, National Research Council, Washington, D.C., 2000, pp. 136–144.

8. *DynaMIT.* Massachusetts Institute of Technology, Cambridge. mit.edu/its/dynamit.html. Accessed Oct. 2005.

9. *DYNASMART.* University of Maryland, College Park. www.dynasmart.umd.edu. Accessed Oct. 2005.

10. Friedrich, M., I. Hofsäss, K. Nökel, and P. Vortisch. A Dynamic Traffic Assignment Method for Planning and Telematic Applications. *Proc., Seminar K, European Transport Conference,* Cambridge, United Kingdom, 2000.

11. *TRANSIMS, TRansportation ANalysis and SIMulation System.* Los Alamos National Laboratory, Los Alamos, N. Mex. transims.tsasa.lanl.gov. Accessed Oct. 2005.

12. Raney, B., and K. Nagel. An Improved Framework for Large-Scale Multi-Agent Simulations of Travel Behavior. In *Towards Better Performing European Transportation Systems* (P. Rietveld, B. Jourquin, and K. Westin, eds.), 2005 (in press). www.vsp.tu-berlin.de/publications. Accessed Oct. 2005.

13. Cetin, N., and K. Nagel. A Large-Scale Agent-Based Traffic Micro-Simulation Based on Queue Model. Presented at the Swiss Transport Research Conference, 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: Journal of the Transportation Research Board, No. 1780,* TRB, National Research Council, Washington, D.C., 2001, pp. 53–63.

17. Abraham, J. E., T. Weidner, J. Gliebe, C. Willison, and J. D. Hunt. Three Methods for Synthesizing Base-Year Built Form for Integrated Land Use–Transport Models. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1902,* Transportation Research Board of the National Academies, Washington, D.C., 2005, pp. 114–123.

18. *SAX, Simple API for XML.* www.saxproject.org. Accessed Oct. 2005.

19. Gamma, E., R. Helm, R. Johnson, and J. Vlissides. *Design Patterns: Elements of Reusable Object-Oriented Software.* Addison-Wesley Professional Computing Series, Boston, Mass., 1995.

20. Marchal, F., and K. Nagel. Modeling Location Choice of Secondary Activities with a Social Network of Cooperative Agents. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1935,* Transportation Research Board of the National Academies, Washington, D.C., 2005, pp. 141–146.

21. Balmer, M., M. Rieser, A. Vogel, K. W. Axhausen, and K. Nagel. Generating Day Plans Using Hourly Origin-Destination Matrices. In *Jahrbuch 2004/05 Schweizerische Verkehrswirtschaft* (T. Bieger, C. Laesser, and R. Maggi, eds.), SVWG, St. Gallen, Switzerland, 2005.

22. Axhausen, K. W. Activity Spaces, Biographies, Social Networks and Their Welfare Gains and Externalities: Some Hypotheses and Empirical Results. Presented at PROCESSUS Colloquium, Toronto, Ontario, Canada, 2005. www.civ.utoronto.ca/sect/traeng/ilute/processus2005/PaperSession/Paper18_Axhausen_ActivitySpacesBiographies_CD_.pdf. Accessed Oct. 2005.

23. Frick, M. A. Generating Synthetic Populations Using IPF and Monte Carlo Techniques: Some New Results. Presented at 4th Swiss Transport Research Conference, Monte Verita, Ascona, Switz., 2004. www.ivt.ethz.ch/vpl/publications/reports. Accessed Oct. 2005.

24. BFS, Swiss Federal Statistical Office. www.bfs.admin.ch/bfs/portal/en/index.html. Accessed Oct. 2005.

25. Vrtic, M., and K. W. Axhausen. Experiment mit einem dynamischen Umlegungsverfahren. *Strassenverkehrstechnik,* Vol. 47, No. 3, 2003, pp. 121–126.

26. *Mobilität in der Schweiz, Ergebnisse des Mikrozensus 2000 zum Verkehrsverhalten.* ARE, Bundesamt für Raumentwicklung und Bundesamt für Statistik, Bern und Neuchâtel, Switzerland, 2001. www.are.admin.ch/are/de. Accessed Oct. 2005.

27. Axhausen, K. W., and M. Frick. Nutzungen, Strukturen, Verkehr. In *Stadtverkehrsplanung: Grundlagen, Methoden, Ziele* (G. Steierwald, H.-D. Künne, and W. Vogt, eds.), Springer, Heidelberg, Germany.

28. RBS, Raumbezugssystem Berlin. *Bevoelkerungsverteilung, Teil-Verkehrszellen und Block Geometrie/Hierarchie.* Statistisches Landesamt Berlin, 2000. www.statistik-berlin.de. Accessed Oct. 2005.

29. *Datengrundlagen Stadtentwicklungsplan Verkehr.* SfSB, Senatsverwaltung für Stadtentwicklung Berlin, 1998. www.stadtentwicklung.berlin.de. Accessed Oct. 2005.

30. 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, Switzerland, 2005. www.ivt.ethz.ch/vpl/publications/reports. Accessed Oct. 2005.

31. Charypar, D., and K. Nagel. Generating Complete All-Day Activity Plans with Genetic Algorithms. *Transportation,* Vol. 32, No. 4, 2005, pp. 369–397.

32. Axhausen, K. W., M. Botte, and S. Schönfelder. *Systematic Measurement of Catchment Areas.* CTPP 2000 Status Report. U.S. Department of Transportation, 2004. www.fhwa.dot.gov/ctpp/sr0804.pdf. Accessed Oct. 2005.