Netw Spat Econ (2012) 12:481–502
DOI 10.1007/s11067-011-9164-9

Behavioral Calibration and Analysis of a Large-Scale Travel Microsimulation

Gunnar Flötteröd · Yu Chen · Kai Nagel

Published online: 5 August 2011

Abstract This article reports on the calibration and analysis of a fully disaggregate (agent-based) transport simulation for the metropolitan area of Zurich. The agent-based simulation goes beyond traditional transport models in that it equilibrates not only route choice but all-day travel behavior, including departure time choice and mode choice. Previous work has shown that the application of a novel calibration technique that adjusts all choice dimensions at once from traffic counts yields cross-validation results that are competitive with any state-of-the-art four-step model. While the previous study aims at a methodological illustration of the calibration method, this work focuses on the real-world scenario, and it elaborates on the usefulness of the obtained results for further demand analysis purposes.

Keywords Multi-agent simulation · Dynamic traffic assignment · Disaggregate demand calibration · Real-world application

G. Flötteröd (✉)
Transport and Mobility Laboratory (TRANSP-OR), Ecole Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland
e-mail: gunnar.floetteroed@epfl.ch

Y. Chen · K. Nagel
Transport Systems Planning and Transport Telematics Laboratory, Berlin Institute of Technology, 10587 Berlin, Germany

Y. Chen
e-mail: chen@vsp.tu-berlin.de

K. Nagel
e-mail: nagel@vsp.tu-berlin.de

1 Introduction

The well-known four-step process, consisting of trip generation, trip distribution (= destination choice), mode choice, and route assignment, has been *the* modeling tool in urban transportation planning for many decades (Ortuzar and Willumsen 2004). However, the four-step process, at least in its traditional form, has many problems with modern issues, such as time-dependent effects, more complicated decisions that depend on the individual, or spatial effects at the micro (neighborhood) scale (Vovsha et al. 2004).

An alternative is to use a microscopic approach, where every traveler is modeled individually. One way to achieve this is to start with the synthetic population and then work the way "down" towards the network assignment. This typically results in activity-based demand models (ABDM, e.g, Bhat et al. 2004; Bowman et al. 1998; Jonnalagadda et al. 2001; Pendyala 2004), which sometimes do and sometimes do not include the mode choice, but typically end with time-dependent origin-destination (OD) matrices, which are then fed to a separate route assignment package. The assignment package computes a (typically dynamic) route equilibrium and feeds the result back as time-dependent zone-to-zone travel impedances. When feedback is implemented, then the activity-based demand model recomputes some or all of its choices based on those travel impedances (Lin et al. 2008).

This type of coupling between the ABDM and the traffic assignment leaves room for improvement (Balmer et al. 2004; Rieser et al. 2007). In particular, it can be argued that route choice is also a behavioral aspect, and in consequence the decision to include route choice into the assignment model rather than into the demand model is arbitrary. Problems immediately show up if one attempts to base a route choice model in a toll situation on demographic characteristics—the demographic characteristics, albeit present in the ABDM, are no longer available at the level of the assignment. Similarly, in all types of intelligent transport system (ITS) simulations, any modification of the individuals' decisions beyond route choice becomes awkward or impossible to implement.

An alternative is to split the assignment into a route choice model and a network loading model and to add the route choice to the ABDM, which leaves the network loading as the sole non-behavioral model component. If it is implemented as a microscopic or mesoscopic traffic flow simulation, then the integrity of the simulated travelers can be maintained throughout the entire modeling process. This has the following advantages:

- Both the route choice and the network loading can be related to the characteristics of the synthetic person. For example, toll avoidance can be based on income, or emission calculations can be based on the type of vehicle (computed in an upstream car-ownership model).
- Additional choice dimensions besides route choice can be included in the iterative procedure of assignment (de Palma and Marchal 2002; Zhou et al. 2007; Nagel and Flötteröd 2012).

- The fully disaggregate approach enables an ex post analysis of arbitrary demand segments. This is an important advantage over any simulation based on OD matrices, where the aggregation is done *prior* to the simulation.

This implies that, at least in principle, all choice dimensions of the ABDM can react to the network conditions, but it also requires to build models of this feedback for all affected choice dimensions. While, for example, route choice only looks at the generalized cost of the trip, departure time choice also includes schedule delay cost, mode choice compares the generalized costs between different modes, location choice includes the attractiveness of the possible destinations, etc. This brings along a vast increase in modeling opportunities, but it also requires substantially more modeling efforts.

In this article, we report on how such an approach can be implemented, calibrated, and analyzed, using the metropolitan area of Zurich as an example (as a sub-region of an "all-of-Switzerland" scenario (Meister et al. 2008)). In previous work (Flötteröd et al. 2009, 2011), the results of the calibrated simulation are compared to 161 counting stations in the Zurich metropolitan area. Despite of the vastly increased scope of the model when compared to a four-step approach, we are able to reproduce traffic counts with an error of 10–15% throughout the entire analysis period. Qualitatively, these results are competitive with any state-of-the art four-step model, but they come along with entirely new modeling perspectives. While the previously published results aimed at an illustration of the deployed calibration method, this work gives a detailed analysis of the real-world scenario and the calibration results, and it elaborates on the usefulness of these results for further demand analysis purposes. Specifically, we investigate how certain characteristic numbers generated by the calibration can be behaviorally interpreted, and how this interpretation facilitates a further trip generation/attraction analysis and the identification of over/under-estimated demand segments.

The quality of the presented real-world results is to a large extent due to new methodological advances on the calibration side: Until recently, the four-step-process was ahead of our approach in this regard because its simple mathematical structure allowed for the development of a broad variety of (more or less automated) demand calibration procedures. In this article, however, we deploy a novel methodology for the calibration of demand microsimulations from network conditions such as traffic counts. The theory for this was developed over the last couple of years (Flötteröd et al. 2011; Flötteröd 2008).

The remainder of this article is organized as follows. Sections 2 and 3 introduce the used microsimulation and the deployed calibration system. The field study is described in Section 4. Section 5 details the mechanisms through which the calibration takes effect and elaborates on the further demand analysis opportunities this brings along. Finally, Section 6 summarizes the article and indicates future research opportunities.

2 Transport microsimulation

The MATSim ("Multi-agent transport simulation toolkit" MATSIM 2011; Raney and Nagel 2006) transport microsimulation is used for the purposes of this study. The MATSim web site provides a wealth of supplementary material that goes beyond the necessarily brief introduction given here.

This simulation is constructed around the notion of **agents** that make independent decisions about their actions. Each traveler of the real system is modeled as an individual agent. The simulation consists of two major building blocks, which are mutually coupled:

- On the demand side, each agent independently generates a so-called **plan**, which encodes its intentions during a certain time period, typically a day. The plan is an output of an activity-based model that comprises but is not constrained to route choice, and its generation depends on the network conditions expected by the agent.
- On the supply side, the plans of all agents are simultaneously executed in a simulation of the physical system. This is also called the **traffic flow simulation** or **mobility simulation**.

The mutual coupling of demand and supply is iteratively resolved, which can be seen as a mechanism that allows agents to **learn**. The simulation iterates between plan generation and traffic flow simulation. It remembers several plans per agent and evaluates the performance of each plan. Agents normally prefer plans with good performance, but they sometimes re-evaluate inferior plans, and they sometimes obtain new plans by modifying copies of existing plans (Nagel and Flötteröd 2012).

The following subsections explain these items in greater detail.

2.1 Choice set generation

A plan contains the itinerary of activities the agent wants to perform during the day, plus the intervening trip legs the agent must take to travel between activities. An agent's plan details the order, type, location, duration and other time constraints of each activity, and the mode, route and expected departure and travel time of each leg.

A specification of the plan choice set for every agent before the iterations is computationally extremely cumbersome because of the sheer number of possible alternatives (Bowman and Ben-Akiva 1998). Such an approach also is conceptually questionable because the accessibility measures that affect the inclusion of a plan in the choice set are an outcome of the iterations, and hence they are a priori unknown. Therefore, the choice set is continuously updated during the iterations. Speaking in the technical terms of MATSim, a plan can be modified by various **modules**. This paper makes use of the following modules:

- The **activity times generator** randomly changes the timing of an agent's plan. In every iteration, there is a 10% chance that this module is used to generate a new plan.
- The **router** is implemented as a time-dependent Dijkstra algorithm that runs based on link travel times obtained from the mobility simulation. In every iteration, there is a 10% chance that this module is used to generate a new plan.
- **Mode choice** is enabled by ensuring that the choice set of every agent contains at least one "car" and one "non-car" plan.

These modules are used in the following way. In every iteration, each agent selects one plan for execution. With a 10% probability, this plan is uniformly selected, the activity times generator is applied, and then the modified plan is executed. Likewise, there is a 10% probability to uniformly select a plan to which the router is applied before the plan is executed. With the remaining 80% probability, no plan-changing module is used, and an existing plan is selected for execution according to the choice model described in the next subsection. The concrete 10% probability values ensure a stable yet relatively fast convergence of the iterated simulation; they are chosen based on experience. At most one module is applied at once to a plan.

The choice set generation is turned off after a pre-specified number of iterations such that the agents select from a stable choice set using the utility-based choice model described next. This choice model is also applied during the choice set generation in order to drive the system towards a plausible state from the very beginning.

2.2 Choice

In order to compare plans, it is useful to assign a quantitative **score** to the performance of each plan. In principle, arbitrary scoring schemes can be used, e.g., prospect theory (Avineri and Prashker 2003). In this work, a simple utility-based approach is used. The elements of the approach are as follows:

- The total score of a plan is computed as the sum of individual contributions consisting of positive contributions for performing an activity and negative contributions for traveling.
- A logarithmic form is used for the positive utility earned by performing an activity a, which essentially has the following form:

$$V_{\text{perf}}(a) = \beta_{\text{perf}} \cdot t_a^* \cdot \ln t_{\text{perf},a} \tag{1}$$

where $t_{\text{perf},a}$ is the actually performed duration of the activity, t_a^* is the "typical" duration of the activity, and β_{perf} is the marginal utility of an activity at its typical duration. These durations are sampled from empirical distributions that are extracted from census data (SFSO 2006). β_{perf} is the same for all activities since in equilibrium all activities at their typical duration need to have the same marginal utility. As long as activity dropping or activity insertion are not allowed, a minimal duration, sometimes used in other publications, has no effect. Concrete values for the parameters are given later in the description of the case study.

- The (dis)utility $V_{\text{travel}}(l)$ of traveling along a leg l is assumed to be linear in the travel time with different valuations of the time for different transport modes. Again, concrete parameter values are given later on.

The total utility of a plan i can thus be written as

$$V(i) = \sum_{a \in i} V_{\text{perf}}(a) + \sum_{l \in i} V_{\text{travel}}(l). \tag{2}$$

It is important to note that the score thus takes into account the complete daily plan. More details can be found in Raney and Nagel (2006) and Charypar and Nagel (2005).

The plan choice is modeled with a multinomial logit model (which clearly calls for enhancements in the future) (Ben-Akiva and Lerman 1985). However, as stated before, it may happen that an agent receives a newly generated plan from one of the aforementioned plan generation modules, which then is chosen for execution without further evaluation. This is necessary because the utility of a plan is determined from its execution, and hence it is not available for newly generated plans.

Summarizing, the probability $P_n(i)$ that agent n chooses plan i is

$$P_n(i) \begin{cases} = 1 & \text{if } i \text{ is newly generated} \\ \sim \exp(V(i)) & \text{otherwise,} \end{cases} \tag{3}$$

where the normalization of the logit model is omitted for notational simplicity.

2.3 Traffic flow simulation

The traffic flow simulation executes the plans of all agents simultaneously on the network and provides output describing what happened to each individual agent during the execution of its plan. The traffic flow simulation is implemented as a queue simulation, which means that each street (link) is represented as a FIFO (first-in first-out) queue with three restrictions (Cetin et al. 2003; Gawron 1998): First, each agent has to remain for a certain time on the link, corresponding to the free speed travel time. Second, the outflow rate of a link is constrained by its flow capacity. Third, a link storage capacity is defined, which limits the number of agents on the link. If it is filled up, no more agents can enter the link, and spillback may occur.

3 Calibration system

The previous section describes a simulation that predicts the performance of a transportation system through an iterative process that couples complex behavioral and physical models. Notably, some aspects of the simulation are what one may call "procedurally modeled" in that there is no explicit mathematical specification of the respective sub-model but rather a sequence of processing steps that build the model output.

This lack of a comprehensive mathematical perspective on the simulation and its outputs has, until recently, rendered the calibration of the system a task based on intuition and, unfortunately, the arbitrariness this brings along. This section outlines the Cadyts ("Calibration of dynamic traffic simulations" Flötteröd 2009, 2011) calibration tool. Because it allows to calibrate arbitrary choice dimensions from traffic counts in a fully disaggregate manner, it lends itself to an application in the Zurich case study.[1]

3.1 Basic functioning

Cadyts makes no assumptions about the form of the plan choice distribution or about the choice dimensions it represents. It combines the prior choice distribution $P_n(i)$ with the available traffic counts $\mathbf{y}$ into a posterior choice distribution $P_n(i|\mathbf{y})$ in a Bayesian manner. The resulting posterior distribution is, essentially, of the following form (Flötteröd et al. 2011):

$$P_n(i|\mathbf{y}) \sim \exp\left(\frac{\partial \mathcal{L}(\mathbf{y})}{\partial P_n(i)}\right) \cdot P_n(i) \tag{4}$$

where $\mathcal{L}(\mathbf{y})$ is the log-likelihood function of the sensor data $\mathbf{y}$.

Some intuition into the workings of this quite general formulation can be obtained by adopting a simplified perspective where congestion is assumed to be light and the traffic counts are independently and normally distributed. In this setting, the above formula simplifies into[2]

$$P_n(i|\mathbf{y}) \sim \prod_{ak \in i} \exp\left(\frac{y_a(k) - q_a(k)}{\sigma_a^2(k)}\right) \cdot P_n(i) \tag{5}$$

where $y_a(k)$ is the available traffic count on link a in simulation time step k, $q_a(k)$ is its simulated counterpart, and $\sigma_a^2(k)$ is the variance of the respective

[1] Cadyts is not constrained to the MATSim microsimulation but is designed to be compatible with a wide variety of transport simulation systems.

[2] The probability of a measurement $y_a(k)$ would be $p(y_a(k)) \sim \exp[-(y_a(k) - q_a(k))^2/(2\sigma_a^2(k))]$. Because of independence, the probability of a measurement set $\mathbf{y}$ would be the product of this, i.e., $p(\mathbf{y}) \sim \prod_{ak} \exp[-(y_a(k) - q_a(k))^2/(2\sigma_a^2(k))]$. From there, $\frac{\partial \mathcal{L}(\mathbf{y})}{\partial P_n(i)} = \frac{\partial \ln p(\mathbf{y})}{\partial P_n(i)} \sim \sum_{ak \in i} \frac{y_a(k) - q_a(k)}{\sigma_a^2(k)}$, where the sum now goes over all ak that are used by plan i; since the expected traffic volume on a link in a given time interval is in uncongested conditions equal to the sum of the choice probabilities of all plans containing that link in that time interval, the derivative of $q_a(k)$ with respect to $P_n(i)$ is one if $ak \in i$ and zero otherwise.

traffic count. The product runs over all links a and time steps k that (i) are contained in plan i in that the plan schedules to cross that link in the given time step and (ii) are equipped with a sensor. (The calibration functions with arbitrary sensor configurations.)

Intuitively, this works like a controller that steers the agents towards a reasonable fulfillment of the measurements: For any sensor-equipped link, the according $\exp(\cdot)$ factor is larger than one if the measured flow is higher than the simulated flow such that the choice probabilities of plans that cross this link are scaled up. Vice versa, if the measured flow is lower than the simulated flow, the according factor is smaller than one such that plans that cross this link are penalized.

What is described here is a calibration of the individual-level choice distributions in the agent population that does not change the parameters of the choice model that generates the prior choice probabilities $P_n(i)$. On the one hand, this is a quite general result in that it is *independent* of the specification of the choice model. On the other hand, this also implies that, without further modifications, rather an improved picture of the current status quo is obtained than stable parameter estimates that could be used for forecast and scenario analysis. Section 3.3 continues the discussion of this topic in the context of a concrete application to the MATSim simulation system, which is described next.

3.2 Application to MATSim

Apart from the immediate execution of newly generated plans, the behavioral model of MATSim is of the multinomial logit form $P_n(i) \sim \exp(V(i))$. Substituting this into the posterior choice model (4) yields

$$P_n(i|\mathbf{y}) \sim \exp\left(V(i) + \frac{\partial \mathcal{L}(\mathbf{y})}{\partial P_n(i)}\right) \tag{6}$$

That is, an implementation of the posterior choice distribution requires nothing but to add a plan-specific utility correction to every considered plan.

For independently distributed traffic count errors with $\mathcal{L}(\mathbf{y}) = \sum_{ak} \mathcal{L}(y_a(k))$, an assumption that is maintained in the following, the above can be written as

$$P_n(i|\mathbf{y}) \sim \exp\left(V(i) + \sum_{ak\in i} \frac{\partial \mathcal{L}(y_a(k))}{\partial P_n(i)}\right) =: \exp\left(V(i) + \sum_{ak\in i} \Delta V_a(k)\right). \tag{7}$$

Here, the plan-specific utility corrections are composed of link- and time-additive correction terms $\Delta V_a(k)$. These terms are computed per sensor location and -time, but independently of which plan they affect. The utility correction of a full plan results from summing up all $V_a(k)$ that are covered by the respective plan.

Returning to the intuitive example given in the previous subsection, the correction terms would be of the form $\Delta V_a(k) = (y_a(k) - q_a(k))/\sigma_a^2(k)$. Again, the functioning of the calibration can be interpreted as a controller in that the utility of plans that improve the measurement reproduction is increased and the utility of plans that impair the measurement reproduction is decreased.

In congested conditions, the computation of the derivatives in Eq. (7) is more involved. Flötteröd et al. (2011) detail this logic based on Flötteröd and Bierlaire (2009), which essentially relies on a linear regression of the *actual* flow across a sensor against the number of vehicles that *intend* to cross that sensor.

As described in Section 2, MATSim functions in two phases, where the first phase builds the choice sets and the second phase simulates the choices based on fixed choice sets. Important from a calibration perspective, plans that are newly generated during the first phase are immediately chosen for execution in the mobility simulation in order to assess their performance. The utility-driven estimator (7) is applied in either phase in the following way:

- During the first phase, a newly generated plan is always selected. If no new plan is generated, then an available plan is selected according to Eq. (7).
- During the second phase, no new plans are generated and the calibrated choice distribution (7) is always employed.

This means that the calibration takes full effect only after the choice set generation is turned off.

Finally, it should be mentioned that Cadyts introduces in the setting described above an almost negligible computational overhead over a plain simulation of the same scenario. The respective performance measures can be found in Flötteröd et al. (2011).

3.3 Scope of the calibration

At this point, a clarification of the scope of the proposed calibration of MATSim using Cadyts is in order. So far, the notion of "calibration" has been used intentionally loosely and without a delineation from a more concrete terminology such as "parameter calibration", "parameter estimation" or "state estimation".

Section 3.2 describes how the simulated travel behavior in MATSim is adjusted to traffic counts through additive modifications of the utility functions, assuming a logit choice model. This approach can be applied more generally for every multivariate extreme value (MEV) choice model because every such model can be phrased in logit form (with a term involving the generating function added to the utility; see, e.g., Ben-Akiva and Lerman 1985). This approach is equivalent to the calibration of an alternative specific constant (ASC) of every single alternative of every traveler.

Adopting this perspective, the calibrated simulation system still solves the original fixed-point problem of attaining consistency between the demand model and the supply model, however, based on calibrated ASCs. The more general formulation (4) does not even require a utility-driven demand model; in this case, the calibrated simulation system deviates from the fixed point formulation of the original demand/supply model in a way that leads to greater consistency with the sensor data.

As described so far, the approach does not calibrate structural model parameters beyond ASCs. For that purpose, one could start by adopting a two-stage approach: In the first stage, Cadyts identifies changes to the utility function values that improve consistency with the traffic counts. In the second stage, these utility changes are then exploited in order to conclude about possible improvements of further structural model parameters. The remainder of this article, in particular Section 5, exemplifies various opportunities along these lines.

However, although the ex post analysis of the utility corrections provided by Cadyts is a insight- and useful exercise, it clearly is desirable to adopt a one-stage approach where structural model parameters are calibrated directly. Although this opportunity has not yet been systematically explored, some theoretical results are by now available that show the feasibility of a direct parameter estimation of choice model coefficients beyond ASCs (Flötteröd 2010). The reminder of this subsection briefly outlines this concept.

The original Cadyts approach (4) results from the maximization of a *posterior entropy function* that essentially represents the plausibility of the simulated travel behavior of all agents given the measurements. Mathematically, this approach can be directly re-phrased as a parameter estimation problem by maximizing the posterior entropy function with respect to its structural parameters. An implementation of this approach has already been made available (Flötteröd 2011), however, its experimental investigations are still ongoing. An apparently highly relevant issue in this context is the consistent treatment of sampled choice sets: While it is well-known in discrete choice theory that this sampling needs to be corrected for when estimating choice model parameters (Ben-Akiva and Lerman 1985), a consistent correction of this type in the context of path flow or OD matrix estimation appears to have not yet been discussed in the literature.

4 Zurich field study

This section describes a real-world case study for the city of Zurich. The setting of the test case is presented and some selected calibration results from a previous study are recalled (Flötteröd et al. 2009, 2011). The utility offsets obtained from this calibration are analyzed in Section 5. This novel analysis shows that the utility corrections, which originally result from a formal solution of the calibration problem, have not only an intuitive meaning but also enable further demand analyses and calibrations.

We consider the Greater Zurich region in Switzerland; the case study network consists of a subset of an all-of Switzerland network with more than 60,000 links (Chen et al. 2008). Figure 1 shows the analysis zone. The synthetic population generated for the study region consists of more than 187,000 agents, which constitutes a random 10% sample of the full population that travels, at

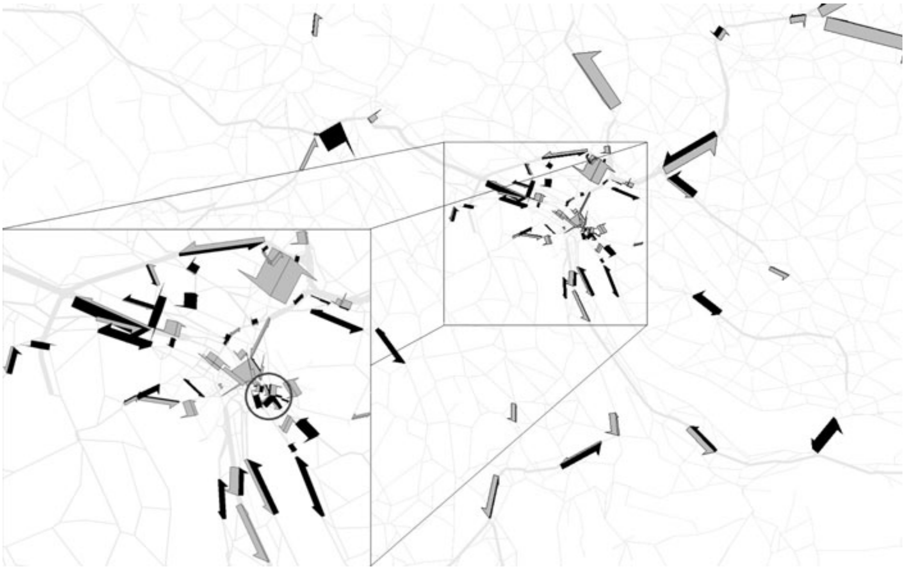

Fig. 1 Spatial layout of the link-based utility offsets at 8–9 am. *Dark* Counts are too high, negative utility offsets try to discourage traffic. *Light* Counts are too low, positive utility offsets try to encourage additional traffic. Width corresponds to the magnitude of the utility offset

any time during the considered 24-h period, within a 30-km circle around the center of the study region. All travelers have complete daily activity patterns based on microcensus information (SFSO 2006). Such activity patterns can include activities of type *home*, *work*, *education*, *shopping*, and *leisure*. The typical durations for those activities are derived from the microcensus data and are specified individually for each member of the synthetic population.

The choice dimensions of all agents are route choice, departure time choice, and mode choice. Table 1 shows the parameters used in the scenario. Activity locations are given opening and closing times in order to keep the agents within some timely limit. The opening and closing times are classified by activity type, i.e., the opening and closing times are distinguished for home, work, education, shop, and leisure activities. There is not yet any distinction based on the location of an activity. For simplicity, a physical network simulation of public transport is replaced by a "teleportation mode" that moves travelers on public transport trips at half the speed of a car in uncongested conditions (Grether et al. 2009; Rieser et al. 2009). This fairly simplistic approach was chosen due to the lack of a proper public transport simulation in MATSim, which, however, will be available in the near future (Rieser 2010).

For calibration purposes, traffic counts from 161 inductive loop sensor stations are available. This data is used in the following way. First, the scenario is simulated with MATSim alone, without using the traffic counts. The results of this "base case" simulation are then compared to the traffic counts. Second, MATSim is run jointly with the calibration in different settings that use

Table 1 Simulation parameters

Parameter	Value
β_{perf} (activity coefficient in Eq. (1))	+12 Eur/h
β_{car} (cost of car travel)	−12 Eur/h
$\beta_{\text{non-car}}$ (cost of non-car travel)	−6 Eur/h
Size of plan choice set	4
Total number of iterations	500
Iterations for choice set generation	300
Min./avg./max. home duration	0.5/14.7/23.0
Min./avg./max. work duration	0.5/6.1/20.0
Min./avg./max. education duration	0.5/5.8/20.0
Min./avg./max. shop duration	0.5/1.7/12.0
Min./avg./max. leisure duration	0.5/2.6/20.0
Home opening/closing time	00:00/24:00
Work opening/closing time	07:00/18:00
Education opening/closing time	07:00/18:00
Shop opening/closing time	08:00/20:00
Leisure opening/closing time	00:00/24:00

one subset of the traffic counts for calibration and the remaining counts for validation. Table 2 gives an overview of the results, which are described below.

The first data column of Table 2 ("reproduction MWSE") compares the measurement data fit of a plain simulation without calibration to that of a simulation where the calibration uses all available measurements at once. The MWSE ("mean weighted square error") shown here is the average quadratic deviation between simulated and observed counts at all sensor stations and in all time steps. All terms in this sum are weighted with one over two times the measured value; this corresponds to the assumption of independently normally distributed measurements with variances equal to the measurements. Table 2 shows that the reproduction MWSE is reduced by 80% through the calibration, which indicates an excellent adjustment to the data.

The second data column of Table 2 shows cross-validation results that were obtained by (i) splitting the sensors in ten disjoint subsets, (ii) running ten calibrations based on the data from nine subsets each, and (iii) comparing each calibration result to the unused sensor data set. A global improvement of almost 30% is obtained.

We stress that the fact that the validation improvement of 28% is lower than the reproduction improvement of 80% is *not* a sign of overfitting: The calibration adjusts directly only the behavior of those agents that may travel across sensors. The behavior of all other agents is implicitly changed through interactions with the immediately adjusted agents in the network (congestion feedback). Having a lower validation improvement than reproduction

Table 2 Simulation and estimation results

	Reproduction MWSE	Validation MWSE
Plain simulation	103.6	103.6
Calibrated simulation	20.9	75.1
Relative difference (%)	−80	−28

improvement indicates that the number of sensor locations is insufficient to "reach" the entire agent population in the calibration—some agents travel simply too far away from the sensors to be meaningfully adjusted. (The same observation holds for OD matrix estimators, which adjust only those OD flows directly that go across sensors.) In summary, rather than pulling only the simulated flows at the sensor locations towards the measurements while ignoring everything else, the calibration pulls the *whole system* towards a more realistic state.

5 Analysis of plan utility offsets

The ability of the Cadyts calibration system to adjust simulated behavior at the level of individual travelers enables an analysis at the fully disaggregate level. This section demonstrates how the utility corrections generated by Cadyts can be used for the further analysis of virtually arbitrary demand segments. The important advantage of this approach over what one could do based on OD matrices is that the definition of a demand segment can be made *after* the simulation/calibration is conducted. This flexibility inevitably gets lost in any approach that aggregates the demand *prior* to the simulation/calibration.

5.1 Direct inspection of utility offsets

One can plot the link- and time-additive correction terms $\Delta V_a(k)$ from Eq. (7); results look like in Fig. 1. From such plots, investigated over all hourly time slices, one obtains the following insights:

- Cadyts compensates for overall bias; i.e., it adjusts the rhythm of daily demand to the counts: Fig. 2 shows the average hourly bias (simulated minus measured counts) over all sensors before the calibration, the average effect of the calibration over all sensor links (all other links have zero utility offsets), and the hourly bias after the calibration. Clearly, the calibration counteracts the bias: The utility corrections are the more positive (i.e., encouraging traffic) the more negative the bias is (i.e., the simulated counts are lower than the measured counts).
  In contrast to other approaches, demand is not considered as fully elastic, but it can be moved between time slices. This is possible only because in MATSim travelers possess different plans with different time structures *and* Cadyts is designed to take advantage of that feature. However, if the demand was elastic, e.g., in that there was a "stay-at-home" plan, then this elasticity would be exploited by Cadyts as well.
- Cadyts compensates for a directional bias; i.e., it reduces regular commuting and increases reverse commuting. This is already visible in Fig. 1, but it will become more evident in the subsequent analysis.
- Cadyts attempts to compensate for a systematic over-prediction in an east-west corridor at the lake (circle in Fig. 1). This feature is visible across all

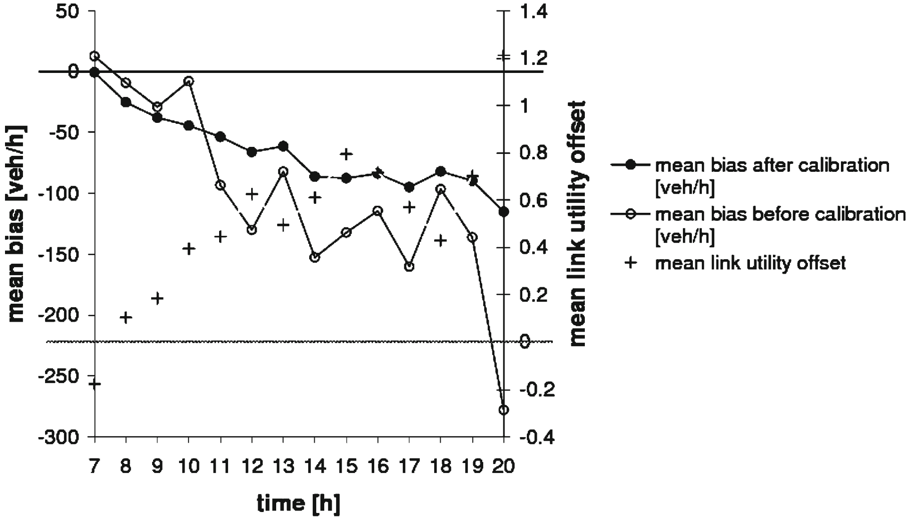

Fig. 2 Mean counts bias and utility correction as a function of time. The counts bias is computed as the mean value of simulated minus measured counts at all sensor locations

time slots. It is, presumably, a network error in the sense that the links possess too much capacity in the simulation.

This is likely to bias the demand estimation results in that the demand is adjusted in an attempt to correct for a supply error. This type of error can be avoided by jointly estimating the demand side and the supply side of the simulation; this is an important topic of future research.

- As a tendency, the corrective signal is the stronger the lower the density of counting stations. This is plausible since with a high density of counting stations several counting stations can collaborate to correct traffic into the desired direction.

5.2 Trip generation/attraction maps

Equation (7) maps the link-based utility corrections on all-day travel plans. This allows to analyze the effect of the calibration on arbitrary demand segments (by considering only the respective subsets of the population) or on arbitrary demand dimensions (e.g., only route choice between two certain regions within a certain time interval). In the following, we analyze the utility corrections that persist after the convergence of the calibrated simulation.

We first adopt a trip-based perspective in that we extract from the agent-based demand model only the trips that fall into the morning rush hour. For each trip, we compute the utility correction according to Eq. (7). We then plot the resulting information in two ways on a map of Zurich, cf. Figs. 3 and 4.

Both plots are generated by putting a 1 km times 1 km grid over the analysis region. In Fig. 3, the shades/patterns of the cells represent the average

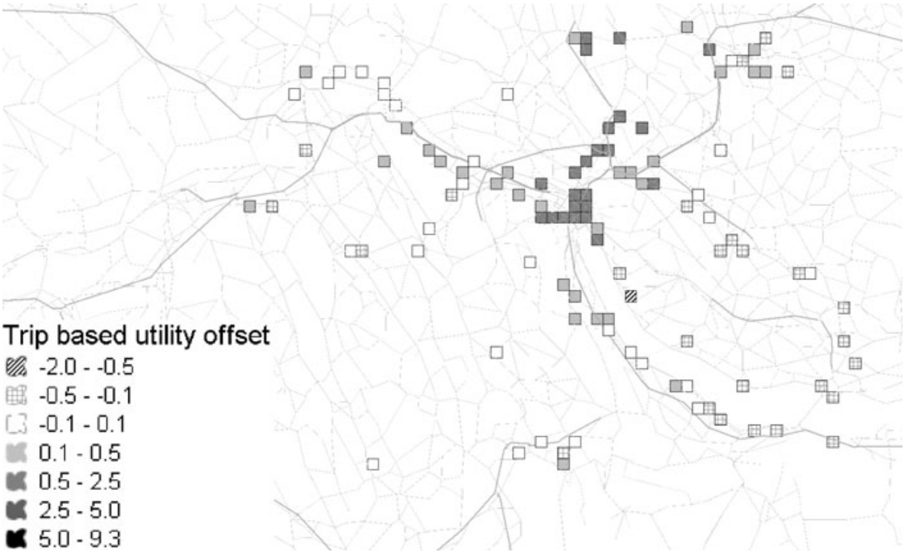

Fig. 3 Spatial distribution of utility corrections for trips generated between 8 and 9 am. Only gridcells with at least 50 generated trips are shown

utility corrections of all trips starting between 8 and 9 am in the respective cell, whereas in Fig. 4 these shades/patterns corresponds to the average utility correction of all trips ending between 8 and 9 am in the respective cell.

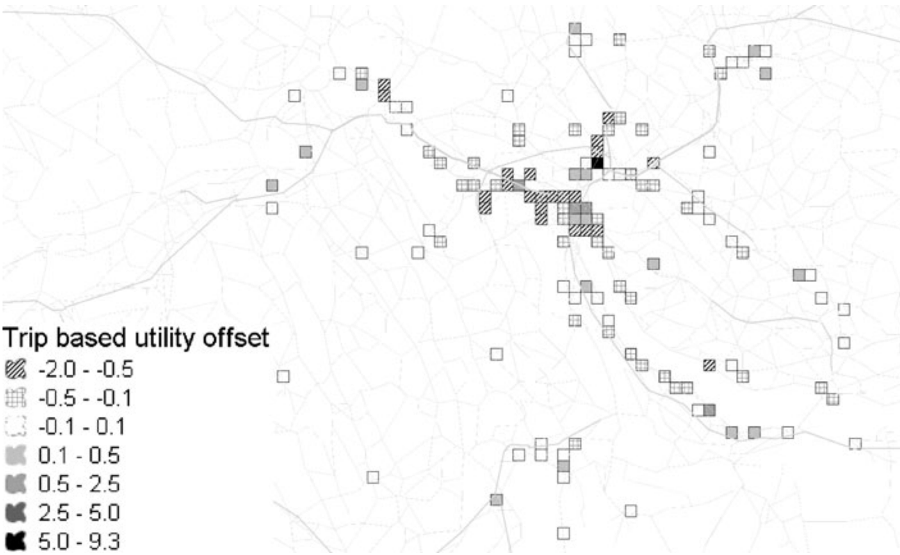

Fig. 4 Spatial distribution of utility corrections for trips attracted between 8 and 9 am. Only gridcells with at least 50 attracted trips are shown

Figure 3 (trip generation) shows positive trip utility offsets for trips originating in the city center, and negative trip utility offsets for trips originating in the surroundings. This can be interpreted as having not enough trip generation between 8 and 9 am in the city center, and having too much trip generation in the surroundings.

Figure 4 (trip attraction) shows negative trip utility offsets for trips arriving in most of the center, while a small area has positive offsets. This area contains the historical city center, the train station, and important parts of two universities. Offsets in some of the far-away surroundings are positive again. This can be interpreted as having too many trips arriving in most of the city center, while there are not enough arrivals in the indicated small area. At the same time, there are not enough arrivals in parts of the surroundings. However, the following analysis shows that the trip-based results described so far need to be taken with great care.

Now we turn to the exploitation of a feature that is unavailable in a purely trip-based (OD matrix driven) traffic simulation: We analyze the *all-day* utility offsets of the *all-day* plans that correspond to the previously described trips.

Figure 5 shows the plan-based counterpart of Fig. 3, i.e., the utility offsets of the entire plans that contain a trip that starts between 8 and 9 am in the depicted gridcells. One observes a qualitatively similar pattern with a somewhat higher overall level of the corrections, which results from the fact

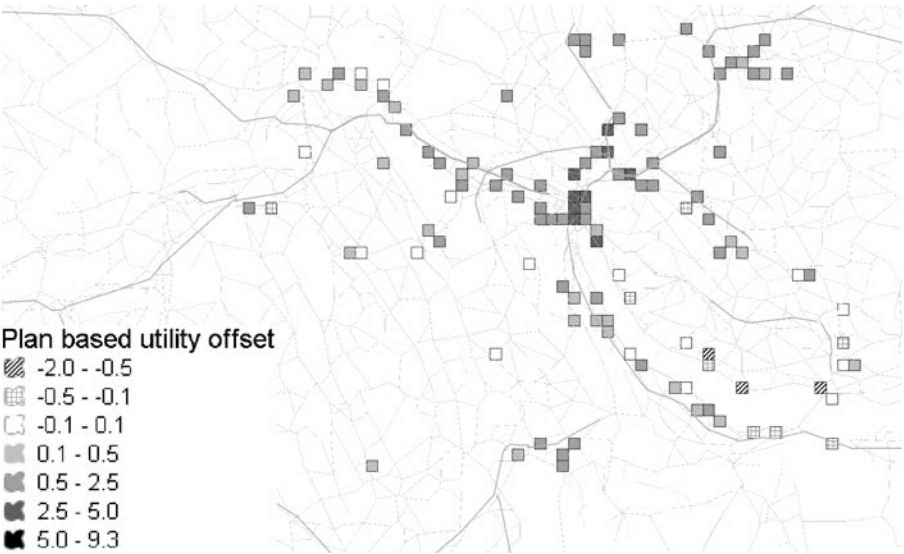

Fig. 5 Spatial distribution of utility corrections for all-day travel plans that have each at least one trip generated between 8 and 9 am. Only gridcells with at least 50 generated trips are shown

that the corrections are now summed up along a whole day (and not just 1 h). Overall, the plan-based perspective confirms the trip-based analysis.

Figure 6 shows the plan-based counterpart of Fig. 4, i.e., the utility offsets of the entire plans that contain a trip that ends between 8 am and 9 am in the depicted gridcells. Here, a striking difference between the plan-based and the trip-based perspective can be observed. Most importantly, the negative utility offsets in the trip-based perspective that discourage travel towards the city center turn into positive utility offsets in the plan-based perspective that encourage travel. Also, the slightly negative trip utility offsets in the city surroundings turn into mostly clearly positive values in the plan based perspective. This difference is explained in the following.

The analysis of all-day plans instead of separate trips allows to account for the dynamical constraints that guide real travel: Behaviorally, it is well known that travelers choose between trip sequences and not between individual trips. Physically, the mass conservation of persons and vehicles must be accounted for. A first conclusion of the comparison between Figs. 4 and 6 is that the negligence of these constraints can lead to drastic misinterpretations.

Regarding the concrete values shown in Figs. 4 and 6, one can conclude that the *trips* ending in the city center between 8 am and 9 am are not the result of an overall demand surplus, but only the result of a demand misallocation, possibly due to imprecise destination or departure time choice modeling (see below): the calibration actually encourages *plans* that end in

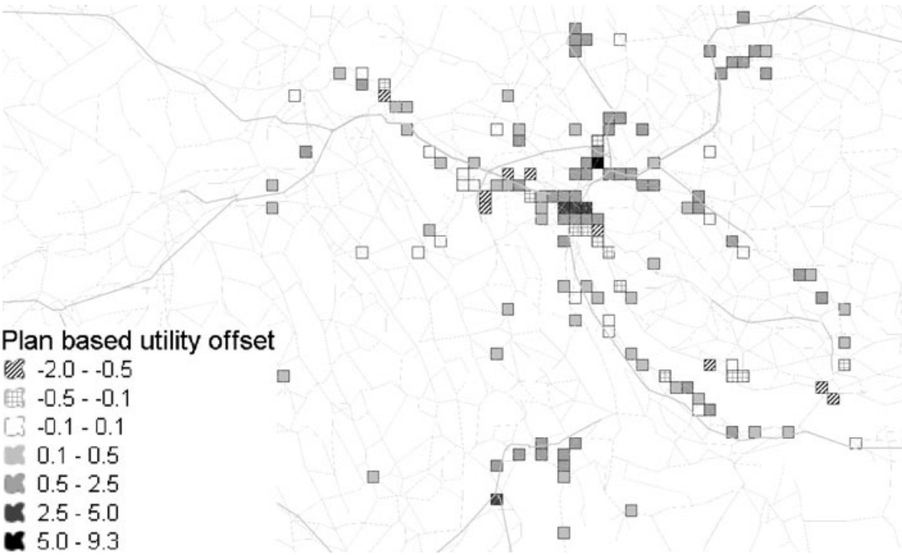

Fig. 6 Spatial distribution of utility corrections for all-day travel plans, which have each at least one trip attracted between 8 and 9 am. Only gridcells with at least 50 attracted trips are shown

the city center between 8 am and 9 am, which is consistent with the general demand underestimation in the simulation as shown in Fig. 2.

The completely different picture in the trip-based perspective may be due to (i) errors in the choice model specification and (ii) errors in the attributes fed into the choice model.

- Choice model specification errors are very likely to be present in the given scenario: The simple multinomial logit plan choice model ignores correlation across alternatives. The choice model coefficients are not estimated from data but inferred on a trial-and-error basis. (As mentioned before, onging work indicates that the latter error source will soon be removed in that the calibration also adjusts choice model parameters (Flötteröd 2010)).
- Errors in the attributes fed into the choice models are likely to exist as well. Perhaps most noteworthy is the assumption of identical opening and closing times for all facility types, cf. Table 1. This is likely to result in an unrealistic morning peak concentration that would be smoothed out by more distributed starting times of, in particular, the work activity.

A more detailed analysis of these maps is the topic of ongoing research and scenario refinements for the city of Zurich. The analysis given here already demonstrates clearly that (i) utility offsets computed from traffic counts can be used for an insightful spatio-temporal demand analysis and that (ii) the new approach of calibrating a fully disaggregate demand of individual travelers can lead to completely different (and structurally far more meaningful) results than what an estimation of independent OD matrices per time slice suggests.

5.3 Identification of underestimated demand segments

This subsection presents an exemplary analysis of how the utility corrections generated by Cadyts can be used to identify demand segments that are likely to be corrupted by modeling errors.

For this, we analyze the travel demand by purpose, where we distinguish trips that head for work, education, shopping, leisure or home, or belong to the "border-crossing" demand segment. Figure 7 shows histograms of the offsets by purpose, with a uniform histogram bin size of 0.25 and accounting only for such trips that cross a sensor at least once (all other trips would do nothing but add a peak at a zero utility correction to the histogram).

The histograms reveal a striking difference between the trips for border-crossing and all other travel purposes. While all other trips are quite symmetrically centered around an almost zero utility offset, the border-crossing trips are much more widely scattered around a mean of approximately +10 utility units.

This means that Cadyts strongly encourages border-crossing traffic but is on average almost indifferent with respect to the other demand segments. This indicates that the border-crossing demand is substantially underestimated in

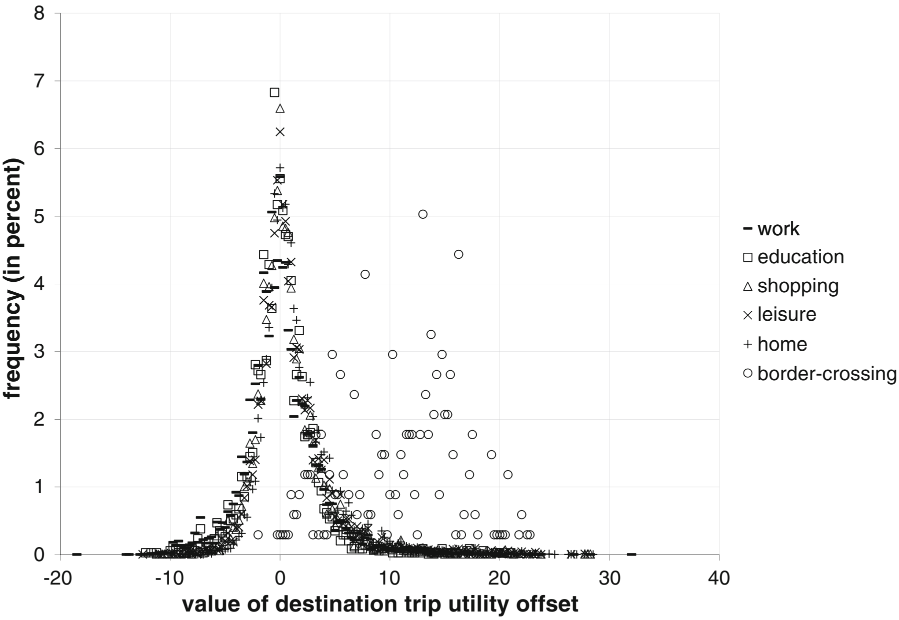

Fig. 7 Histogram of trip utility offsets by purpose

the synthetic population of the Zurich scenario. This observation motivated a re-examination of the demand modeling of this scenario, which indeed revealed an inconsistency: The initial demand contains, statistically, all trips generated by persons living in Switzerland, plus all trips generated by vehicles crossing the borders of Switzerland. As a result, all border-crossing traffic by Swiss drivers is, statistically, counted twice, while non-border-crossing traffic by non-Swiss drivers is missing. It is plausible to assume that, 50 km away from the border, the second segment is larger than the first, and that the second segment mostly comprises of through traffic, which looks somewhat similar to the border-crossing traffic. Here, the calibration has revealed a structural incompleteness in the demand modeling that should be corrected for in future work.

The wide histogram scatter of the utility corrections for border-crossing traffic can in part be explained with the relatively low total number of border-crossing travelers simulated, which naturally leads to a higher variability in the histogram. However, the wide scatter of utility values may also indicate that a further disaggregation of this demand segment is necessary. This is quite plausible given the above observation that the initial demand modeling in some sense compensates for one demand segment through another. We leave the further analysis of these details to future studies.

In summary, this section demonstrates that the utility corrections computed by Cadyts for every single synthetic traveler can be utilized for an ex post analysis of the simulation system in various ways. It needs to be stressed again that the manual/visual inspection conducted here has by no means pushed this approach to its limits: A logical next step is to utilize the utility corrections not only for the calibration of the plan choice patterns in a given population but also for an adjustment of the size of the different demand segments within that population.

6 Discussion and summary

A standard question in conjunction with calibration is in how far the results are useful for prediction. Based on the results of the last sections, one can argue that the results are useful for short-term prediction: both in a real-time setting or for a short-term policy measure, the link offsets could be frozen and then used in the prediction. As discussed in Flötteröd (2008), care needs to be taken that the offsets are only used for choice and not for choice set generation, i.e., not for routing.

Clearly, this approach runs into problems when anything in the system that is presumably related to the link offsets changes. A simple example would be the addition of a lane to such a link. For such situations, a calibration of "higher level" behavioral parameters is necessary. We are currently investigating two approaches:

- Calibration of the parameters of the utility function, such as $\beta_{\text{non-car}}$, from traffic counts and supplementary observations (Flötteröd 2010).
- Calibration of location choice, in particular "secondary" activity location choice. This would directly correspond to OD matrix estimation in the four-step procedure, except that it would calibrate full daily plans.

Apart from the calibration of utility functions, an analysis of the utility offsets reveals further calibration opportunities. Since the plan-specific utility offsets can be interpreted as encouragements (when positive) or discouragements (when negative) of the respective travel behavior, the total levels of arbitrary demand segments can be analysed in hindsight. While this article only indicates this opportunity through the analysis of selected demand segments in a single scenario, it appears feasible to develop a calibration method that also corrects such inconsistencies in a statistically consistent manner.

In summary, this article demonstrates that a fully disaggregate transport microsimulation that represents travel demand at the level of individual persons can be applied to the realistic simulation of large metropolitan systems. The agent-based simulation goes beyond traditional transport models in that it equilibrates not only route choice but all-day travel behavior, including departure time choice and mode choice. A novel calibration method is applied to the calibration of the microscopic travel demand from traffic counts. The method does not only generate a clear improvement in measurement and validation data fit, it also adjusts the demand in a behaviorally interpretable way. It does so by computing utility corrections to which the utility-driven travel demand simulator reacts with more realistic behavior. A detailed analysis of these utility corrections clarifies their behavioral interpretation, shows ways in which they can be applied for demand analysis, and indicates possibilities for their further exploitation in the automatic calibration of disaggregate travel demand models.

References

Avineri E, Prashker J (2003) Sensitivity to uncertainty: need for paradigm shift. Transp Res Rec 1854:90–98

Balmer M, Cetin N, Nagel K, Raney B (2004) Towards truly agent-based traffic and mobility simulations. In: Autonomous agents and multiagent systems (AAMAS'04). New York, NY. URL: www.aamas-conference.org

Ben-Akiva M, Lerman S (1985) Discrete choice analysis. MIT Press series in transportation studies. MIT Press, Cambridge

Bhat C, Guo J, Srinivasan S, Sivakumar A (2004) A comprehensive econometric microsimulator for daily activity-travel patterns (cemdap). Transp Res Rec 1894:57–66

Bowman J, Ben-Akiva M (1998) Activity based travel demand model systems, In: Marcotte P, Nguyen S (eds) Equilibrium and advanced transportation modelling. Kluwer, Dordrecht, pp 27–46

Bowman J, Bradley M, Shiftan Y, Lawton T, Ben-Akiva M (1998) 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, pp 171–184

Cetin N, Burri A, Nagel K (2003) A large-scale agent-based traffic microsimulation based on queue model. In: Proceedings of Swiss transport research conference (STRC), Monte Verita, CH. See www.strc.ch. Earlier version, with inferior performance values: Transportation Research Board Annual Meeting 2003 paper number 03-4272

Charypar D, Nagel K (2005) Generating complete all-day activity plans with genetic algorithms. Transportation 32(4):369–397

Chen Y, Rieser M, Grether D, Nagel K (2008) Improving a large-scale agent-based simulation scenario, VSP working paper 08-15, Transport Systems Planning and Transport Telematics Laboratory, Berlin Institute of Technology

de Palma A, Marchal F (2002) Real case applications of the fully dynamic METROPOLIS toolbox: an advocacy for large-scale mesoscopic transportation systems. Netw Spat Econ 2(4):347–369

Flötteröd G (2008) Traffic state estimation with multi-agent simulations. PhD thesis, Berlin Institute of Technology, Berlin, Germany

Flötteröd G (2009) Cadyts—a free calibration tool for dynamic traffic simulations. In: Proceedings of the 9th Swiss transport research conference. Monte Verita/Ascona, Switzerland

Flötteröd G (2010) A general methodology and a free software for the demand calibration of DTA models, In: Proceedings of the 3rd international symposium on dynamic traffic assignment, Takayama, Japan

Flötteröd G (2011) Cadyts web site. http://transp-or.epfl.ch/cadyts. Accessed 22 July 2011

Flötteröd G, Bierlaire M (2009) Improved estimation of travel demand from traffic counts by a new linearization of the network loading map. In: Proceedings of the European transport conference. Noordwijkerhout, The Netherlands

Flötteröd G, Chen Y, Rieser M, Nagel K (2009) Behavioral calibration of a large-scale travel behavior microsimulation. In: proceedings of 12th international conference on travel behaviour research. Jaipur, India

Flötteröd G, Bierlaire M, Nagel K (2011) Bayesian demand calibration for dynamic traffic simulations. http://transci.journal.informs.org/cgi/content/abstract/trsc.1100.0367v1

Gawron C (1998) Simulation-based traffic assignment. PhD thesis, University of Cologne, Cologne, Germany. URL: www.zaik.uni-koeln.de/AFS/publications/theses.html

Grether D, Chen Y, Rieser M, Nagel K (2009) Effects of a simple mode choice model in a large-scale agent-based transport simulation. In: Reggiani A, Nijkamp P (eds) Complexity and spatial networks. In search of simplicity, advances in spatial science, chapter 13. Springer, New York, pp 167–186

Jonnalagadda N, Freedman J, Davidson W, Hunt J, (2001) Development of microsimulation activity-based model for San Francisco: destination and mode choice models. Transp Res Rec 1777:25–35

Lin DY, Eluru N, Waller S, Bhat C (2008) Integration of activity-based modeling and dynamic traffic assignment. Transp Res Rec 2076:52–61

MATSIM (2011) MultiAgent Transport SIMulation. URL: http://matsim.org/. Accessed 22 July 2011

Meister K, Rieser M, Ciari F, Horni A, Balmer M, Axhausen K (2008) Anwendung eines agentenbasierten Modells der Verkehrsnachfrage auf die Schweiz. In: Proceedings of Heureka '08. Stuttgart, Germany

Nagel K, Flötteröd G (2012) Agent-based traffic assignment: going from trips to behavioral travelers. In Bhat C, Pendyala R (eds) Travel behaviour research in an evolving world. Emerald Group Publishing, Bingley (forthcoming)

Ortuzar J, Willumsen L (2004) Modelling transport. Wiley, New York

Pendyala R (2004) Phased implementation of a multimodal activity-based travel demand modeling system in Florida. Volume II: FAMOS users guide. Research report, Florida Department of Transportation, Tallahassee. See www.eng.usf.edu/~pendyala/publications

Raney B, Nagel K (2006) An improved framework for large-scale multi-agent simulations of travel behavior. In: Rietveld P, Jourquin B, Westin K (eds) Towards better performing European transportation systems. Routledge, London, pp 305–347

Rieser M (2010) Adding transit to an agent-based transportation simulation—concepts and implementation. PhD thesis, Technical University of Berlin

Rieser M, Nagel K, Beuck U, Balmer M, Rümenapp J (2007) Truly agent-oriented coupling of an activity-based demand generation with a multi-agent traffic simulation. Transp Res Rec 2021:10–17

Rieser M, Grether D, Nagel K (2009) Adding mode choice to a multi-agent transport simulation. Transportation Research Record: Travel Demand Forecasting 2132:50–58

SFSO (2006) Ergebnisse des Mikrozensus 2005 zum Verkehr. Swiss Federal Statistical Office, Neuchatel

Vovsha P, Bradley M, Bowman J (2004) Activity-based travel forecasting models in the United States: progress since 1995 and prospects for the future. In: Proceedings of the EIRASS conference on progress in activity-based analysis. Maastricht, The Netherlands

Zhou X, Mahmassani H, Zhang K (2007) Dynamic micro-assignment modeling approach for integrated multimodal urban corridor management. Transportation Research Part C: Emerging Technologies 16(2):167–186