Behavioral calibration of a large-scale travel behavior microsimulation

Gunnar Flötteröd (corresponding author)
Transport and Mobility Laboratory
Ecole Polytechnique Fédérale de Lausanne
1015 Lausanne, Switzerland
Email: gunnar.floetteroed@epfl.ch

Yu Chen
Transport Planning and Transport Telematics
Technical University of Berlin (TUB)
10623 Berlin, Germany
Email: chen@vsp.tu-berlin.de

Kai Nagel
Transport Planning and Transport Telematics
Technical University of Berlin (TUB)
10623 Berlin, Germany
Email: nagel@vsp.tu-berlin.de

approx. 6000 words (estimated from Latex code) + 4 figures + 2 tables

August 31, 2010

submitted for presentation only

ABSTRACT
This article reports on the application and calibration of a fully disaggregate (agent-based) transport simulation for the metropolitan area of Zurich. The application of a novel calibration technique yields cross-validation results that are competitive with any state-of-the-art four-step model. The added value of the proposed modeling/calibration approach is that the transport simulation equilibrates not only route choice but all-day travel behavior, which is in its entirety calibrated from traffic counts.

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 (1). 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 (2).

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, (3, 4, 5, 6), 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 (7).

This type of coupling between the ABDM and the traffic assignment leaves room for improvement (8, 9). 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 traffic flow microsimulation, 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 (also see (10, 11)).

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 destination, 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, using the metropolitan area of Zurich as an example (as a sub-region of an "all-of-Switzerland" scenario (12)). The results 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 % to 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.

The quality of the presented results is to a large extent due to new methodological advances on the calibration side: Until recently, the 4-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 present the first real-world application of 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 (13, 14). The article presents cross-validation results that confirm that the calibration does not simply "drag" the demand towards a good measurement fit but indeed realizes meaningful structural demand adjustments.

The remainder of this article is organized as follows. Sections 2 describes the used microsimulation, and Section 3 drafts the principles of the deployed demand calibration tool. The field study is described in length in Section 4. Section 5 details the mechanisms through which the calibration takes effect, and Section 6 discusses the approach. Finally, Section 7 summarizes the article.

2. OUTLINE OF TRANSPORT MICROSIMULATION

The MATSim ("Multi-agent transport simulation toolkit", (15, 16)) transport microsimulation is used for the purposes of this study. 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 in our simulation. 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 choose the plan with the best performance, but they sometimes re-evaluate inferior plans, and they sometimes obtain new plans by modifying copies of existing plans.

The following subsections explains 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 computational intractable because of the sheer number of possible alternatives. 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.

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. Note that 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 (17). 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:

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

  where $t_{perf,a}$ 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. 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.
- The (dis)utility $V_{travel}(l)$ 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.

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

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

It is important to note that the score thus takes into account the complete daily plan. More details can be found in (16, 18).

The plan choice is modeled with a multinomial logit model (which calls for enhancements in the future) (19). The choice model has one additional twist during the choice set generation phase: If it happens that an agent receives a newly generated plan from one of the aforementioned plan generation modules, then this plan 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 (20, 21): 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 this link.

3. OUTLINE OF CALIBRATION

The previous section describes a simulation system 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" (14, 22)) 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.

3.1. Basic functioning

Cadyts makes no assumptions about the form of the plan choice distribution (3) 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 sense.

Assuming (only for the sake of an intuitive formulation) congestion to be light and the traffic counts to be independently normal distributed, the posterior choice distribution can be shown to be approximately of the following form (13):

$$P_n(i|\mathbf{y}) \sim \prod_{ak \in i} \exp\left(\frac{y_a(k) - q_a(k)}{\sigma_a^2(k)}\right) \cdot P_n(i) \tag{4}$$

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 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.

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

$$P_n(i|\mathbf{y}) \sim \exp\left(V(i) + \sum_{ak \in i} \frac{y_a(k) - q_a(k)}{\sigma_a^2(k)}\right) =: \exp\left(V(i) + \sum_{ak \in i} \Delta V_a(k)\right). \tag{5}$$

That is, an implementation of the posterior choice distribution requires nothing but to add link- and time-additive correction terms $\Delta V_a(k)$ to the utility of every considered plan. 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.

As described in Section 2, MATSim functions in two phases, where the first phase builds the choice set 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 (5) 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 (5).
- During the second phase, no new plans are generated and the calibrated choice distribution (5) is always employed.

4. ZURICH FIELD STUDY

This section describes results from a real-world case study for the city of Zurich. First, the basic setting of the test case is presented in Section 4.1. Second, the interactions between simulation and calibration are investigated in Section 4.2. Finally, Section 4.3 discusses the validation results for the calibrated simulation system.

TABLE 1 Simulation parameters.

parameter	value	matsim key
$\beta_{perf.act.}$	12 Eur/h	`performing`
β_{car}	-12 Eur/h	`traveling`
$\beta_{non-car}$	-6 Eur/h	`travelingPt`
β_{scale}	1	`BrainExpBeta`
size of plan choice set	4	`maxAgentPlanMemorySize`
total number of iterations	500	
iterations for choice set generation	300	
home opening time	00:00	
home closing time	24:00	
work opening time	07:00	
work closing time	18:00	
education opening time	07:00	
education closing time	18:00	
shop opening time	08:00	
shop closing time	20:00	
leisure opening time	00:00	
leisure closing time	24:00	

4.1. Description of test case and uncalibrated simulation results

An all-of-Switzerland network with 60 492 links and 24 180 nodes is used. It is based on a Swiss regional planning network, which has been made ready for simulation purposes based on additional OpenStreetMap network data (23). For some intuition regarding the network, see Figure 3.

A synthetic population of travelers for all of Switzerland is available from a previous study (12, 24). All travelers have complete daily activity patterns based on microcensus information (25). Such activity patterns can include activities of type *home*, *work*, *education*, *shopping*, *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 initial demand used for the simulations is based on the aforementioned demand of whole Switzerland, but consists only of all agents who cross a 30 km (18.6 miles) circle around the center of Zurich at least once during their daily travel, including those agents who stay within that circle for the whole day. In order to obtain a higher computational speed, a random 10 % sample is chosen for simulation, which consists of 187 484 simulated travelers.

All agents iteratively adapt 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. Public transit is simulated as described in Refs. (26, 27), that is, it is assumed that it provides door-to-door connectivity at twice the car free speed travel times.

Hourly traffic counts from 161 inductive loop sensors are available for an entire day. The deviation between measured and simulated traffic counts is both graphically and quantitatively evaluated. For visual inspection, scatter plots such as those given in Figure 1 (left) are used. Every point represents one pair of measured/simulated traffic counts, where the measured value defines the x-coordinate and the simulated value defines the y-coordinate. If all measurements were perfectly reproduced by the simulation, all points would lie on the diagonal with slope one. Deviations from that diagonal signalize inconsistencies between measurements and simulations.

Figure 1 (left) shows results after 500 iterations of uncalibrated simulation. Most points are within an (admittedly loose) band of a factor of two in both directions, which indicates that the simulation captures the overall situation fairly well. However, there clearly is room for improvement.

A quantitative analysis of the measurement reproduction quality is conducted in terms of the mean relative error

$$\mathrm{MRE}(k) = \left\langle \frac{|y_a(i) - q_a(k)|}{y_a(k)} \right\rangle_a \tag{6}$$

where the average $\langle \cdot \rangle$ over all measurement locations a is evaluated separately for each hour k of the day, $y_a(k)$ is the measured volume on link a in hour k, and $q_a(k)$ is its simulated counterpart. Figure 2 (top) shows these values for the uncalibrated base case. The simulation deviates strongly from the reality during the night hours, i.e., from midnight until 6 am. However, during daytime the hourly MRE is consistently below 30 %. It needs to be stressed that these results are not intended to model the nightly conditions because the according travel demand has been deliberately ignored in this study.

4.2. Inserting the calibration into the simulation

According to Section 3.2, the calibration affects all utility-based choices in the simulation by modifying the utility according to (5). This applies to all choices but the selection of newly generated plans, which are always executed. This implies that these parts of the demand remain uncalibrated during the first iteration phase that builds the choice sets. Only in the second iteration phase, where stable choice sets are used, the calibration takes full effect.

The first data column of Table 2 ("reproduction MWSE error") compares the measurement data fit of a plain simulation without calibration to that of a simulation where the calibration takes effect. The used error measure is defined as

$$\mathrm{MWSE} = \left\langle \frac{(y_a(k) - q_a(k))^2}{2\sigma_a^2(k)} \right\rangle_{ak} \tag{7}$$

where $\sigma_a^2(k)$ is the variance assigned to the sensor data on link a in hour k. It is calculated as

$$\sigma_a^2(k) = 0.5 \cdot \max\{y_a(k), (25\,\text{veh/h})^2\}, \tag{8}$$

which also is the specification used in the application of (5). It reflects two considerations. First, there is the assumption that the variance of a measurement is proportional to the measured value. Second, the variance is limited to a minimal positive value, which ensures that very small measurements are not over-weighted and avoids numerical problems in the evaluation of (5) and (7). The particular numbers used in this specification have been obtained by trial-and-error. Because of the previously discussed underestimation of the nightly demand, only measurements from 6:00 to 19:59:59 (as from now called the analysis period) are used by the calibration and evaluated in (7).

TABLE 2 Simulation and estimation results.

	reproduction MWSE error	validation MWSE error
plain simulation	103.6	103.6
estimated simulation	20.9	75.1
relative difference	- 80 %	- 28 %

Table 2 shows that the reproduction MWSE error is reduced by 80%, which indicates an excellent adjustment to the data. This impression is visually confirmed by the scatterplots of Figure 1 (right), which are obtained from the last iteration of the calibrated simulation. A comparison with the uncalibrated scatterplots on the left shows a substantial improvement in measurement fit in that the data points are substantially more centered around the main diagonal. Figure 2 (bottom) shows that the calibration enforces a MRE that is consistently between 10 % and 15 % during the analysis period, which is a reduction by half. One can also see that the MRE is increased outside of the analysis period when compared to the uncalibrated case. This is likely to result from the omission of certain demand segments, which the calibration compensates for by "drawing" agents from outside of the analysis period through an adjustment of their departure times. From this, one can also conclude that a better all-day base demand outside of the analysis period is likely to improve the results within the analysis period as well.

Overall, the calibration generates a substantial improvement in measurement fit. However, this alone does not prove that the calibrated agent behavior becomes more realistic because there are many plausible and not-so-plausible combinations of plan choice distributions that reproduce the measurements equally well. The next section provides cross-validation results that indicate that the calibrated demand is indeed more realistic.

4.3. Cross-validation results

While the previous section demonstrates that the calibration greatly improves the measurement reproduction, this section demonstrates that it does so in a way that also improves the realism of the global traffic situation. This is an important issue that applies to demand calibration from traffic counts in general because this problem is highly under-determined, which implies that there is a large number of demand configurations that reproduce the traffic counts equally well. Cadyts resolves this under-determination by taking the choice logic that is implemented in the simulation system itself as the prior information about the demand. The traffic counts are then added to this information in order to obtain an improved posterior choice distribution.

For cross-validation, the 161 sensor locations are randomly assigned to ten disjoint **validation data sets** of roughly equal size. For each validation data set, there is a corresponding **measurement data set** that contains the traffic counts from all sensors that are not represented by the respective validation data set. For every measurement/validation data set pair, one calibration is conducted, where only the measurement data is made available to the calibration and the corresponding validation data is used to evaluate how well the calibrated demand generates a spatiotemporal extrapolation of the traffic counts.

The second data column of Table 2 gives the resulting cross-validation MWSE values ("valuncalibrated calibrated

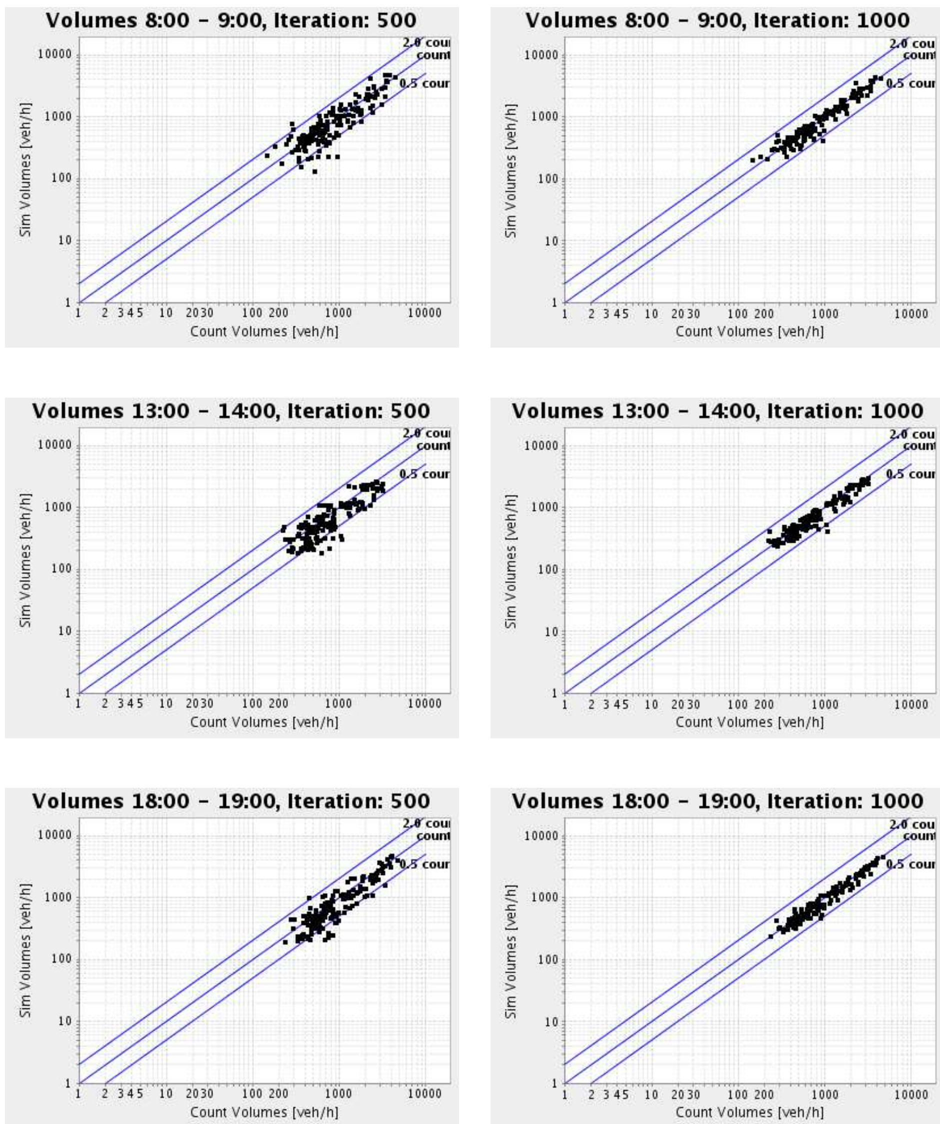

FIGURE 1 Scatter plots. Left: before the calibration. Right: after the calibration.

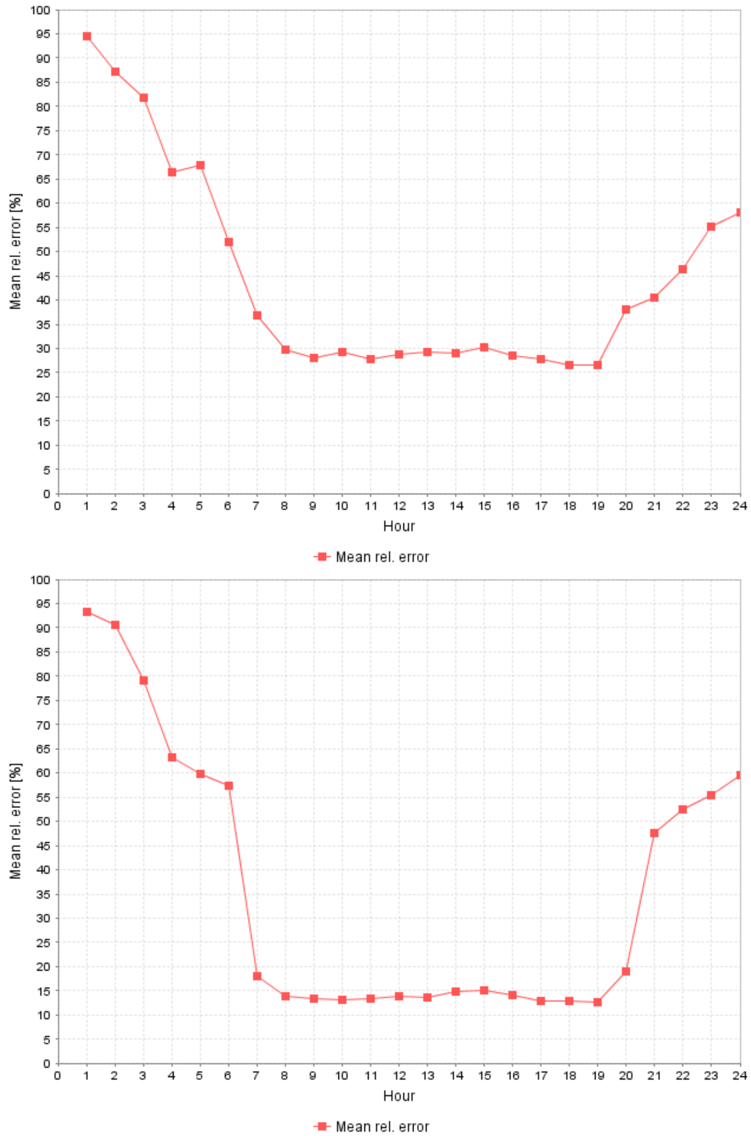

FIGURE 2 Top: Mean relative error (MRE) for uncalibrated base case. Bottom: Mean relative error (MRE) after calibration

idation MWSE error"), i.e., these numbers are derived from the measurements that were *not* included into the calibration. A global improvement of almost 30% is obtained. This indicates that the local information that is contained in the measurement data is used by the calibration in a way that changes the network-wide agent behavior such that more realistic network conditions result even far away from the sensor locations.

Note that the fact that the validation improvement of 30% 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 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.)

These results show that the calibration conducts demand modifications that are structurally meaningful in that they do not only fit the sensor data well but also lead to a global improvement in the system's realism. At this point, the difficulty of the calibration problem that is solved here needs to be stressed. The calibration adjusts simultaneously the route choice, mode choice, and departure time choice of hundreds of thousands of individual travelers in a purely simulation-based environment on a network with many ten thousand links. The number of iterations required to obtain stable and realistic results (500) is in the order of a plain simulation, and the computational overhead introduced by the calibration is below ten percent. All presented experiments were computed within less than 21 hours on a single computing node. The authors are not aware of any other calibration technique that comes close to such results.

5. SPATIAL STRUCTURE OF THE CORRECTIONS

One can plot the link- and time-additive correction terms $\Delta V_a(k)$ from (5); results look like in Figure 3. 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: Figure 4 shows the average hourly bias per sensor before the calibration, the average effect of the calibration per sensor link (all other links have offset zero), and the hourly bias after the calibration. Clearly, the calibration counteracts the bias, and, within the calibrated time period, the resulting bias is moved closer towards zero.

  In contrast to other approaches, demand is not considered as fully elastic, but it will be moved to other 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.

- Cadyts compensates for a directional bias; i.e. it reduces regular commuting and increases reverse commuting. This is also visible in Figure 3.

- Cadyts attempts to compensate for a systematic over-prediction in an east-west corridor at the lake (orange circle in Figure 3). This feature is visible across all time slots. It is,

FIGURE 3 Spatial layout of the induced link-based utility offsets at 8am–9am. Red: Counts are too high, trying to discourage traffic. Green: Counts are too low, trying to encourage additional traffic. Width corresponds to the strength of the signal.

FIGURE 4 Counts bias and utility correction as a function of time

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 station several counting stations can collaborate to correct traffic into the desired direction.

6. DISCUSSION

A standard question in conjunction with calibration is in how far the results are useful for prediction. Based on the results of the last section, 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 (13), 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 parameter would be useful. We are currently investigating two approaches:

- Calibration of the parameters of the utility function, such as $\beta_{non-car}$.
- 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.

7. 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. Crucial to the quality of the simulation is a proper calibration of the demand, for which traffic counts are shown to be a valuable data source. In particular, traffic counts from 161 sensors are used in a novel calibration methodology to adjust the route choice, mode choice, and departure time choice of hundreds of thousands of individual travelers on a network with many ten thousand links. The calibrated simulation system is successfully evaluated by cross-validation.

Future work will concentrate on the following items:

- Ongoing improvements of the Zurich base case with respect to all modeling aspects.
- Extension of the calibration system to the identification of structural demand parameters.

Finally, it should be mentioned that the deployed Cadyts calibration tool is not constrained to the MATSim microsimulation but is designed to be compatible with a wide variety of transport simulation systems.

ACKNOWLEDGMENTS

An earlier version of this paper, without Sections 5 and 6, was presented at IATBR'09 in Jaipur (India).

REFERENCES

[1] Ortuzar, J. and L. Willumsen, *Modelling Transport*. Wiley, 2004.

[2] Vovsha, P., M. Bradley, and J. Bowman, 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, 2004.

[3] Bhat, C., J. Guo, S. Srinivasan, and A. Sivakumar, A Comprehensive Econometric Microsimulator for Daily Activity-Travel Patterns (CEMDAP). *Transportation Research Record*, Vol. 1894, 2004, pp. 57–66.

[4] Bowman, J., M. Bradley, Y. Shiftan, T. 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*, Elsevier, Oxford, Vol. 3, 1998, pp. 171–184.

[5] Jonnalagadda, J., N. Freedman, W. Davidson, and J. Hunt, Development of microsimulation activity-based model for San Francisco: destination and mode choice models. *Transportation Research Record*, Vol. 1777, 2001, pp. 25–35.

[6] Pendyala, R., *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, 2004, see www.eng.usf.edu/~pendyala/publications.

[7] Lin, D.-Y., N. Eluru, S. Waller, and C. Bhat, Integration of activity-based modeling and dynamic traffic assignment. *Transportation Research Record*, Vol. 2076, 2008, pp. 52–61.

[8] Balmer, M., N. Cetin, K. Nagel, and B. Raney, Towards truly agent-based traffic and mobility simulations. In *Autonomous agents and multiagent systems (AAMAS'04)*, New York, NY, 2004.

[9] Rieser, M., K. Nagel, U. Beuck, M. Balmer, and J. Rümenapp, Truly agent-oriented coupling of an activity-based demand generation with a multi-agent traffic simulation. *Transportation Research Record*, Vol. 2021, 2007, pp. 10–17.

[10] de Palma, A. and F. Marchal, Real case applications of the fully dynamic METROPOLIS toolbox: An advocacy for large-scale mesoscopic transportation systems. *Networks and Spatial Economics*, Vol. 2, No. 4, 2002, pp. 347–369.

[11] Zhou, X., H. Mahmassani, and K. Zhang, Dynamic micro-assignment modeling approach for integrated multimodal urban corridor management. *Transportation Research Part C*, Vol. 16, No. 2, 2007, pp. 167–186.

[12] Meister, K., M. Rieser, F. Ciari, A. Horni, M. Balmer, and K. Axhausen, Anwendung eines agentenbasierten Modells der Verkehrsnachfrage auf die Schweiz. In *Proceedings of Heureka '08*, Stuttgart, Germany, 2008.

[13] Flötteröd, G., *Traffic State Estimation with Multi-Agent Simulations*. Ph.D. thesis, Berlin Institute of Technology, Berlin, Germany, 2008.

[14] Flötteröd, G., Cadyts – a free calibration tool for dynamic traffic simulations. In *Proceedings of the 9th Swiss Transport Research Conference*, Monte Verita/Ascona, 2009.

[15] MATSIM www page, *MultiAgent Transport SIMulation*. http://matsim.org/, accessed 2009.

[16] 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.), Routledge, 2006, pp. 305–347.

[17] Avineri, E. and J. Prashker, Sensitivity to uncertainty: Need for paradigm shift. *Transportation Research Record*, Vol. 1854, 2003, pp. 90–98.

[18] Charypar, D. and K. Nagel, Generating complete all-day activity plans with genetic algorithms. *Transportation*, Vol. 32, No. 4, 2005, pp. 369–397.

[19] Ben-Akiva, M. and S. Lerman, *Discrete Choice Analysis*. MIT Press series in transportation studies, The MIT Press, 1985.

[20] Cetin, N., A. Burri, and K. Nagel, A Large-Scale Agent-Based Traffic Microsimulation Based On Queue Model. In *Proceedings of Swiss Transport Research Conference (STRC)*, Monte Verita, CH, 2003, see www.strc.ch. Earlier version, with inferior performance values: Transportation Research Board Annual Meeting 2003 paper number 03-4272.

[21] Gawron, C., *Simulation-based traffic assignment*. Ph.D. thesis, University of Cologne, Cologne, Germany, 1998.

[22] Cadyts, *Cadyts web site*. http://transp-or2.epfl.ch/cadyts, accessed 2009.

[23] Chen, Y., M. Rieser, D. Grether, and K. Nagel, *Improving a large-scale agent-based simulation scenario*. VSP working paper 08-15, Transport Systems Planning and Transport Telematics Laboratory, Berlin Institute of Technology, https://svn.vsp.tu-berlin.de/repos/public-svn/publications/vspwp/2008/08-15/, accessed 2009, 2008.

[24] Balmer, M., K. Axhausen, and K. Nagel, A demand generation framework for large scale micro simulations. *Transportation Research Record*, Vol. 1985, 2006, pp. 125–134.

[25] SFSO, *Ergebnisse des Mikrozensus 2005 zum Verkehr*. Swiss Federal Statistical Office, Neuchatel, 2006.

[26] Grether, D., Y. Chen, M. Rieser, and K. Nagel, Effects of a Simple Mode Choice Model in a Large-scale Agent-Based Transport Simulation. In *Complexity and Spatial Networks. In Search of Simplicity* (A. Reggiani and P. Nijkamp, eds.), Springer, Advances in Spatial Science, 2009, chap. 13, pp. 167–186.

[27] Rieser, M., D. Grether, and K. Nagel, Adding mode choice to a multi-agent transport simulation. *Transportation Research Record: Travel Demand Forecasting 2009*, Vol. 2132, 2009, pp. 50–58.