Behavioral calibration of a large-scale travel behavior microsimulation

Gunnar Flötteröd*, Yu Chen, Marcel Rieser, Kai Nagel†

November 9, 2009

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 [20]. 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 [26].

An alternative is to use a microscopic approach, where every traveler is modeled individually. This typically starts with a synthetic population of individuals, adds activity patterns and activity locations to each individual, lets the synthetic travelers choose their mode, and ends with a route assignment procedure.

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, [5, 6, 16, 21], 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

*Transport and Mobility Laboratory (TRANSP-OR), Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland, gunnar.floetteroed@epfl.ch

†Transport Planning and Transport Telematics Laboratory (VSPVT), Technical University of Berlin (TUB), Germany, {chen,rieser,nagel}@vsp.tu-berlin.de

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 [17].

This type of coupling between the ABDM and the traffic assignment leaves room for improvement [3, 23]. 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 [11, 27]).

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 [19]). 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 [12, 13]. 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. Finally, Section 5 summarizes the article and gives an outlook on future research directions.

2 Outline of transport microsimulation

The MATSim (“Multi-agent transport simulation toolkit”, [18, 22]) 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 [1]. 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.

- 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

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

It is important to note that the score thus takes into account the complete daily plan. More details can be found in [9, 22].

The plan choice is modeled with a multinomial logit model (which clearly calls for enhancements in the future) [4]. 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

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

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 [8, 14]: 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 rather awkward task that was based on intuition and, unfortunately, the arbitrariness this brings along. This section outlines the Cadyts ("Calibration of dynamic traffic simulations" [7, 13]) 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 utmost 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 [12]:

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

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

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

That is, an implementation of the posterior choice distribution requires nothing but to add a link-additive correction term 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.

What is described here is a calibration of the realized choices. Another important aspect of the calibration is to reveal structural information about the choice model itself, e.g., in terms of coefficients of the utility function. In this regard, the additive utility modifications can be seen as corrections of the alternative specific constants of their respective plan alternatives. However, the possibility to exploit traffic counts for the calibration of demand parameters is not limited to this, and it appears plausible to also deploy it to correct other utility parameters. This is an important subject of future research.

4 Zurich field study

This section describes results from an ongoing 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.

4.1 Description of test case and uncalibrated simulation results

Figure 1 gives an overview of the Zurich analysis zone, and Figure 2 shows the according road network. 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 [10].

A synthetic population of travelers for all of Switzerland is available from a previous study [2, 19]. 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

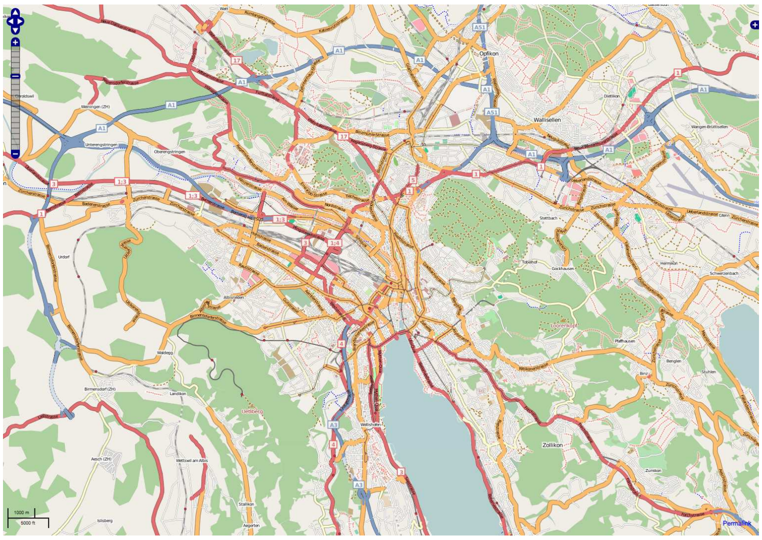

Figure 1: Zurich analysis zone

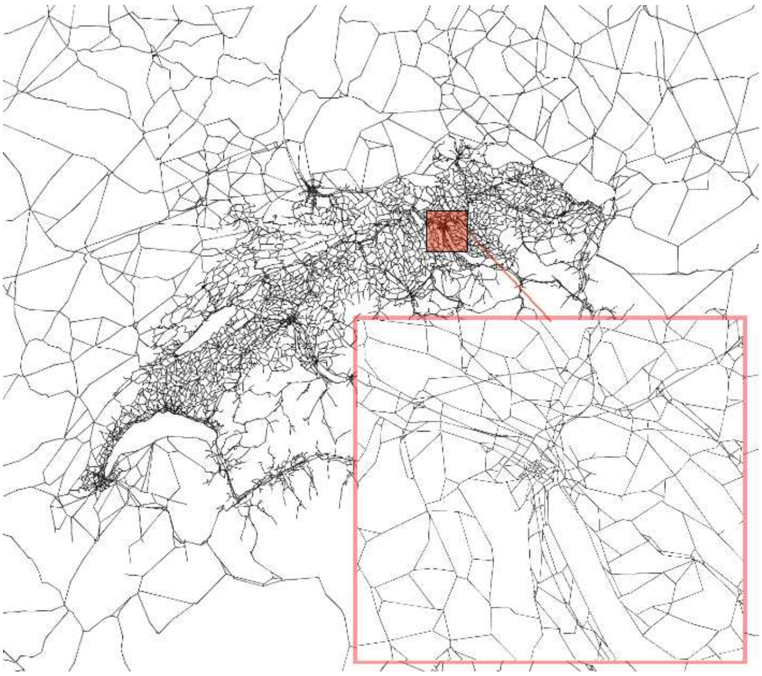

Figure 2: Zurich network

Table 1: Simulation parameters

parameter	value
$\beta_{perf.act.}$	12 Eur/h
β_{car}	-12 Eur/h
$\beta_{non-car}$	-6 Eur/h
size of plan choice set	4
total number of iterations	500
iterations for choice set generation	300

Table 2: Opening and closing times

activity type	opening time	closing time
home	00:00	24:00
work	07:00	18:00
education	07:00	18:00
shop	08:00	20:00
leisure	00:00	24:00

(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 behavioral 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. Table 2 summarizes the opening and closing times available to perform activities. Public transit is simulated as described in Refs. [15, 24], 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 3 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 3 shows results after 500 iterations of uncalibrated simulation. The scatterplots reveal a minor underestimation of the volumes in the simulation, which can be explained by the limited number of activities accounted for in the generation of the initial demand. However, the overall bias is moderate.

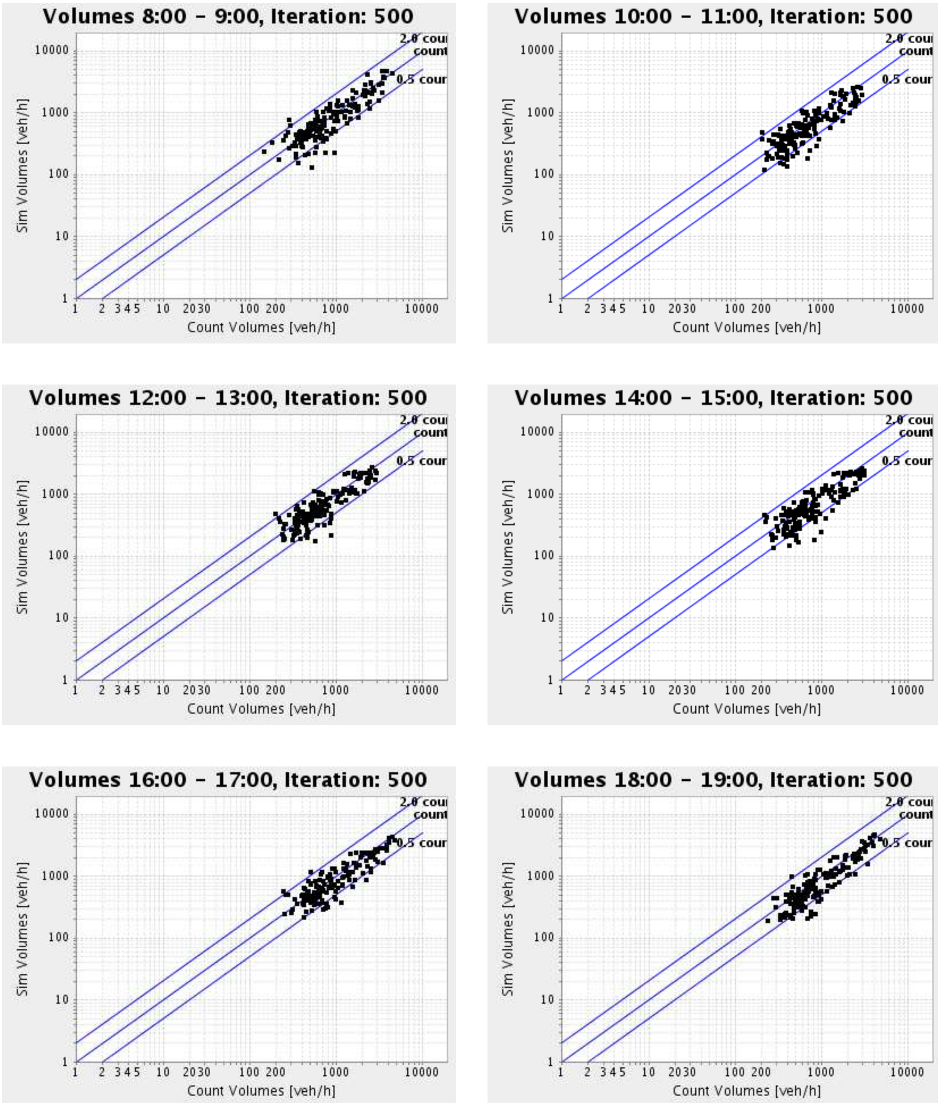

Figure 3: Scatter plots for uncalibrated base case

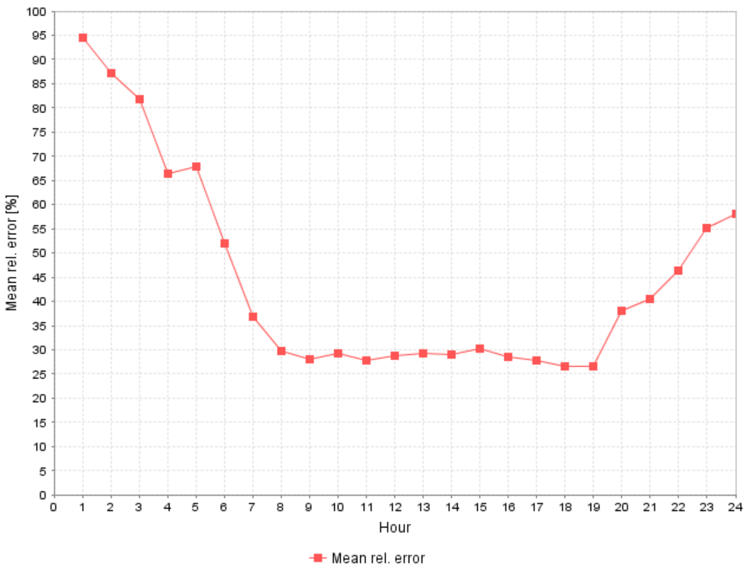

Figure 4: Mean relative error (MRE) for uncalibrated base case

The line above (below) the main diagonal represents simulation values of twice (half) the observed traffic counts (note that the plots are double-logarithmic). Most points are within this (admittedly loose) band, 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 4 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 evolution of the calibrated simulation over the iterations is visualized in Figure 5, which shows the mean weighted square error MWSE of all measurements over the iteration number. This 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

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

which also is the specification used in (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).

Since the system starts already in an equilibrated state that has been attained after 500 uncalibrated iterations, all systematic changes of MWSE in Figure 5 can be attributed to the calibration. The MWSE is quickly reduced from more than 100 to around 45 in the first 100 iterations. After this, the curve flattens. It is plausible to assume that in the first iterations, the calibration "fills up" the measurement locations by arbitrary plans, and that in the following iterations the simulation rearranges the plans such that behaviorally more reasonable plans take the place of other plans that have been "used" by the calibration before.

The choice set generation phase finishes at iteration 300, which clearly generates a jump in the system behavior: Since the immediate execution of newly generated plans is omitted, the calibration can affect the whole plan choice distribution, which results in another improvement of MWSE from around 35 to little more than 20. The variability of MWSE is reduced to almost zero after iteration 300, which also is a consequence of the reduced variability in the executed plans once the choice set generation is turned off.

The scatterplots of Figure 6 are obtained from the last iteration of the calibrated simulation. A comparison with the uncalibrated scatterplots of Figure 3 shows a substantial improvement in measurement fit in that the data points are substantially more centered around the main diagonal. Figure 7 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

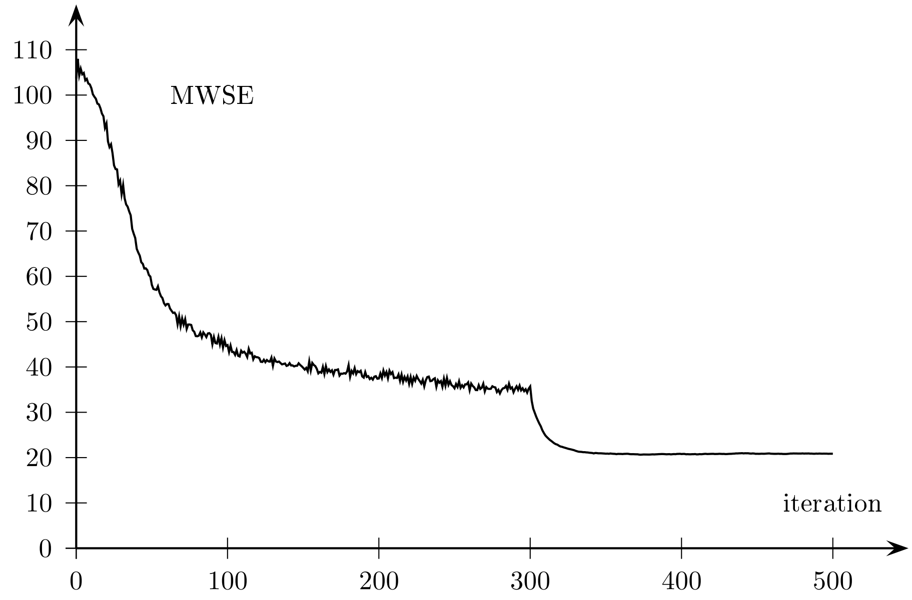

Figure 5: Mean weighted square error (MWSE) using all counting stations

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 clearly 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.

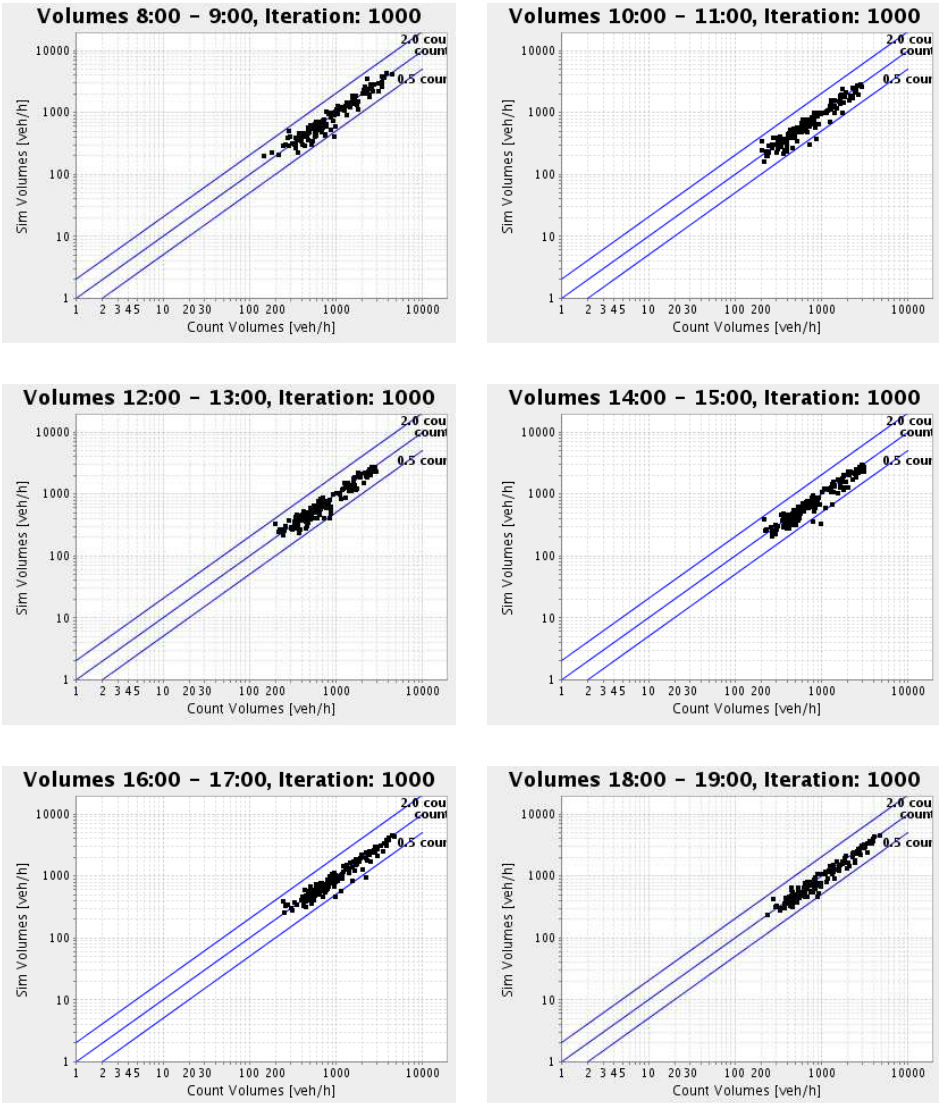

Figure 6: Scatter plots after calibration

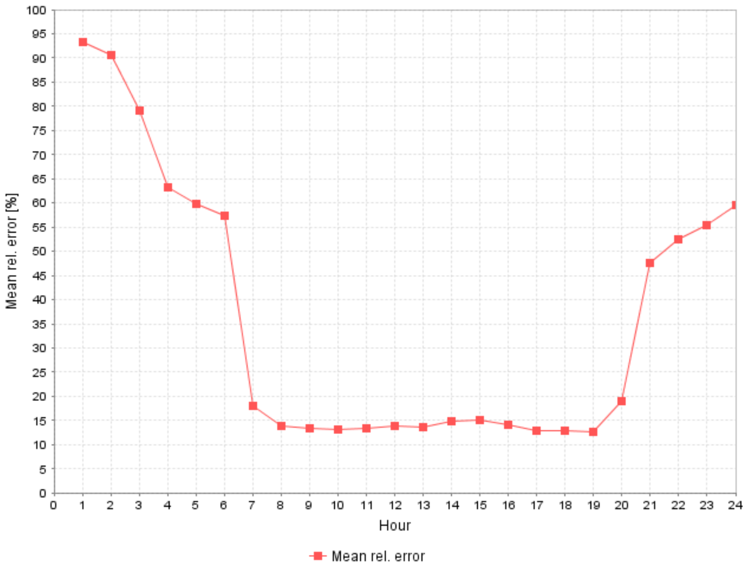

Figure 7: Mean relative error (MRE) after calibration

Figure 8 shows the MWSE trajectories of the measurement data for all ten experiments over the iterations, where all trajectories are normalized to their values at iteration zero for better comparability. Figure 9 shows the same type of curves for the validation data. The similar dynamics of the measurement MWSE values indicate that the calibrated simulation exhibits well-behaved dynamics and generates reproducible results. Overall, the measurement reproduction error is reduced by around than 80 % in all cases.

The validation MWSE curves exhibit a greater variability, which can be explained by the lower number of measurements that enter the averaging in (7). Again, the variability is substantially decreased once the choice set generation is turned off in MATSim. The different experiments attain different values, which can be explained by the fact that here disjoint sets of sensor data are evaluated. Overall, an improvement of 15 % to 45 % is attained. This clearly indicates that the local information that is contained in the measurement data is used by the calibration in a way that affects the network-wide agent behavior such that more realistic network conditions result even far away from the sensor locations. One also has to keep in mind that the relative positioning of the sensors affects the validation results in that the extrapolation power of the calibration is limited by the spatiotemporal correlations in the network conditions: if the validation sensors are too far away, they simply are not affected any more by the calibration, no matter how good it is.

These results show clearly 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

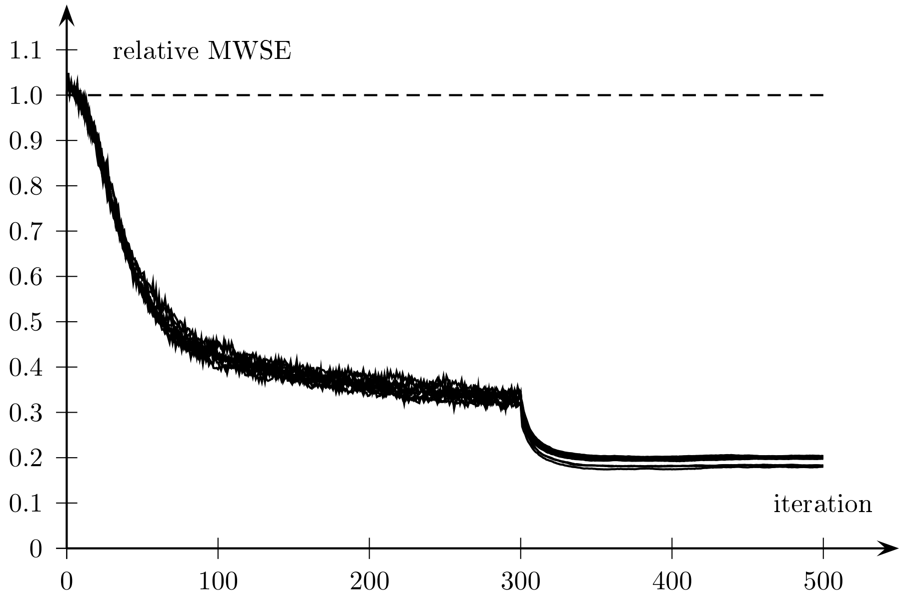

Figure 8: Validation results – measurement reproduction

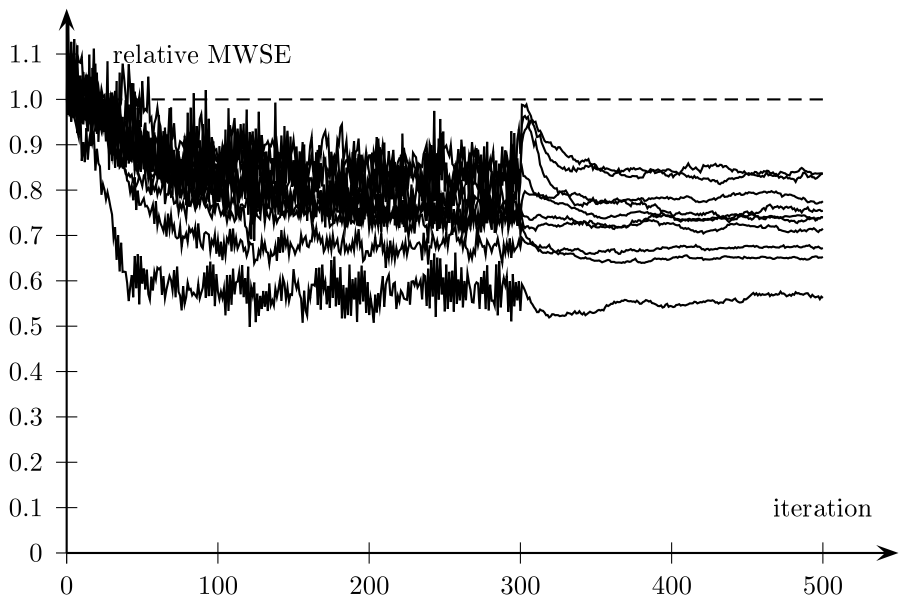

Figure 9: Validation results – measurement extrapolation

a purely simulation-based environment on a network with many ten thousand links. The number of iterations required to obtain stable and realistic results is in the order of a plain simulation, and the computational overhead introduced by the calibration is negligible. The authors are not aware of any other calibration technique that comes close to such results.

5 Summary and outlook

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.

References

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

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

[3] M. Balmer, 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, July 2004.

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

[5] C.R. Bhat, J.Y. Guo, S. Srinivasan, and A. Sivakumar. A comprehensive econometric microsimulator for daily activity-travel patterns (cemdap). *Transportation Research Record*, 1894:57–66, 2004.

[6] J.L. Bowman, M. Bradley, Y. Shiftan, T.K. Lawton, and M. Ben-Akiva. Demonstration of an activity-based model for Portland. In *World Transport Research: Selected Proceedings of the 8th World Conference on Transport Research 1998*, volume 3, pages 171–184. Elsevier, Oxford, 1998.

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

[8] N. Cetin, 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.

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

[10] Y. Chen, 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.

[11] A. de Palma and F. Marchal. Real case applications of the fully dynamic METROPOLIS tool-box: An advocacy for large-scale mesoscopic transportation systems. *Networks and Spatial Economics*, 2(4):347–369, 2002.

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

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

[14] C. Gawron. *Simulation-based traffic assignment*. PhD thesis, University of Cologne, Cologne, Germany, 1998.

[15] D. Grether, Y. Chen, M. Rieser, and K. Nagel. Effects of a simple mode choice model in a large-scale agent-based transport simulation. Submitted to "paralimes", in press.

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

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

[18] MATSim web site. http://www.matsim.org, accessed 2009.

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

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

[21] R.M. Pendyala. 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.

[22] B. Raney and K. Nagel. An improved framework for large-scale multi-agent simulations of travel behavior. In P. Rietveld, B. Jourquin, and K. Westin, editors, *Towards better performing European Transportation Systems*, pages 305–347. Routledge, 2006.

[23] M. Rieser, 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*, 2021:10–17, 2007.

[24] Marcel Rieser, Dominik Grether, and Kai Nagel. Adding mode choice to a multi-agent transport simulation. Paper 09-2758, Transportation Research Board Annual Meeting, Washington, D.C., 2009.

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

[26] P. Vovsha, M. Bradley, and J.L. 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, May 2004.

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