

Automatic Calibration of Agent-Based Public Transit Assignment Path Choice to Count Data

Manuel Moyo Oliveros*, Kai Nagel*

**Technische Universität Berlin, Transport Systems Planning and Transport Telematics
Salzufer 17-19, 10587 Berlin, Germany. <http://www.vsp.tu-berlin.de>*

Abstract

This work describes the calibration of a schedule-based transit assignment inside an iterative microscopic agent-based simulation. The calibration challenge implies that the behavioral rules should be modified in order to move the simulation closer to observed passenger counts. First, route choice set of agents is enriched with travel parameter utilities randomization. Secondly, the calibration interacts directly into the performance evaluation of individual daily plan of activities, so that the plan is also evaluated for its contribution to the count reproduction. In this way, appropriate plans from the calibration perspective can persist along simulation iterations. The Berlin public transport system with day-based counts is used as test scenario. The results show that the calibration approach can work with large scale scenarios, and that it is able to deal with the inter-temporal aspects implied by counts.

1. Introduction

The implementation of a microscopic agent-based model for public transport simulation represents a helpful instrument for mobility research. If the simulation includes also a travel behavioral approach, it constitutes an advantageous analysis tool for the knowledge of passengers' route decisions and to recognize the effects of those decisions on the whole transport system.

In microscopic models, a special challenge has been calibration. More technically: Given a set of macroscopic observations, how should the physical or behavioral microscopic rules of the agent-based simulation be modified in order to move the simulation closer to the observations? The example considered here is a situation where one has daily passenger counts for certain or all transit lines, demand for 1% of the population from trip diaries, but no route information from the survey. The task is to find passenger connections and possibly modify the passenger demand such that the simulation matches the counts.

This topic belongs to a class of problems that are quite common in agent-based simulations. Agent-based simulations were always built around the notion of "emergence", that is, they are expected to be particularly useful where certain macroscopic properties, in our case congestion, vehicle overloading, and resulting delay patterns, cannot be derived in analytical ways from the microscopic input data (including the behavioral rules), and in consequence one needs to run the simulation in order to obtain them. However, because the connection from input data to emergent properties is by simulation, the mathematical connection is not as well established as in normal numerical modeling.

2. Related works

Several works for demand estimation of large scenarios based on passenger counts can be found. However, most of them are based on transit OD matrix adaptations. Rongviriyapanich, Nakamura, and Okura (2000) used on-off count data for two bus routes in Tokyo to evaluate OD estimation techniques used originally for road networks. The Entropy Maximization Method is reported as the most practical of all considered techniques, if a prior OD matrix is available. In the same way, Wen Chi (2005) used boarding and alighting counts of the Hong Kong metro network for station-to-station OD matrix assignment calibration and validation. Random choice function coefficients generation methods in Monte Carlo simulation are also presented. Farrol and Livshits

(1998) calibrated a scenario of the Greater Toronto Area based on a survey in 1996. Their results present the adjusted weights to access time, wait time, and the transfer penalty in an EMME/2 implementation. Lu (2008) used automatically-obtained passenger boarding and alighting counts for a bus line in Columbus, Ohio to review five OD flow estimation methods. The results show that the output quality depends to a great extent on the quality of the base OD matrix. Li (2009) presented the statistical inference of large transit route matrices using on-off counts. Considering a given occupancy of a passenger on a stop, the probability of alighting on a posterior stop of the transit line is calculated with a Markov chain model. A Bayesian analysis draws inference about unknown parameters.

3. Agent-based transit simulation

An important challenge for the agent-based approach for the recent past has been computational: Finding an implementation that is both close to the agent-based concepts and fast enough for real world scenarios. For transit micro-simulation, the open source framework MATSim (Multi-Agent Transport Simulation) (Balmer, Raney & Nagel, 2005) (Raney & Nagel, 2006) is employed in this work. Its capacity to simulate large scale scenarios and its modular architecture make it appropriate for calibration research. Its simulation approach follows these steps:

- Initialization: Synthetic travelers with activity-based day plans are assumed to be given; see below how they were generated for the present study. This includes generating suitable connections through the public transit network for each trip that is part of the passenger’s plan. The calculation of routes for passengers is described in Sec. 4.3 of (Rieser & Nagel, 2009) and in Sec. 7.4 of (Rieser, 2010). The routing process looks for least compound cost paths with a trade-off between walk time, waiting time, in-vehicle travel time, travel distance, and vehicle change counts. These travel preferences are represented respectively in an economic approach by these parameters:

Table 1. Transit Travel Parameters

Name for present paper	Name in MATSim config	Current MATSim default value	Effective value (including Opportunity Cost of Time)
Marginal Utility of Travel Time Walk (MUTTW)	travelingWalk	-6/h	-12/h
Marginal Utility of Transit Waiting Time (MUTWT)	waitingPt	-6/h	-12/h
Marginal Utility of Travel Time Transit (MUTTT)	travelingPt	-6/h	-12/h
Marginal Utility of Travel Distance Transit (MUTDT)	monetaryDistanceCostRatePt *	0	0
Utility of Line Switch (ULS)	marginalUtilityOfMoney	-1	-1
	utilityOfLineSwitch		

- The “effective value” includes the marginal Opportunity Cost of Time, which needs to be added to all marginal (dis)utilities of time. It denotes the implicit penalty incurred by cutting short preceding or succeeding activities. In MATSim, it is approximately given by the parameter called “performing”, and its default is $-6/h$. The utilities are taken as dimensionless quantities; $-6/h$, say, means “minus six utils. per hour” or “minus 6/3600 utils. per second”. Note that the original approach weights all temporal contributions equally.
- Synthetic reality (aka network loading): All plans are executed simultaneously in a simulation of the physical system (“synthetic reality”) which includes a detailed simulation of the public transit system (Rieser, 2010),

(Rieser & Nagel, 2009). This simulation can and should describe travel incidents such as boarding delays, failures because of overloading, delays because of late incoming vehicles, etc.

- Scoring: Each plan is scored based on its own performance in the synthetic reality. Following a utility-based approach (Charypar & Nagel, 2005), the utility $V(i)$ of a plan i is calculated as the sum of positive utilities (in a logarithmic form) achieved by carrying out activities, plus the sum of negative utilities of traveling between activities locations

$$V(i) = \sum_{act \in i} \beta_{perf} \cdot t_{act}^* \cdot \ln t_{perf,act} + \sum_{leg \in i} V_{tr,leg} \quad (1)$$

where:

β_{perf} is the activity marginal utility at its typical duration.

t_{act}^* is the activity typical duration.

$t_{perf,act}$ is the activity duration in the simulation (in sec).

$V_{tr,leg}$ is the utility (typically negative) of a leg.

- Choice set modification: The synthetic travelers try to construct new plans given their experience from the previous simulated day (iteration). The re-planning strategies are methods to construct new plans with mechanisms whereby some plans properties are modified. A typical strategy in this calibration study is `ReRoute`. When this strategy performs, a complete new route is calculated for the passenger. The new plans are added to the agents' choice sets. Plans that achieved a poor score according to the utility-based approach are removed from choice sets.
- Choice: All other agents select between already known plans. Configurable selection strategies define the choice criteria. Typically the `ChangeExpBeta` strategy is used in standard simulations in order to approximate the simulation to a logit model. See section 3.2.1 of (Nagel & Flötteröd, 2009).
- Go to: synthetic reality (above).

The iterative process is continued until some appropriate convergence criterion is reached. The loop describes day-to-day learning; one may, in particular for the investigation of short-term customer reactions, also include within-day or en-route learning.

4. Greater Berlin Scenario

4.1. Demand preparation

The information about public transport demand in the Berlin and Brandenburg area was granted by the BVG (Berlin public transport company). The demand was transformed from a macroscopic representation into an activity-based model description. This was realized in a work by Neumann, Balmer, and Rieser (2009). There, the plans of 598,891 persons who use all transport modes were generated. This synthetic population was adopted for the purposes of this work. In order to set the research focus on the transit calibration, these preparatory steps were realized:

- Routing: as the original passenger's survey did not include data about selected routes, the connections between activity locations were calculated before the simulation. Basic route diversity was obtained by calculating 3 plans per agents with three travel priorities: strong walk penalty, strong transfer penalty and moderate values of them.
- Filtering: All persons who did not include the public transport mode at all in their plans were discarded. Some persons who intended to use public transport were also discarded, namely those whom the transit router

calculated a direct walk to their destinations instead of a transit connection. 231,369 persons remained at the end.

4.2. Counts

Daily occupancy counts for stops were also granted by the BVG. The occupancy measurements are defined on a per day basis. It means that all counts from simulation are not distributed in 24 time bins, but in one single time bin for a whole day. These preparation steps were necessary for them:

- Filtering: Since stops and schedule data originated from different projects, the counts were validated to use only counts that could match nominally and geographically with the stops inside the schedule file with a high degree of certainty. Only 2723 from the original 7125 stop counts remained.
- Stop zone conversion: The counts data were not related to stops, but to stop zones. A stop zone describes a set of near stops that usually have the same name but each stop can be used by different transit lines or routes in different directions. Observed stop zone-based counts were not fragmented to assign occupancy values to each component stop because there was not any reference at all to do it. Instead of it, the simulation worked doing normal stop-based occupancy analysis but an extra module was developed to do stop zone-based analysis by aggregating the simulated particular stops occupancies into their respective zone.

4.3. Transit Schedule

The transit schedule contains the supply data of the scenario in the scenario. The complete Berlin public transport network with 329 transit lines was used in the simulation. However, not all lines were considered for the calibration procedure. After matching the data of counts and transit stops, only 218 lines that contained at least one of the considered stops with occupancy counts were labeled for the calibration.

4.4. Initial transit simulation

First, routes for all agents were generated. Passengers got their transit connections calculated with default parameter values of the transit router. The MUTTT and the MUTWT parameters added the Opportunity Cost of Time to their default values. The Opportunity Cost of Time represents the implicit punishment for not performing an activity and it represents an additional penalty value of -6/3600sec as explained earlier. After all routes for passengers were calculated, a subsequent standard simulation run was executed.

Fig. 1 shows the analysis with scatter plots for the initial and final iterations. In scatter plots, every dot represent both observed and simulated values for a stop at the given time bin. The observed values are in the scale of X axis and the simulated values are in Y axis. Thus, a point in the diagonal would represent the state of exact match for both values. Expressed differently, for a perfect reproduction of counts in the simulation, all dots must be situated exactly over the diagonal with slope one. On the contrary, the more distant a dot from diagonal one is, the bigger is its observed-simulated values inconsistency.

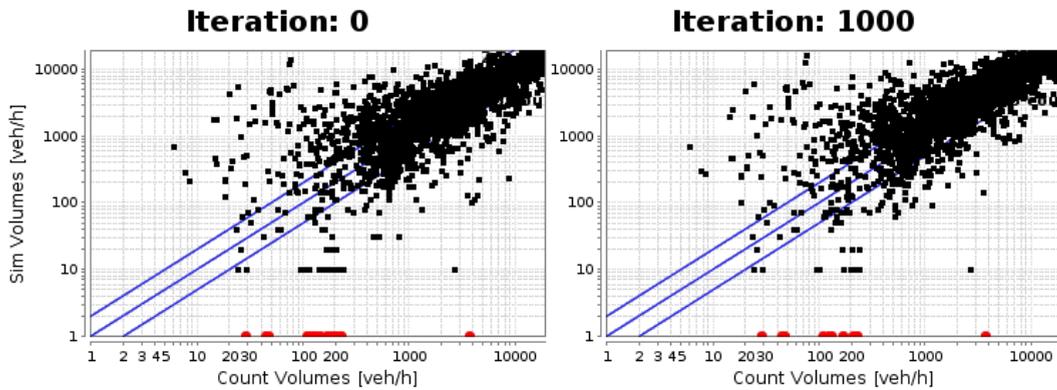


Fig. 1 Scatter plot for initial situation of Greater Berlin scenario: standard transit simulation with MATSim transit router

One can notice that in both iterations, dots bring together around both sides of the diagonal, which is interpreted as under- and over-occupation. The Mean Relative Error (MRE) for the simulated day and selected lines starts with 89.6% and ends with 97.8% in iteration 1000. The reason why anything at all changes over the iterations lies in the fact that there is also a car traffic assignment which changes over the iterations. This can, for example, cause synthetic travelers with mixed car/transit plans to obtain a different time structure because of changing car mode travel times.

5. A calibration solution approach: Cadyts

5.1. Cadyts configuration

For transport simulations, but with the clear potential to be more general, the demand estimation problem was addressed by G. Flötteröd and co-workers (e.g. Flötteröd, 2008; Flötteröd, Bierlaire, and Nagel, 2011a). He implemented his methodological approach into the open source software Cadyts (Flötteröd, 2013). Very intuitively, the approach uses the freedom that is left when individual decisions are modeled as random draws from a discrete choice model: decisions that are congruent with the observations become preferred over those that are not.

Cadyts theoretical principles can be found in (Flötteröd, Bierlaire & Nagel, 2011a). Here, only the calibration steps are briefly introduced:

1. Initialization. The calibration process starts by registering observed counts at stops. Each entry has this data structure:

- **Id:** defines the unique identifier for the count station.
- **Start time:** defines the inclusive initial time for the count time bin.
- **End time:** defines the exclusive final time for the count time bin.
- **Value:** defines the number of observed mobile agents during the time bin.
- **Minimal standard deviation (`minStdDev`):** defines the smallest allowed standard deviation for the observed counts. After an initial experimentation, it was set to 4 for tests described in this work; see below for more information.

2. Choice. After a given simulation iteration, the calibrator calculates the utility correction per agent that

influences the plan choice to match the counts reproduction in a more efficient way. The choice is based on the calculated utility correction and reported to the calibrator. For the purpose of this work, a special implementation makes Cadyts act as a weighted accumulated scoring function.

3. Update. The calibrator is informed about the network loading conditions after the simulation iteration.

The Cadyts posterior choice model is outlined in Section 3.1 of (Flötteröd, Chen, Rieser & Nagel, 2009). Specially the section 3.2 presents the formulation of its distribution embedded with the MATSim demand simulation multinomial logit model. For the general formulation reproduced next, “congestion is assumed to be light and the traffic counts are independently and normally distributed” (Flötteröd, Chen & Nagel, 2011b). The core equation for the purposes here is:

$$P(i|y) \sim \exp \left(V(i) + \sum_{ak \in i} \frac{y_a(k) - q_a(k)}{\sigma_a^2(k)} \right) \quad (2)$$

where:

$P(i|y)$ is the posterior plan choice distribution given y .

$V(i)$ is the standard score of a plan i as formulated in Eq.(1).

$y_a(k)$ is the actual observed count at location a during time k .

$q_a(k)$ is the simulated count at location a during simulated time k .

$\sigma_a^2(k)$ is the inverse weight of the measurement.

The inverse weight $\sigma_a^2(k)$ should optimally come from specific knowledge about each sensor, but in its absence, it is calculated as: $\sigma_a^2(k) = \max(\text{varianceScale} * y_a(k), \text{minStdDev}^2)$, i.e. assumed to be proportional to the measured value in order to be consistent with an assumption of Poisson distributed measurements (Flötteröd, 2010). For the present paper, *varianceScale* is left at its default value of 1.0, meaning that all measurements below $y_a = 16$ are under-weighted because of *minStdDev* = 4.

5.2. Initial brute force calibration with fixed route choice

The first calibration attempt over the scenario is presented in this section. It was realized with the brute force setting, which enforces the counts reproduction with best effort, usually overriding the effect of other evolutionary mechanisms like the standard scoring function. The brute force is not a plausible method in a behavioral context, but it is used only as preliminary test to prove the capacity of calibration settings over the synthetic demand in its initial state.

Similar to Moyo O. and Nagel (2012), a fixed set of three transit connection alternatives per passenger were pre-calculated according to three diverse travel criteria: least number of interchanges, least amount of walking, and a balance of them. For this first calibration, Fig. 2 shows the comparison of counts adequacy between initial iteration 0 and final iteration 1000 in scatter plots.

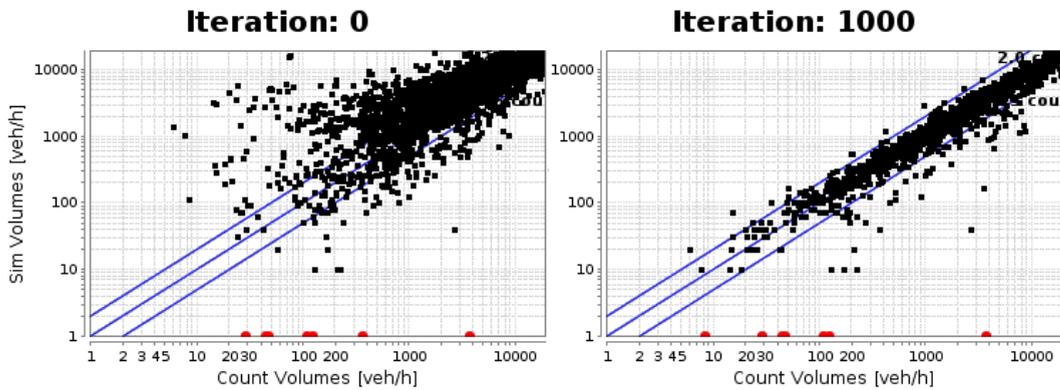


Fig.2: Brute force calibration of Greater Berlin scenario with fixed route choice

One can see that the brute force calibration ends up with an acceptable match of counts, even with a large number of agents and day-based counts. The brute force setting pushes the simulation into the reproduction of occupancy volumes by selecting the plans most consistent with the observations. The MRE starts with value 128.5 % and it is reduced to 15.3% at iteration 1000. However, while the approach works well, the brute force option and the scoring function switching-off are incongruous from a behavioral modeling perspective. Moreover, the method depends on the fixed set of pre-calculated transit connections. Without plans mutation, Cadyts can only shift between existing plans and this would not allow the calibration procedure to guide the search into the reproduction of real measurements.

6. Transit route diversity research

It is assumed that the calibration core procedure leaves the generation of choice alternatives to the simulation counterpart. Cadyts only influences the count reproductions by evaluating and preferring some alternatives that are presented to it. At the end, a favorable calibration output depends to a large extent on the generation of sufficient choice diversity. A reduced number of routes, or a choice set that do not correspond to realistic passengers routes, will affect negatively the expected count match.

Two ways have been so far considered to apply Cadyts to agent-based public transit assignment, the second one described in this paper for the first time. In a first approach (Moyo O. & Nagel, 2012; also described above), multiple transit connection alternatives per passenger were pre-calculated according to three diverse criteria: least number of interchanges, least amount of walking, and some balance of the criteria. Cadyts would then select those plans that help to match the observed occupancy values.

While that approach worked well, it depended on the fixed set of pre-calculated transit connections. This is a very basic attempt to create route diversity. For that reason, a second approach is presented here, where new plans are generated as part of the choice set modification step (Step 4). For this, a transit routing algorithm is used that randomizes the utility coefficients of in-vehicle time, walk time, wait time, and number of interchanges, before searching for a route.

With the goal of discarding the pre-calculation of the route set and generate discretionary connections instead, a special routing module was implemented in MATSim. The “randomized transit router” takes MUTWT, MUTTW, MUTTT, MUTDT and ULS default values and generates new random route search costs from them. Every time that each parameter gets a new random value, a new different connection can be generated. In this way, very diverse passenger travel priorities are simulated.

The generation of route diversity from random values is illustrated with an example of a simulated person who needs to travel from the TU Transport Systems Planning and Transport Telematics Institute to the transit hub *Alexander Platz* in Berlin. First, the transit router calculates an initial route at iteration “0” with default parameter values (They include here the Opportunity Cost of Time). With random re-routing as re-planning strategy, after 10 more iterations, the same number of combinations of random values is also generated. Table 2 shows the values created for each iteration.

Table 2. Travel parameter random value generation example.

Random values combination	MUTTW	ULS	MUTWT	MUTTT
0 (default values)	-12.0 / 3600	-1.0	-12.0 / 3600	-12.0 / 3600
1	-2.6 / 3600	-1.9	4.5 / 3600	-18.3 / 3600
2	-13.0 / 3600	-1.3	-19.4 / 3600	-9.0 / 3600
3	-59.9 / 3600	-3.2	-9.6 / 3600	-43.8 / 3600
4	-39.1 / 3600	-2.6	-9.0 / 3600	-59.6 / 3600
5	-26.0 / 3600	-4.5	-29.2 / 3600	-34.3 / 3600
6	-20.7 / 3600	-3.7	-10.1 / 3600	-27.9 / 3600
7	-7.6 / 3600	-0.6	-0.3 / 3600	-2.6 / 3600
8	-46.8 / 3600	0.0	-29.7 / 3600	-18.5 / 3600
9	-33.8 / 3600	-1.9	-19.9 / 3600	-53.2 / 3600
10	-26.5 / 3600	-1.9	-14.9 / 3600	-16.0 / 3600

At each iteration, the respective value combination is given to the transit router to calculate a new additional connection for the passenger. These 11 routes are graphically represented on Fig. 3. There, the origin is marked with letter “0” and the destination with “D”. For 11 queries, the randomized transit router calculated 8 different routes with diversity in number of transfers and walk distances. Actually, the first route in light green with number 0 is a direct walk to the destination, which means that the initial values are not good enough to find a connection for this example. The same color is employed for the starting and ending walking distances in the other routes. Four iterations calculated the same route marked with a red line. It can be explained with the fact that that connection correspond to the use of a near bus line that links to the high frequency and fast suburban metro lines to *Alexander Platz*.

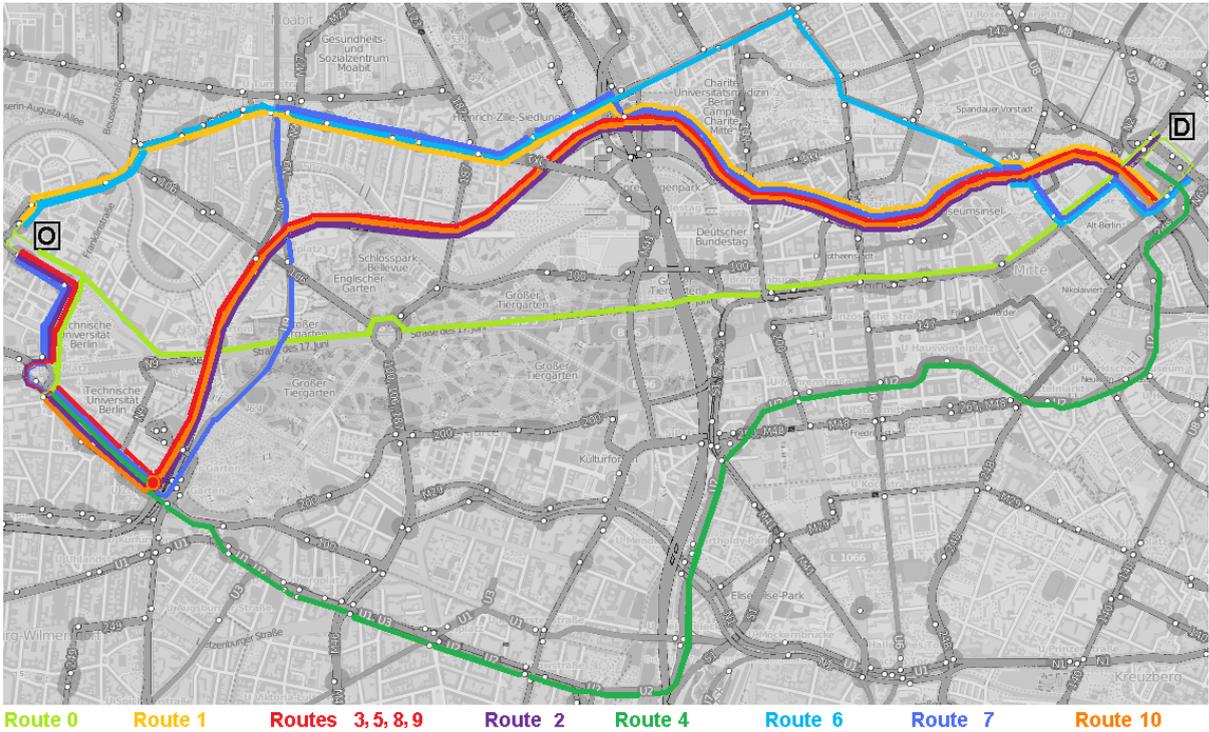


Fig 3: Randomized transit router example: 11 route queries from TU Transport Systems Planning and Transport Telematics Institute to transit hub Alexander Platz in Berlin.

7. Cadyts calibration as scoring function

Cadyts calculates utility corrections for plans in order to guide the re-planning and choice process in the direction of the counts match. The new implementation of MATSim-Cadyts integration presented here consists in the integration of the calibration linear plan effect with the other simulation scoring components (performing and traveling). Thus, the counts match is also included as part of the plan evaluation. More formally, Cadyts core function of posterior choice distribution (Eq. 2) is added to the utility of plan $V(i)$ (Eq. 1) during the scoring process as a new weighted term:

$$V(i) = \sum_{act \in i} \beta_{perf} \cdot t_{act}^* \cdot \ln t_{perf,act} + \sum_{leg \in i} V_{tr,leg} + \left(w \sum_{ak \in i} \frac{y_a(k) - q_a(k)}{\sigma_a^2(k)} \right) \quad (3)$$

where:

w is the weight of Cadyts correction inside the accumulated scoring function.

Having Cadyts as part of the scoring function leads to these advantages:

- Brute force is technically abandoned which returns the calibration standpoint to a behavioral model.
- Both plans performance and their count match contribution can be evaluated together with compound utility formulation.
- Good plans from the calibration perspective can persist along iterations, which influences positively on the calibration feature. A scoring model based solely on travel disutility and activity performance evaluation jeopardizes the existence of plans that are plausible for the counts concordance. That is, plans that are dismissed are usually the ones that are considered the worst from the standard behavioral scoring context,

overlooking their contribution to counts reproduction.

- The calibration effect on the general model can be adjusted. Instead of an explicit brute force setting, the configurable weight can regulate the strength of the calibration in relation to the other scoring parts.

8. Coupling route diversity and Cadyts scoring function

8.1. Approach and implementation

Achieving route diversity through random routes generation might seem inadequate from the classical assignment models perspective. Certainly it would be impractical if it were implemented as a stand-alone module for route choice model. However, its implementation is justified because random paths are created on the base of proved standard travel values as initial seed. But most important, if the search of random candidate solutions is combined with a selection mechanism (like Cadyts correction inside the scoring function) where new alternatives for each agents are evaluated and the worst are discarded, this coupling constitutes a composite co-evolutionary algorithm that directs the choice distribution to a count match convergence. The integration of both approaches is graphically represented in Fig.4 and outlined next.

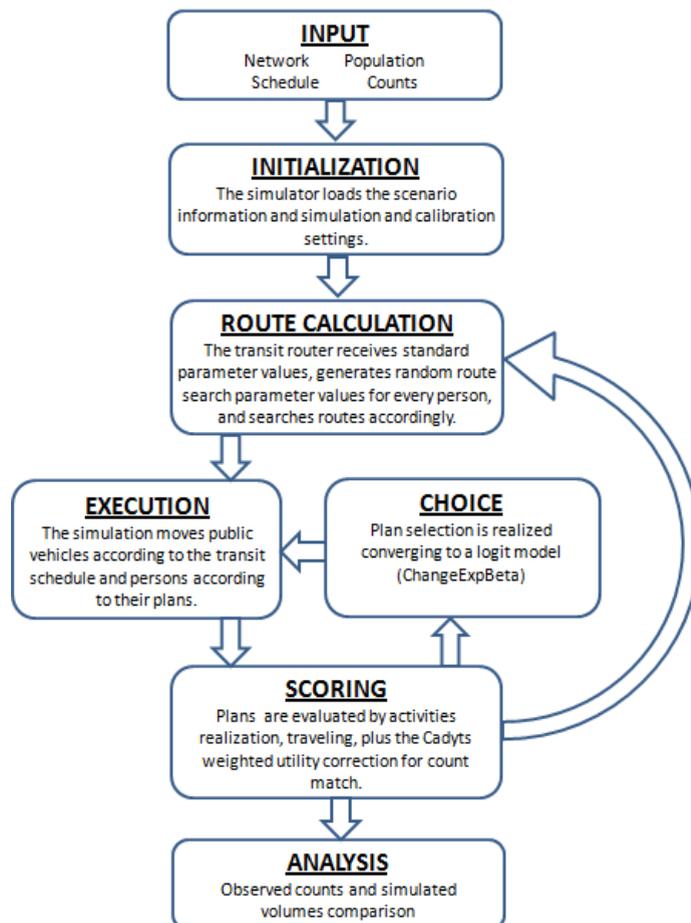


Fig 4. Diagram of randomized routing plus Cadyts inside the scoring function

- Initialization: Usual scenario data are loaded, including the revealed occupancy counts. Some of the configurable parameters in this step are: calibration weight, maximal number of plans per agent, execution probability of each strategy, and use of stop zone conversion for analysis.
- Initial routing: The router generates the first alternative route plan and it is selected.
- Execution: The selected plan is executed.
- Scoring: The executed plan is evaluated according to the accumulated scoring function (Eq. 3).
- Re-planning: New routes based on randomized generalized cost coefficients might be calculated (according to the re-routing strategy probability) and worst plans from behavioral and count convergence perspectives are discarded.
- Selection: A plan is selected and scored if it was never executed before. If all plans of an agent are scored, the *ChangeExpBeta* strategy selects the plan (generating a logit distribution).
- Iteration: The process goes back to “execution”.
- Analysis: It includes the MRE calculation and generation of counts juxtaposition graphs.

For the implementation of the approach on the Greater Berlin scenario, further adaptations and settings are reported:

- Cadyts included since version 1.1.0 a special implementation for MATSim. The so-called “*MatsimCalibrator*” is started necessarily with pre-defined and fixed time bins of 3600 seconds. For this reason its upper, more flexible instance was used instead. The “*AnalyticalCalibrator*” allowed the creation of a calibrator object with the configured 86400 seconds sized time bins per station.
- Calibrated hours: the complete day from 0 to 24 (in concordance to the 24 hours counts).
- Maximal number of plans per agent: 5
- Simulation and calibration settings: Cadyts utility correction is embedded in the scoring function. The scoring performs as usual for each plan that is selected at a given iteration. The simulation re-planning module is configured to distribute its execution probabilities like this: randomized re-routing with 10% until iteration 400 and *ChangeExpBeta* as plan selection strategy with 90%. *ChangeExpBeta* keeps performing until iteration 1000, along with the scoring approach that includes Cadyts correction utility.

8.2. Results

In order to evaluate how Cadyts enforces the counts reproduction, a number of parametric calibration runs is realized over the scenario. The runs are realized with incrementing Cadyts weights (w in Eq.3), namely 0, 1, 10, 100, and 1000 to test the effect of the increasing calibration force over the general accumulated scoring. Fig. 5 shows the analysis in scatter plots for each run. The first plot corresponds to the initial iteration, which is the same for all weights. Then, the output plot (iteration 1000) for each different calibration weight is depicted.

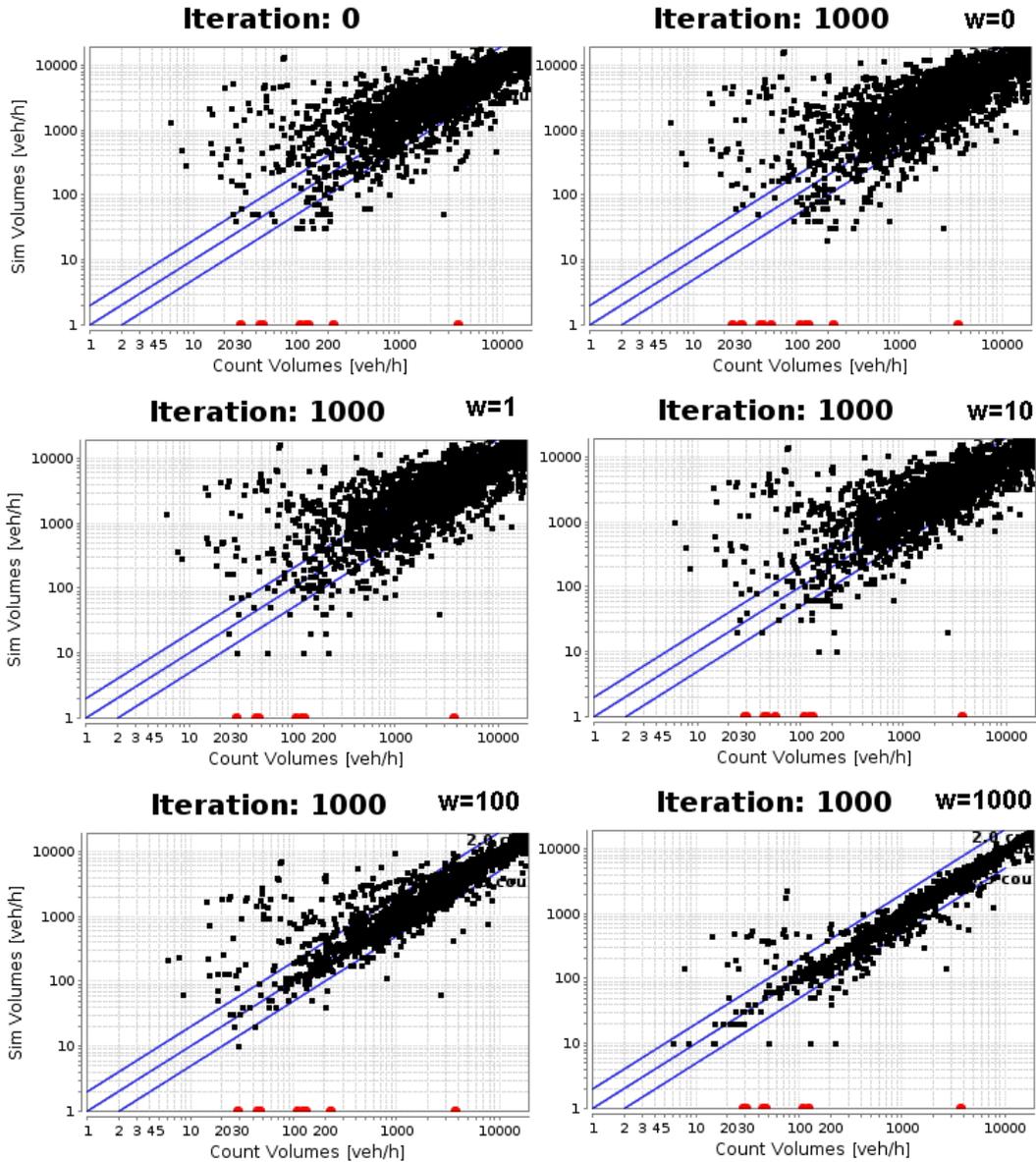


Fig.5. Calibration with randomized parameters route search and Cadyts as part of the scoring function with increasing weight.

One can see that for low calibration weights (like 0, 1, and 10) hardly any improvement is achieved regarding counts match. On the contrary, it is noticeable that a very strong value like 1000 corresponds almost to the brute force calibration.

Another simulation exercise is described next, whereby the synthetic demand is duplicated, that is, each agent is cloned. Moreover, instead of 5 plans per agent, the choice set size was increased to 10. With it, more diverse routes are available for passengers and also for the calibrator. If an excessive number of passenger occupancies are simulated at a stop, some of those passengers can be forced by the calibrator to travel through other stops without counts. The goal for this test is to show how the same calibration settings and the same observed counts

can produce better affinity of observed and simulated counts in demand change scenario. Fig. 6 shows the initial and final plots for the calibration with cloned agents.

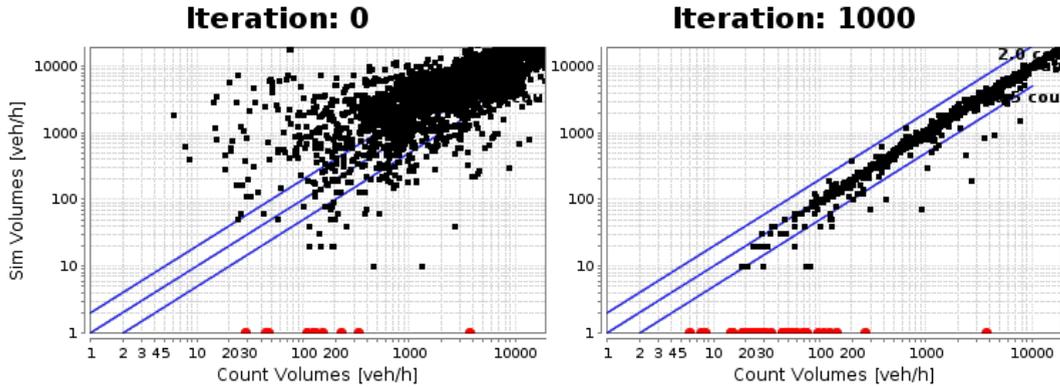


Fig. 6: Calibration of Greater Berlin scenario with duplicated demand and Cadyts weight 1000.

Another analysis consists on tracking the MRE value for observed and simulated counts divergence along iterations. Table 3 shows the initial and final MRE value for all runs presented in this work.

Table 3. Comparison of initial and final MRE values for calibration runs.

Test	Initial MRE	Final MRE
Simulation initial situation (without calibration)	89.6	97.8
Brute Force calibration with fixed choice set	128.5	15.1
Random routing and Cadyts weight 0	100.0	107.2
Random routing and Cadyts weight 1	100.0	104.4
Random routing and Cadyts weight 10	100.0	89.8
Random routing and Cadyts weight 100	100.0	41.5
Random routing and Cadyts weight 1000	100.0	15.7
Random routing and Cadyts weight 1000 with agent cloning	151.3	5.0

One can notice that higher Cadyts weights achieve lower final MRE values. In the same way, Fig. 7 compares graphically evolution of the MRE values along iterations of all calibration tests.

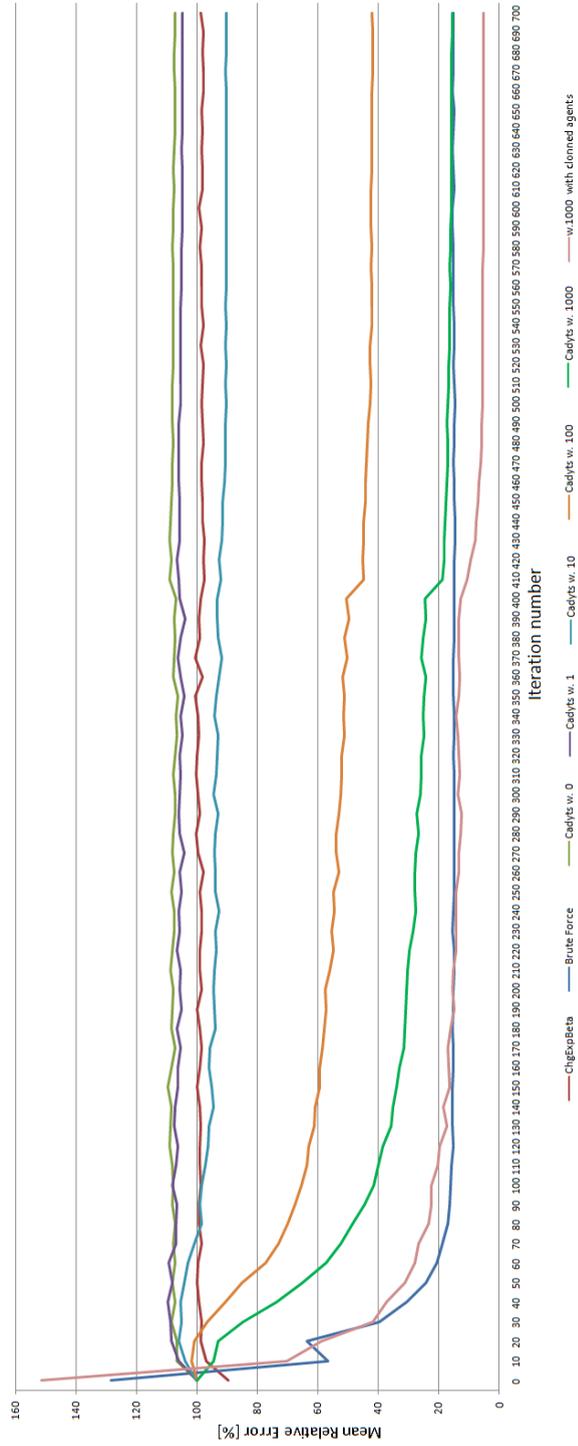


Fig. 7. Evolution of MRE values along iterations with normal simulation, brute force calibration and Cadyts as scoring function with different weights and duplicated demand

The calibration with brute force with fixed choice set starts with a high MRE value (128.5%), in comparison with the value (100.0%) of all runs made with randomized routing. This can be explained by the fact that the fixed choice set starts always with the first plan selected, that corresponds to the pure high transfer resistance. However, the fixed plans and the plans created with random routing tend to stabilize and come relative to close values (42% and 15% respectively) around iteration 600.

The analysis also shows that low Cadyts weight values have barely a visible effect. In contrast, one may see the evident improvement for values 100 and 1000 just after the first iterations. In both strong calibrations, a sudden error reduction that happens just after iteration 400 is noteworthy. It corresponds to the stop of plan innovation (with randomized router) and the start of full calibration. The strongest weight value (1000) deserves special attention because it reaches the same MRE value of brute force calibration from iteration 600 on.

The calibration of duplicated demand starts with the worst count reproduction, but the effect of the strongest Cadyts weight value and the larger number of agents produces at the end the best output.

9. Conclusions and future work

The calibration of a large scenario was carried out, preparing the necessary input data and adapting the integration code between transit simulation and calibration. This is possible thanks to the adaptability of Cadyts (proved with its many different implementations) and the robustness of the transit simulation in MATSim.

The results demonstrate that the calibration approach is able to work with very large scale, real world scenarios, and that it is able to deal with the inter-temporal aspects implied by the available counts.

The next challenge will be how to make these findings useful for prediction. The approach for this will be to extract behavioral parameters, possibly per individual, which would explain behaviorally the choices that are most consistent with the measurements.

Acknowledgements

This work is funded in part by the National Council on Science and Technology of Mexico (CONACyT) and the German Academic Exchange Service (DAAD). The computing calculations are possible with the help of Prof. H. Schwandt's group from the Institute of Mathematics at TU Berlin. Gunnar Flötteröd provided very useful suggestions and support for this work. The Berlin public transit company BVG provided the passenger occupancy counts data.

References

- Balmer, M., Raney, B., and Nagel, K. (2005). Adjustment of activity timing and duration in an agent-based traffic flow simulation. In Timmermans, H., editor, *Progress in activity-based analysis*, pages 91-114. Elsevier, Oxford, UK.
- Charypar, D., and Nagel, K. (2005). Generating complete all-day activity plans with genetic algorithms. *Transportation*, 32(4):369-397.
- Farrol, B. & Livshits, V. (1998). Analysis of Individual Transit Trips in EMME/2: Calibration of 1996 TTC Trips Disaggregate Assignment. *Urban Transportation Research and Advancement Centre*, Report 80.
- Flötteröd, G. (2008). *Traffic state estimation with multi-agent simulations*. PhD thesis, Berlin Institute of Technology.
- Flötteröd, G. (2010). Cadyts manual. Version 1.1.0. http://transp-or.epfl.ch/cadyts/Cadyts_manual_1-1-0.pdf.
- Flötteröd, G. (2013). Cadyts-calibration of dynamic traffic simulations. <http://transp-or.ep.ch/cadyts>. Accessed 2013.
- Flötteröd, G., Bierlaire, M., & Nagel, K. (2011a). Bayesian demand calibration for dynamic traffic simulations. *Transportation Science*, 45(4):541-561.
- Flötteröd, G., Chen, Y., & Nagel, K. (2011b). Behavioral calibration of a large-scale travel behavior microsimulation. Annual Meeting Preprint 11-2890, Transportation Research Board (TRB), Washington D.C.
- Flötteröd, G., Chen, Y., Rieser, M., & Nagel, K. (2009). Behavioral calibration of a large-scale travel behavior microsimulation. In *Proceedings of the 12th Conference of the International Association for Travel Behaviour Research (IATBR)*, Jaipur, India.

- Li, B. (2009). *Markov models for Bayesian analysis about transit route origin-destination matrices*. Transportation Research Part B. Loughborough University, 43:301-310.
- Lu, D. (2008). *Route level bus transit passenger origin-destination flow estimation using APC data: Numerical and empirical investigations*. Master's thesis, The Ohio State University.
- Moyo O., M. & Nagel, K. (2012). *Automatic calibration of microscopic, activity-based demand for a public transit line*. Annual Meeting Preprint 12-3279, Transportation Research Board (TRB), Washington D.C.
- Nagel, K. & Flötteröd, G. (2009). Agent-based traffic assignment: Going from trips to behavioral travelers. In *Proceedings of The 12th Conference of the International Association for Travel Behaviour Research (IATBR)*, Jaipur, India.
- Neumann, A., Balmer, M., & Rieser, M. (2012). Converting a static macroscopic model into a dynamic activity-based model for analyzing public transport demand in Berlin. In *Proceedings of the 13th Conference of the International Association for Travel Behaviour Research (IATBR)*, Toronto, Canada.
- Raney, B. & Nagel, K. (2006). An improved framework for large-scale multi-agent simulations of travel behaviour. In Rietveld, P., Jourquin, B., and Westin, K., editors, *Towards better performing European Transportation Systems*, pages 305-347. Routledge, London.
- Rieser, M. (2010). *Adding transit to an agent-based transportation simulation concepts and implementation*. PhD thesis, TU Berlin.
- Rieser, M. & Nagel, K. (2009). Combined agent-based simulation of private car traffic and transit. In *Proceedings of the 12th Conference of the International Association for Travel Behaviour Research (IATBR)*, Jaipur, India.
- Wen Chi, F. S (2005). *Calibration and validation of transit network assignment models*. Master's thesis, The University of Hong Kong.
- Rongviriyapanich, T., Nakamura, F., & Okura I. (2000). Use of On-Off Counts for OD Estimation an approach towards more cost-effective bus surveys. *Infrastructure Planning Review*, 17:p. 623-632. Japan.