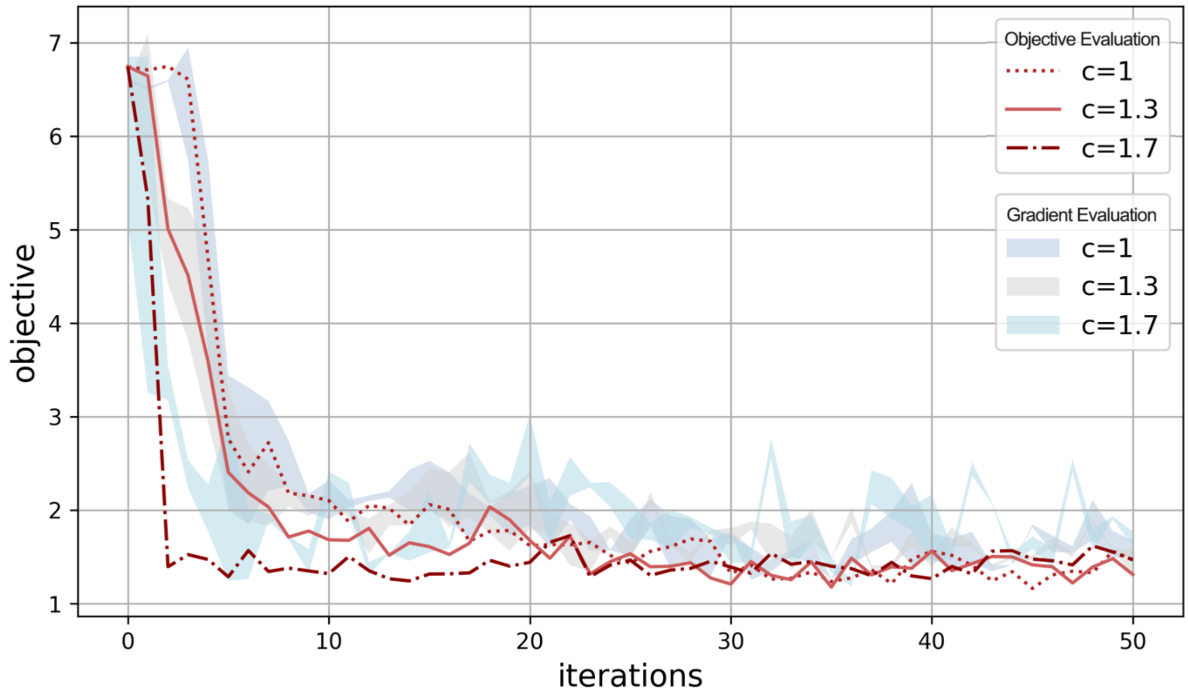


Calibration of Agent-based Transport Simulation with SPSA method

Hong Liang

Master thesis
Spatial Development and Infrastructure Systems

July 2018

Institut für Verkehrsplanung und Transportsysteme
Institute for Transport Planning and Systems

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Acknowledgment

I would first like to express my special thanks of gratitude to my thesis advisor Prof. Dr. Kay Axhausen as well as my supervisor Dr. Basil Vitins and Sebastian Hörl who gave me the golden opportunity to do this interesting thesis and offered me so many useful advice, supports and insightful comments.

Secondly, I would also like to thank my parents and friends, especially Tian, who was always by my side to give me support when I was under pressure and helped me a lot in finalizing the thesis in the limited time frame.

Master Thesis

Calibration of Agent-based Transport Simulation with SPSA Method

LIANG Hong
IVT, ETH Zürich
Stefano-Franscini-Platz 5, 8093
Zürich

Phone: +41 78 951 21 62
E-mail: hliang@student.ethz.ch

July 2018

Abstrct

Nowadays, transport simulation becomes an indispensable tool for transport planners, traffic engineers and also policy decision makers. Calibration is a crucial step in building a reliable simulation model. The calibration here refers to the parameter optimization. The goal of the optimization is to make the analyzed model output as close as possible to the observations made in reality. In this thesis, SPSA (Simultaneous Perturbation Stochastic Approximation) is chosen to be the calibration algorithm to solve the calibration problem. In order to minimize the difference of mode share distribution of census data and simulation output, eight parameters of the utility function in MATSim are calibrated. Golden Section Method (GSM) is also applied combined with SPSA to improve the efficiency of the algorithm. As a result, SPSA is proved to be an efficient method to solve calibration problem for agent-based model.

Key words:

Calibration; SPSA; agent-based model; MATSim

Preferred citation style

Liang, H. (2018) Calibration of agent-based transport simulation with SPSA method, *Master Thesis*, IVT, ETH Zurich, Zurich.

Table of Content

List of Tables

List of Figures

Acronyms and Abbreviations

ABDM	Agent-based demand model
DTA	Dynamic Traffic Assignment
GSM	Golden Section Method
MATSim	Multi-Agent Transport Simulation
SPSA	Simultaneous Perturbation Stochastic Approximation

1 Introduction

In modern life, traffic systems are more and more complex due to the rapid urban growth and accompanying increased demand for transport. The travel behavior and decisions of each influence the traffic pattern. Considering the large population in the city, the complexity of traffic system makes the prediction and management of transport system enormous challenges. Therefore, transport simulation is nowadays getting indispensable for planning, designing and operating transportation system. With the traffic simulation, the effect of possible measures on different situations can be tested, which usually aim at optimizing the traffic flow, increasing accessibility or decreasing congestion and pollution. However, before that, a satisfactory traffic model should first be an accurate representation of the observed traffic conditions. Therefore, model calibration is a vital step in building a reliable traffic simulation. A complicated model usually requires various input parameters. The calibration here refers to the parameter optimization, through which, the analyzed model output is as close as possible to the observations made in reality.

Transport simulation has many categories. While the traditional four-step process has been used as a modeling tool for many decades, the agent-based demand models (ABDM) is gaining popularity and shows the prominent advantages of flexibility and accuracy. Instead of generating aggregative origin-destination (OD) matrices, ABDM is a fully disaggregate approach to simulate every traveler individually, which enables an ex-post analysis of arbitrary demand segments. [15] For instance, MATSim is an agent-based dynamic traffic assignment model, where the agents optimize their daily activity plans according to the score. The score is calculated using MATSim's scoring function. The detailed procedure is explained in Section 4. The parameters of the scoring function determine the travel behavior of each agent and the traffic pattern of the whole system, which are, therefore, the focused calibration variables in this thesis. However, due to the complex behavior and the long running time of MATSim, efficiency is particularly crucial for calibration. Simultaneous perturbation stochastic approximation (SPSA) has been first developed by Spall [25], which addresses explicitly large-scale, stochastic problems. Compared to other large population-based global optimization methods, SPSA is attractive because of its efficient gradient approximation by perturbing all variables at once, thus requiring much less run time. The algorithm process is introduced in Section 3.

With the help of new technologies, the data can be gathered by sensors, traffic counts, GPS data and other information sources and gives us a lot of details about real traffic conditions. In this thesis, the main focused traffic characteristic is the mode share distribution along travel distances, which is one of the essential features of traffic patterns. Hence, the calibration problem in this thesis is the problem of finding a parameter set of the utility function in MATSim, with which, the error of the mode share distribution of simulation output and sensor data can be as small as possible.

This research aims to implement SPSA into MATSim, test the feasibility of the calibration algorithm and improve the efficiency by the modification of the step rules. The rest of the article is organized as follows. In the next section, the related work about the calibration problem for DTA model and the application of SPSA algorithm is reviewed. Section 3 introduces the general SPSA algorithm and the possible modification to improve the performance of the algorithm. In Section 4, the fundamental principle and functions of MATSim are presented including the general structure of MATSim, the ABDM, and the utility function. After the Section 5 of the description of the study case Zurich, in Section 6, the specific implementation of SPSA is proposed, which includes the procedure and the discussion of the parameter and objective function. Lastly, the results, analysis, and summary are shown in Section 7 and 8.

2 Literature Review

For solving the problem of calibrating Dynamic Traffic Assignment (DTA) model, many calibration methodologies have been explored. Because of the nonlinear nature of calibrating demand and supply parameters in DTA system, extended Kalman filter (EKF) was explored comparing also with limiting EKF (LimEKF) and unscented Kalman filter. [2] Frederix (2011) proposed the use of marginal computation (MaC) with the use of kinematic wave theory principles to estimate dynamic O-D matrix on congested networks using a DNL or DTA model. [16] Also to achieve the accurate estimation and prediction of O-D matrix, another application of principal component analysis (PCA) was introduced by Djukic (2012) with dramatically reduced computational costs. [11]

Spall (1992) developed simultaneous perturbation stochastic approximation (SPSA), which computes a gradient approximation calculating only two measurements of objective function per iteration, regardless of the dimension of the calibration problem. This technique is especially attractive when dealing with a large number of parameters simultaneously and solving problems where the objective function is not precise, which is the case for DTA model calibration. Some works of calibration with SPSA algorithm were successfully achieved in both microscopic traffic simulation applications [3] and mesoscopic simulation-based DTA systems [4] by Balakrishna. Ben-Akiva (2012) also proved the feasibility of calibrating over forty thousand parameters simultaneously in DTA system with SPSA. [6] Vaze (2011) compared SPSA with other calibration methods, for instance, genetic algorithms (GA). As a result, SPSA was shown to be the most effective algorithm for DTA system and small network. [27]

In spite of the widely accepted virtues of SPSA algorithm, researchers also proposed some modifications to improve the performance of the algorithm. Lee and Ozbay (2009) introduced an enhanced SPSA in conjunction with the Bayesian sampling approach, which enables the user to enhance calibration by considering statistical data distribution. Another enhanced SPSA algorithm was raised by Lu (2015), called weighted SPSA, which incorporates the information of spatial and temporal correlation in a traffic network to limit the impact of noise and improve convergence rate and robustness.[22] Cipriani (2011) also described three modifications: (1) replacement of basic SPSA step rule by the Golden Section Method (GSM); (2) the use of third degree polynomial interpolation (PI) of objective function to compute the new solution along the descendant direction; (3) Asymmetric design (AD) during the gradient approximation.

Specific to the problem of calibrating MATSim, Cadyts (Calibration of Dynamic Traffic Simulations) [12] and Opdyts (Optimization of Dynamic Traffic Simulations) [13] are the two main approaches developed by Flötteröd. Cadyts calibrates DTA model by adjusting the plan choice probabilities for all agents, while Opdyts identifies the best parameters out of a candidate set without running the MATSim simulation until convergence for every single parameter setting. Flötteröd (2012) proposed a Bayesian optimization for solving calibration of a large-scale travel microsimulation. [15] Beyeler in his latest work also introduced a practical multi-fidelity Bayesian batch optimization for calibration of DTA models and applied it into MATSim. [9]

3 Heuristic Calibration Methods

3.1 General SPSA

SPSA (Simultaneous Perturbation Stochastic Approximation) is a stochastic gradient approximation algorithm. Compared with other path search algorithm, it is able to minimize the error function when the relationship between the objective function and the variables to be optimized is unknown and can only be estimated with noisy observations. SPSA starts from an initial estimation of the variable vector and iteratively traces a sequence of variable estimations which make the objective function converges to zero based on the gradient approximation. Its greatest advantage is the efficiency of calibrating a large number of parameters simultaneously. Because in SPSA all of the variables in the decision vector are perturbated at the same time, and the approximation of gradient needs only two function evaluations regardless of the number of variables.

The standard iterative form of SPSA is:

$$\theta_{k+1} = \theta_k - a_k \hat{g}_k(\theta_k) \quad (1)$$

where θ_k is the estimate of the variable set in the kth iteration of the algorithm, $\hat{g}_k(\theta_k)$ is the approximation of the gradient at θ_k. a_k is a non-negative coefficient, which determines the kth step size and usually gets smaller as k becomes larger. The standard step size rule has the following form:

$$a_k = \frac{a}{(A+k+1)^{\alpha}} \quad (2)$$

where a, A and α are the algorithm parameters.

The approximation of gradient is the essential step of the stochastic approximation. The traditional approach perturbs each variable in θ_k one at a time, while the SPSA approach perturbs all of the variables at the same time. Assuming that θ is m-dimensional, the gradient approximation step can be expressed as follows:

$$\hat{g}_k(\theta_k) = \begin{bmatrix} \frac{z(\hat{\theta}_k + c_k \otimes \Delta_k) - z(\hat{\theta}_k - c_k \otimes \Delta_k)}{2c_k \Delta_{k1}} \\ \vdots \\ \frac{z(\hat{\theta}_k + c_k \otimes \Delta_k) - z(\hat{\theta}_k - c_k \otimes \Delta_k)}{2c_k \Delta_{km}} \end{bmatrix}$$

$$= \frac{z(\hat{\theta}_k + c_k \otimes \Delta_k) - z(\hat{\theta}_k - c_k \otimes \Delta_k)}{2c_k} [\Delta_{k1}^{-1}, \Delta_{k2}^{-1}, \cdots, \Delta_{km}^{-1}]^T \qquad (3)$$

where $z()$ is the objective function. As SPSA is an aggregated optimization algorithm, the value of the objective function should be a scalar instead of a vector. The concrete definition of objective function for this specific MATSim calibration problem is explained in Section 6.2. While the approximation of gradient, θ_k is perturbed by the value of Δ_k times c_k, where Δ_k is an m-dimensional perturbation vector. Its components are chosen randomly as either -0.5 or 0.5 with the same probability of ½. Also, c_k is a positive scalar and has the similar form to a_k that usually gets shrunken when k is larger, but it determines the amplitude of the perturbation:

$$c_k = \frac{c}{(k+1)^{\gamma}} \qquad (4)$$

The sequences a_k and c_k are critical elements in the optimization process. c_k defines the region where two measurements of the objective function are calculated to compute the gradient approximation. If a_k or c_k are chosen too small, the algorithm might be stuck in the current position regardless of whether it is optimal or not. To the contrary, if a_k is too larger, the algorithm might take a large step far away from the optimal solution. Also, an over-large c_k might lead to the algorithm not converged after many iterations. Spall (1992) proved that the SPSA algorithm converges when the parameters of the algorithm fulfill some conditions. But within the framework, the user is free to choose the parameters. These conditions include:

$$\alpha - 2\gamma > 3\gamma - \frac{\alpha}{2} \geq 0$$

$$\alpha > 0; \ \gamma > 0$$

$$c > 0; \ a, A > 0$$

The more detailed conditions are given in Spall (1992) [25]. The proposed value for α and γ are chosen to be 0.602 and 0.101.

A two-dimensional example of SPSA algorithm is explained in Appendix B two-dimensional example of SPSA algorithm.

3.2 Golden Section Method

Since the parameters of the SPSA algorithm are crucial for the effectiveness, the basic step rule can return an uneven convergence of the objective function with relatively poor results even after many iterations when the parameters are not chosen suitably. Therefore, each calibration problem requires users put effort into testing and adjusting the calibration parameters. In order to improve this ineffectiveness, Cipriani (2011) proposed the replacement of the basic SPSA step rule by the Golden Section Method (GSM) [10]

GSM is a technique which iteratively narrows the search area. In the beginning, two points are chosen to define the range where an optimum is searched. In this calibration problem, the first point p is the current vector θ_k, and the second point q is the candidate vector in the next iteration in the standard SPSA iteration form as follows:

$$\begin{aligned} p &= \theta_k \\ q &= \theta_k - a_k \hat{g}_k(\theta_k) \end{aligned} \tag{5}$$

where a_k and $\hat{g}_k(\theta_k)$ still satisfy the Equation (2) and Equation (3) respectively.

A high value of a_k means that a large area needs to be searched, which leads to more number of simulations. A small a_k might make the search area is too small to find the minimum in it. Commonly, a is chosen so that a_k is equal to the maximum step size, which means the maximum change of the variables in one iteration. As the standard step size rule, the maximum step size also decreases during iterations.

Once the points p and q are determined, two inner points x_1 and x_2 are calculated based on the golden ratio

$$\varphi = \frac{\sqrt{5}-1}{2} \tag{6}$$

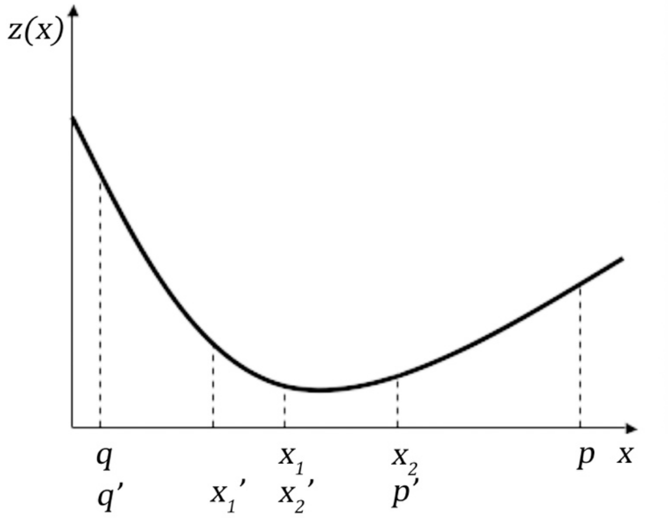
so that

$$\begin{aligned} x_1 &= p - \varphi(p-q) = \theta_k - a_k \hat{g}_k(\theta_k)\varphi \\ x_2 &= q + \varphi(p-q) = \theta_k - a_k \hat{g}_k(\theta_k)(1-\varphi) \end{aligned} \tag{7}$$

Now the objective value $z(x_1)$ and $z(x_2)$ can be found and compared. There are two different cases: $z(x_1) \geq z(x_2)$ and $z(x_1) < z(x_2)$.

If $z(x_1) < z(x_2)$, then the interval where the minimum lies is expected to be $[x_1, p]$, otherwise the minimum occurs in $[q, x_2]$. The interval is the new search area for the next iteration, and the procedure is repeated for the new point p' and q'. An example of this procedure for one-dimensional θ is shown in Figure 1.

Figure 1 Illustration of the GSM procedure



This procedure is repeated until the interval is reduced to the level of a chosen value ε. To make the solution more and more accurate during iterations, ε is also set to decrease with the number of iterations.

4 Deloyed Silumation System (MATSim)

MATSim is an activity-based, multi-agent simulation framework implemented in JAVA. [19] In this section, the basic concepts behind MATSim and its framework are introduced.

4.1 Agent-based Demand Model (ABDM)

Demand model is the central part of traffic simulation. Theoretically, travel is viewed as derived from the demand for activity. However, in practice travel has been modeled with trip-based rather than activity-based methods. As the most classic trip-based demand model, the four-step model has been used in traffic simulation for many years. The four-step model, including trip generation, trip distribution, modal split, and assignment, uses origin-destination (OD) matrix as the principal database rather than activity surveys. In other words, the activity characteristics and the information of each traveler are ignored in this model.

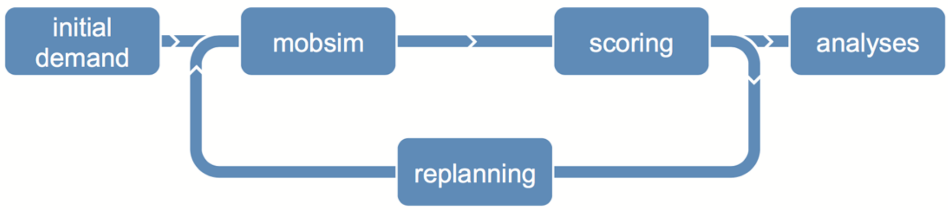
Activity-based travel theory was started by Hagerstrand (1970) when he first elaborated the temporal-spatial constraint of traveling and indicated that the aim of travel is enabling people to engage in desired activities. [18] The basic ideas of activity-based travel theory are (1) the demand for travel is derived from demand for the activities which require travel; (2) human behavior is constrained in time and space; (3) the household significantly affects individual activity and travel decisions and (4) activity and travel decisions occur dynamically.

Based on this, agent-based demand model came into use. The meaning of “agent” here is a traveler modeled in the simulation and all agents build the population. Each agent has its daily plan of activity chain, which includes the information of activity type, departure time, activity duration, route, mode, destination and so on. Instead of just producing the traffic, the agents try to manage their day in a profitable way. Through the individual decision-making process of modification of the activity chain, each agent tries to optimize its plan. The agent travels in the system according to the plan and interacts with each other. The derived traffic is generated. The major advantage of agent-based model, compared to the traditional four-step model, is the ability to simulate each traveler individually. The more explicit modeling is not just a more precise representation of the real world, but also offers more opportunities to make various ex-post analysis and research.

4.2 MATSim framework

MATSim is an activity-based, multi-agent simulation framework, which is designed for large-scale scenarios. As introduced in Section 4.1, every agent tries to optimize its daily activity schedule iteratively. In MATSim, the optimization is based on the co-evolutionary principle. Due to the limited transportation infrastructure, all agents compete with all other agents and change their plan accordingly. Then after iterations, an equilibrium is reached eventually, where the agents cannot further improve their plans. In other words, the result of the co-evolutionary algorithm in MATSim is user equilibrium rather than system equilibrium.

Figure 2 MATSim loop



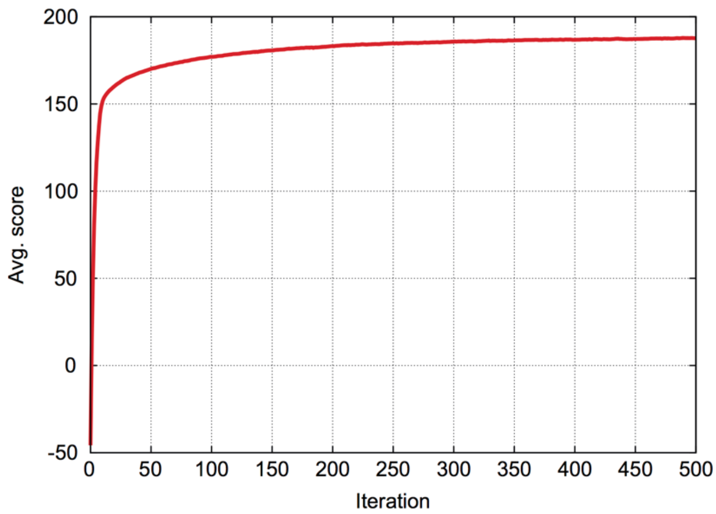
Source: Horni, A., K. Nagel and K. W. Axhausen, (2016). *The Multi-Agent Transport Simulation* MATSim, Ubiquity Press, London

The entire MATSim iterative process is presented in Figure 2. It starts with an initial demand, which is generated from the initial activity chains of all agents in the study area. The initial activity chains are usually derived from empirical data through sampling or discrete choice modeling. In every iteration, each agent selects a plan from its memory, which contains a fixed number of plans. After the simulation of the network loading with “mobsim” (mobility simulation), each agent gets a score according to the executed plan, which can differ from the initial selected plan at the beginning of the iteration. The agent sometimes has to adjust the activity chain due to the result of interaction with other agents in the system, for example, the congestion. The score can be interpreted as an econometric utility. (See Section 4.3)

On the replanning stage, a certain share of the agent is allowed to clone the selected plan and modify it and store the new plan in its memory. The considerations of modification in MATSim can be departure time, route, mode, and destination. The further possible considerations, such as activity adding or dropping, are currently under development. This process corresponds to the decision making by each agent. Once the number of plans in the agent’s memory reaches the ceiling, the plan with the lowest score is removed from the memory. The agent, who has no chance to modify the plan, selects the best from the existing plans in the memory.

An iteration ends up with the evaluation of the agents' experiences and scoring the executed plan. The scoring function, also called the utility function, is explained in Section 4.3. The MATSim loop is repeated until the average population score gets stabilized, which means the iterative simulation gets to the closed point of user equilibrium. A typical score curve is shown as Figure 3.

Figure 3 Typical score progress



Source: Horni, A., K. Nagel and K. W. Axhausen, (2016). *The Multi-Agent Transport Simulation* MATSim, Ubiquity Press, London

As the result of the simulation, the output files of MATSim include Log file, score statistics, leg travel distance statistics and so on. Besides these summarized output information, every action in the simulation generates an event, including "an agent finishes an activity", "an agent starts a trip", "a vehicle enters a road segment", "a vehicle leaves a road segment", "an agent arrives at a location" and so on. These events and plans of agents are also recorded for analysis.

4.3 Utility function

The score is to describe the performance of a plan. A higher score means better performance, to the contrary, a worse plan corresponds to a lower score. However, whether the performance of a plan is good or not, is decided by the travelers' preference. In other words, some people may prefer a crowded but fasted trip rather than a comfortable but slower trip, while some may choose the other. Therefore, the typical way to describe the subjective performance is to use econometric utility functions.

In MATSim, the basic utility functions are founded by Charypar and Nagel (2005). The utility of a plan S_{plan} is calculated as the sum of all activity utilities $S_{act,q}$ plus the sum of all travel utilities $S_{trav,mode(q)}$:

$$S_{plan} = \sum_{q=0}^{N-1} S_{act,q} + \sum_{q=0}^{N-1} S_{trav,mode(q)} \qquad (8)$$

where N is the number of activities, q is the trip that follows activity q.

The utility of an activity q is calculated as the sum of utilities of performing activity $S_{dur,q}$, the spend waiting time $S_{wait,q}$, the penalty for not staying long enough $S_{early.dp}$ and the penalty for a "too short" activity $S_{short.dur,q}$:

$$S_{act,q} = S_{dur,q} + S_{wait,q} + S_{early.dp} + S_{short.dur,q}: \qquad (9)$$

While the agents earn positive utilities during activities, traveling, on the other hand, costs people money and time. Therefore, the utility of a trip is usually negative. The travel utility for a trip q is given as:

$$\begin{aligned} S_{trav,q} &= C_{mode(q)} + \beta_{trav,mode(q)} \cdot t_{trav,q} + \beta_m \cdot \Delta\, m_q \\ &+ \left(\beta_{d,mode(q)} + \beta_m \cdot \gamma_{d,mode(q)}\right) \cdot d_{trav,q} + \beta_{transfer} \cdot x_{transfer,q} \qquad (10) \end{aligned}$$

where:

$C_{mode(q)}$ is a mode-specific constant,
$\beta_{trav,mode(q)}$ is the marginal utility of traveling time by mode,
$t_{trav,q}$ is the travel time between activity location q and $q+1$,
β_m is the marginal utility of money,
$\Delta\, m_q$ is the change in monetary budget caused by fares, or tolls for the complete leg,
$\beta_{d,mode(q)}$ is the marginal utility of distance,
$\gamma_{d,mode(q)}$ is the mode-specific monetary distance rate,
$d_{trav,q}$ is the travel distance between activity location q and $q+1$,
$\beta_{transfer}$ is public transport transfer penalties,

$x_{transfer,q}$ is a binary variable representing whether a transfer occurred the previous and current leg.

The calculation of the utility function is the basis of the decision-making process of each agent. These parameters of the utility function decide what kind of plans are good and affect directly the individual decision of each agent, which is the determinant of traffic demand. However, those parameters are usually hard to obtain from observations or experience directly. The calibration is, therefore, extremely crucial.

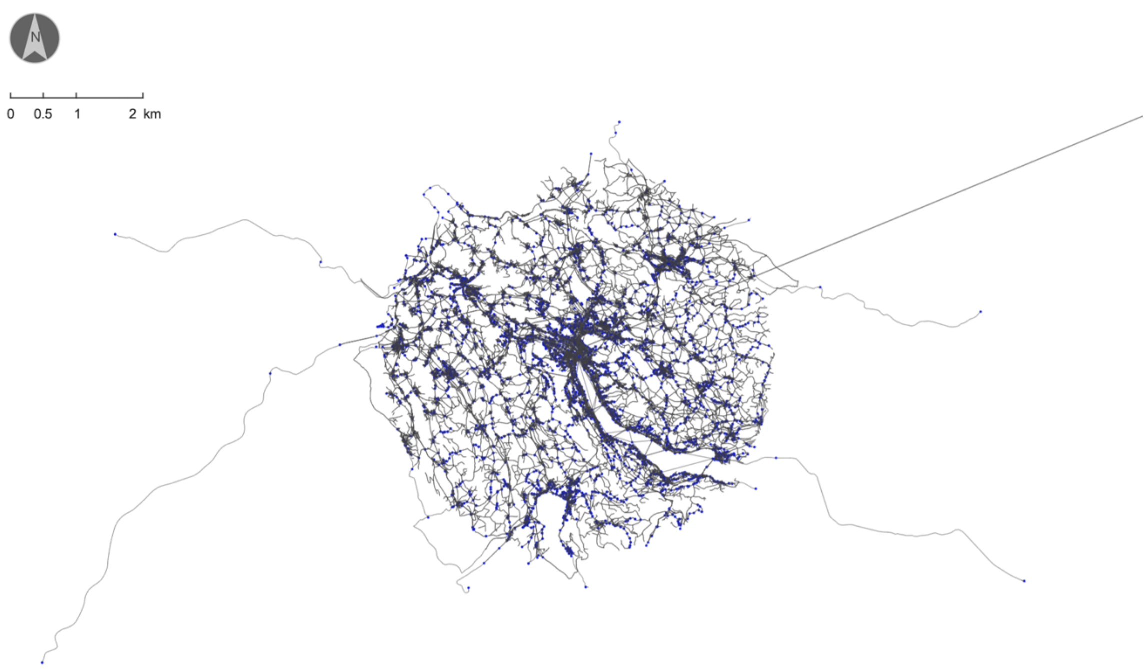
5 Case Study Zurich

For the implementation of the algorithm, two different scenarios of Zurich are used. Both of them are provided by the Institute for Transport Planning and Systems (IVT) at ETH Zürich. The small-scale Zurich scenario of 0.1% population serves as a test-case scenario to keep running time at an acceptable level, whereas the Zurich scenario of 10% population is used to evaluate the robustness and the performance of the algorithm on a larger scale scenario.

5.1 Network

The study area is confined by a circle with a radius of 30 kilometers and a center at "Bellevue", which is located in the central area of Zurich. The network consists of over 70'000 nodes and over 157'000 directional links, which includes the information of location, direction, length, free speed, capacity and the travel modes allowed on the link. For the public transport, there are over 8'000 stop facilities and 648 transit lines in the system.

Figure 4 The network of Zurich scenario



Source: The Zurich scenario provided by IVT, ETH Zürich

5.2 Demand

In the activity-based model, the travel demand is derived from the activities required by agents. In order to obtain a higher computational speed, 0.1% and 10% population are chosen for two scenarios. The activities are executed at the facilities in the system. Different facilities correspond to different activity types, including home, work, education, leisure, and shop. Due to the decrease in population, the number of households and facilities should also be reduced correspondingly. The specific number is as follows:

Table 1 Description of Zurich scenarios

Scenario	Population	Household	Facility
0.1% sample	1566 agents	1565 households	4363 facilities
10% sample	158485 agents	149592 households	147563 facilities

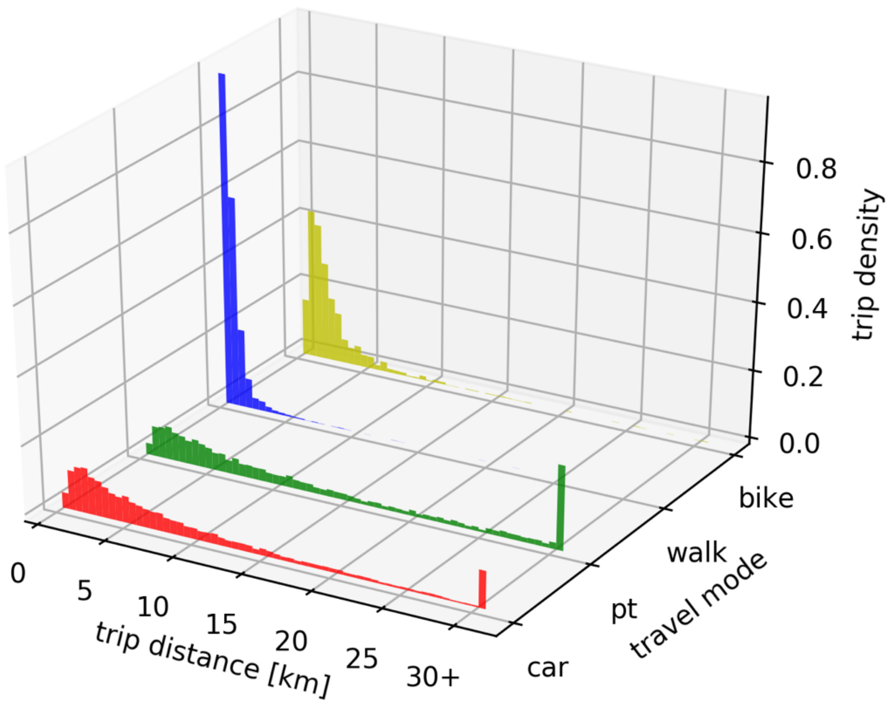
5.3 Configuration file

The configuration options of MATSim can be described and set in the configuration file. Specifically, the attributes and parameter of data containers, global modules, mobility simulation, replanning strategies and observational modules can be defined. Importantly, the iteration count is set to 80, where the average score of the population is already getting stable. It is also worth mentioning that the flow capacity factor of the mobility simulation (*qsim*) needs to be set according to the sample scenarios. For example, the flow capacity factor should be 0.1 for 10% sample, and 0.001 for 0.1% sample. At the replanning stage, the replanning strategy needs to be defined. In this thesis, the change of travel mode is allowed with the weight of 0.2.

5.4 Census data

The census data is collected from the microsensors in the region of Zurich. The data records 148'092 trips, which include the information of travel distance, travel mode and the coordinate of the start and end point. Among these trips, there are 23'641 trips that departure and also arrive in the study area.

Figure 5 Trip count distribution of census data for different modes



As shown in Figure 5, the trip distributions by trip distance are entirely different for four different modes. The distributions of walk and bike concentrate in the short distance. Over 70% of walking trips are shorter than 1kelometer. To the contrary, the distributions of car and public transport are more even along travel distance. But for long-distance trips, there are still more than 10% and 20% of trips by car and by public transport respectively over 30 kilometers.

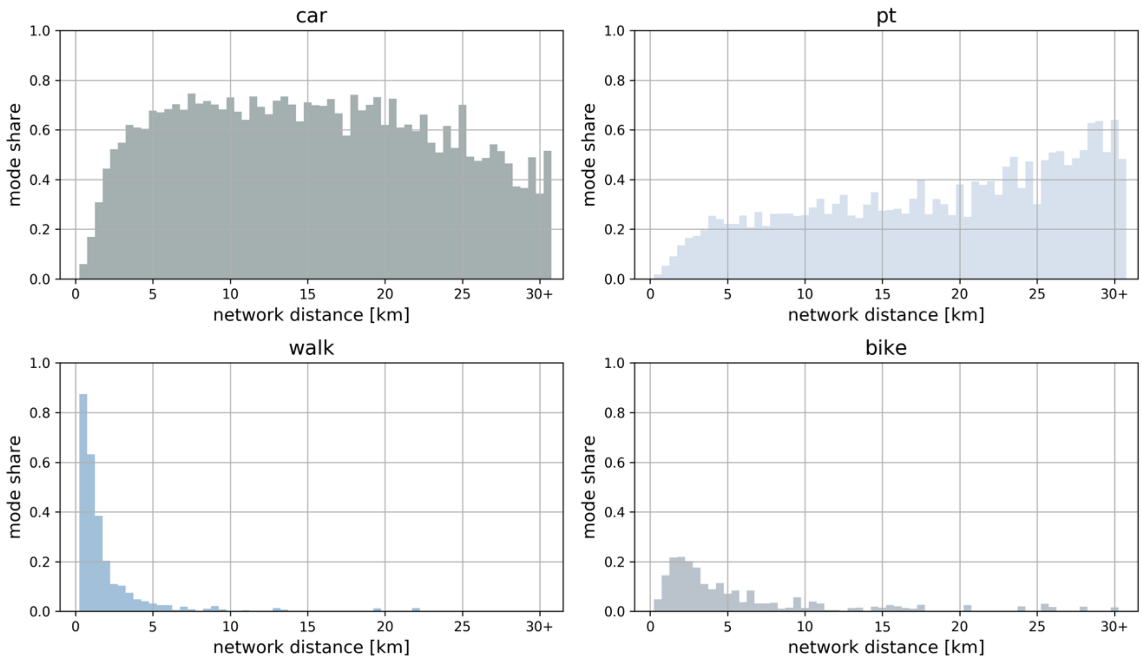
Mode share is an important indicator in developing a sustainable transport within a city or a region. A mode share $S_{m,d}$ is the percentage of travelers using a particular type of transportation, which can be written as follows:

$$S_{m,d} = \frac{n_{m,d}}{\sum_{m \in M} n_{m,d}}; \qquad (11)$$

where $M = \{car, public\ transport, walk, bike\}$, $n_{m,d}$ is the number of trips by mode m at distance d.

The mode share distribution by distance reflects how travel distance affects travelers' decision of mode choice. Figure 6 shows the result from census data. Private car is the most attractive travel mode within most distances, around half of travelers travel by car in this area. As the second attractive travel mode in general, people have more intentions of taking public transport for long distances than short distances. Walk and bike represent a small fraction of all trips. Most traveler (over 70%) choose to walk within 1 kilometer, while travelers prefer riding a bike between two and five kilometers.

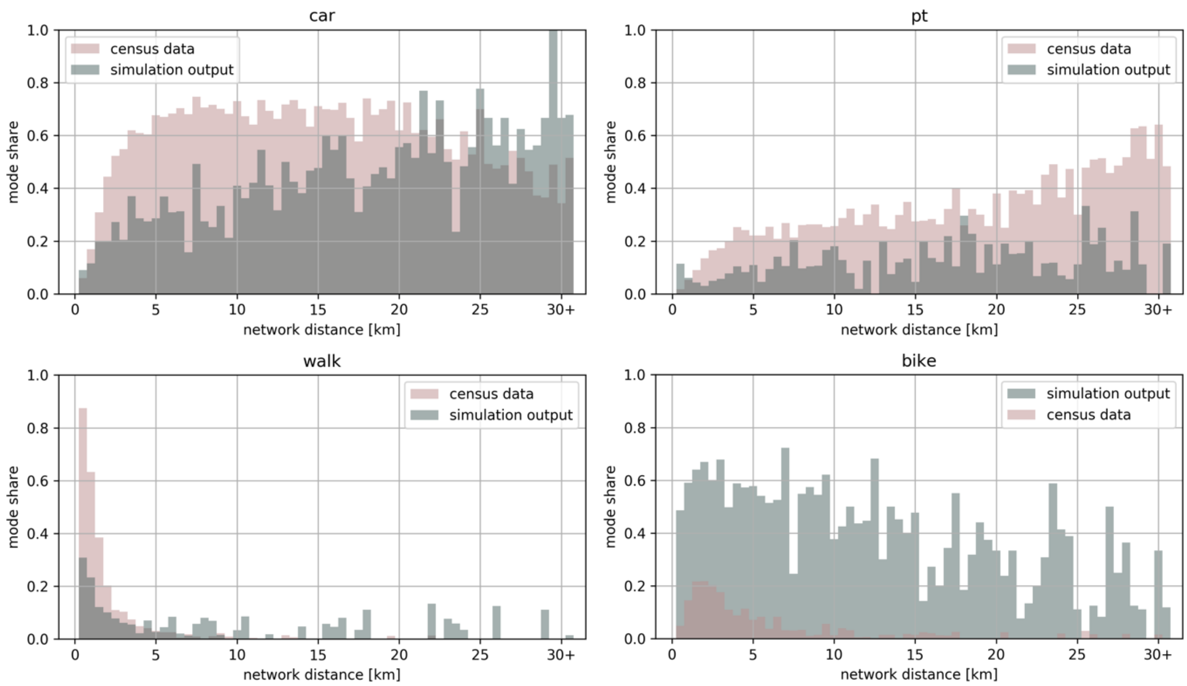
Figure 6 Mode share distribution of census data for different modes



5.5 Uncalibrated simulation

The uncalibrated result is the output of the simulation with the initial parameters of the utility function. As shown in Figure 7, the travel mode of bike is much more attractive than the census data due to the unsuitable parameters of the utility function. The huge distance between census data and simulation output need to be reduced as much as possible by the calibration process.

Figure 7 Comparison between mode share distribution of census data and uncalibrated simulation output



6 Calibration

The calibration process with SPSA algorithm is implemented in the framework of MATSim in JAVA. The calibration procedure, variables, objective function and the measurements of performance are clarified in this section.

6.1 Calibration variables

As introduced in section 3.1, the major advantage of SPSA algorithm over other stochastic approximation is the ability to calibrate a large number of variables at the same time in an efficient way. However, in order to implement the algorithm and test the feasibility of calibrating MATSim, the study starts with a small number of variables. In this thesis, which is the first stage of the study, the calibration variables are chosen to be the constant parameter $C_{mode(q)}$ and the marginal utility by traveling time $\beta_{trav,mode(q)}$ in Equation (10). So, there are eight variables need to be calibrated in total considering the four travel modes in the system, which are:

C_{car}, $\beta_{trav,\,car}$, C_{pt}, $\beta_{trav,\,pt}$, C_{walk}, $\beta_{trav,\,walk}$, C_{bike}, $\beta_{trav,\,bike}$.

Since the constant parameter $C_{mode(q)}$ has no physical meaning in reality, it can be set as any value from negative infinity to positive infinity. Whereas the marginal utility $\beta_{trav,mode(q)}$ has to be non-positive. In the MATSim system, trips are regarded as the derivate of the demand for activities. While the agents can work to earn money and go hiking for health and pleasure during activities, they have to spend time and money on the trips. Therefore, $\beta_{trav,mode(q)}$ should be negative to represent the disutility of trips.

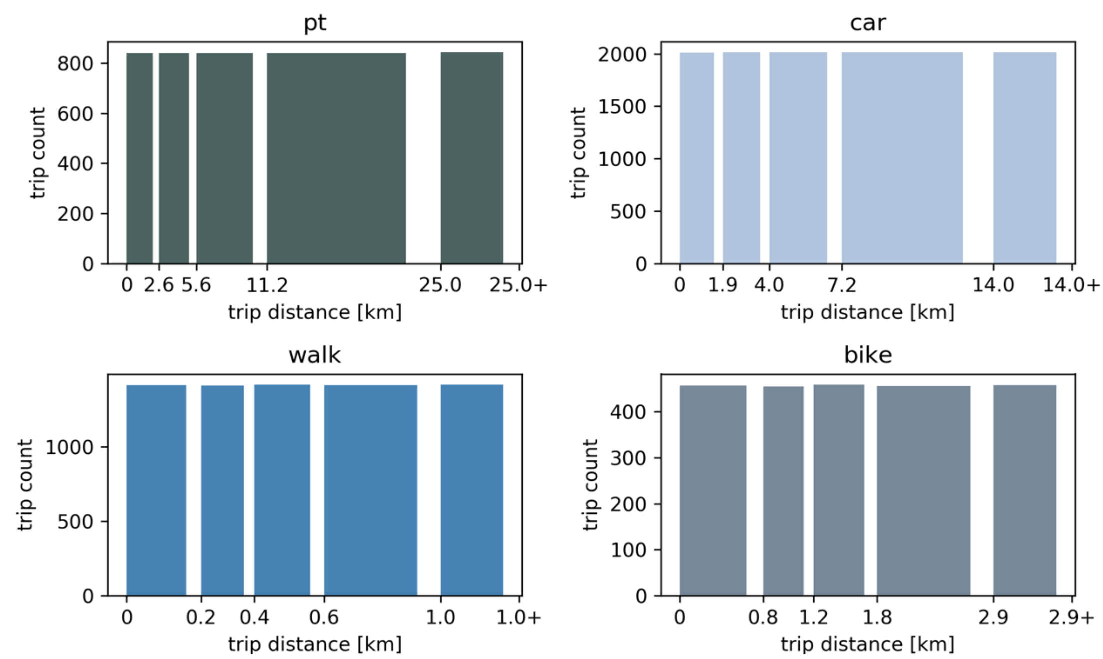
6.2 Objective function

The choice of calibration objective plays an essential role in the calibration problem. Because it determines what values are required to be minimized in the process. In general, the goal of the calibration process in this thesis is to minimize the errors of mode share distribution between census data and simulation output. As explained the Equation (3) in Section 3.1, the result of the objective function $z()$ should be a scalar. Hence, as an aggregated result, it is not worth of calculating all the distance between census data and simulation output in each small distance bin as shown in Figure 7. In other words, to simplify the problem, the number of bins for mode share distribution should be reduced, and the suitable bin size needs to be found.

As described in Section 5.4, the mode share distributions in reality are skewed and concentrate at the different distances for different modes. Therefore, it is reasonable to have uneven and different bin sizes for four travel modes.

In general, more bins with smaller bin size mean a higher resolution and a higher accuracy. For example, the most trips by walk occur when the trip distance is less than 2 kilometers. Therefore, the bins at the short distance should be narrower than at the long distance. A suitable way to determine the bin size is to find the boundaries that split the data into a predefined number of buckets containing an equal number of trips. In this thesis, the number of bins is chosen to be five. Hence, each bin should contain exactly 20% trips for each mode as shown in Figure 8.

Figure 8 Equi-depth histogram of trip distribution for different modes



As there are five bins for each of four travel modes, twenty differences between census data and simulation result need to be minimized. So, the objective function can be determined in different ways. For now, two options are under consideration, which are the sum of twenty absolute errors (Equation 12) and the maximum error among them (Equation 13).

$$z(\hat{\theta}_k) = \sum_{m \in M, d \in D} \left| S_{m,d} - \tilde{S}_{m,d} \right| \qquad (12)$$

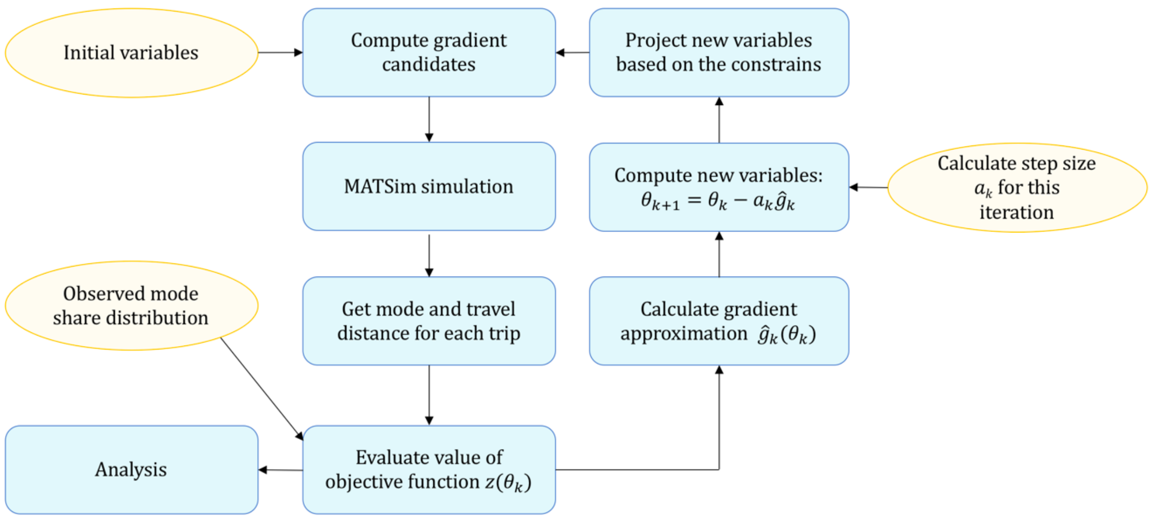
$$z(\hat{\theta}_k) = \max_{m \in M, d \in D} \left| S_{m,d} - \tilde{S}_{m,d} \right| \qquad (13)$$

The comparison of these two objective function is introduced in Section 7.1.

6.3 Calibration procedure

As illustrated in Figure 9, the calibration process starts from the initial variables and the computing two gradient candidates. Then three MATSim simulations with three variable sets, including initial variables and two gradient candidates, run at the same time. Based on the simulation output, the objective values of three variable vectors can be calculated by comparison with census data. Those objective values then at the next stage are used to evaluate the gradient approximation. At the step of computing new variables, step size a_k needs to be determined and also GSM can be implemented. The new variables have to satisfy the constraints, then the loop of the calibration process is repeated until the objective value, and variables get stabilized.

Figure 9 Calibration process



6.4 Measures of performance

Once the simulation results are obtained, they could be compared with the observation measurements qualitatively and quantitatively, in order to measure the performance of the calibration. Different methods of calculating simulation errors are applied in this thesis in order to quantify and aggregate the calibration results from different aspects. The quantification methods include: (a) Normalized Root-Mean-Square Error (RMSN), (b) Root-Mean-Square Percent Error (RMSPE), (c) Mean Absolute Normalized Error (MANE) and (d) Mean Percentage Error (MPE).

The RMSN and RMSPE quantify the overall error of the simulation. The large errors are more heavily penalized by these measures than the small errors. RMSN is formulated as follows:

$$RMSN = \frac{\sqrt{N \sum_{m\in M,\, n\in[1,N]} (Y^S_{m,n} - Y^O_{m,n})^2}}{\sum_{m\in M,\, n\in[1,N]} Y^O_{m,n}} \tag{14}$$

where

$N =$ number of bins
$M =$ travel modes $\{car, pt, walk, bike\}$
$Y^o_{m,n} =$ observation, and
$Y^S_{m,n} =$ simulated value of mode m in bin n.

RMSPE is calculated as follows:

$$RMSPE = \sqrt{\frac{1}{N} \sum_{m\in M,\, n\in[1,N]} \left[\frac{Y^S_{m,n} - Y^O_{m,n}}{Y^O_{m,n}}\right]^2} \tag{15}$$

The MANE indicates the overall error of simulation as well, but it is not particularly sensitive to large errors. The MANE is expressed as follows:

$$MANE = \frac{1}{N} \sum_{m\in M,\, n\in[1,N]} \left[\frac{|Y^S_{m,n} - Y^O_{m,n}|}{Y^O_{m,n}}\right] \tag{16}$$

The MPE indicates the existence of systematic under- or overprediction of the simulation, which is calculated as follows:

$$MPE = \frac{1}{N} \sum_{m\in M,\, n\in[1,N]} \left[\frac{Y^S_{m,n} - Y^O_{m,n}}{Y^O_{m,n}}\right] \tag{17}$$

While the rest three indicators are limited to be positive, MPE can be both positive and negative. A negative value means an underprediction of the simulation, whereas the positive value indicates an overprediction.

7 Results

The performance of SPSA algorithm on calibrating MATSim is tested in different scenarios in this section. The small scenario of 0.1% population is used for the comparison between different objective functions, different step size rules and the sensitivity analysis of the important parameters in the algorithm. The robustness of the algorithm is proved with the large scenario of 10% population. The concrete parameters of the algorithm chosen for the case study are listed in the Appendix C Parameters of SPSA algorithm for case study.

7.1 Comparison between different calibration objective functions

As introduced in Section 6.2, the sum of all differences and the maximum difference between census data and simulation output are the two options of the objective function in this thesis. Table 2 illustrates the performance of two objective functions. Compared with the uncalibrated result, the results of both two options reach to a similar satisfactory level.

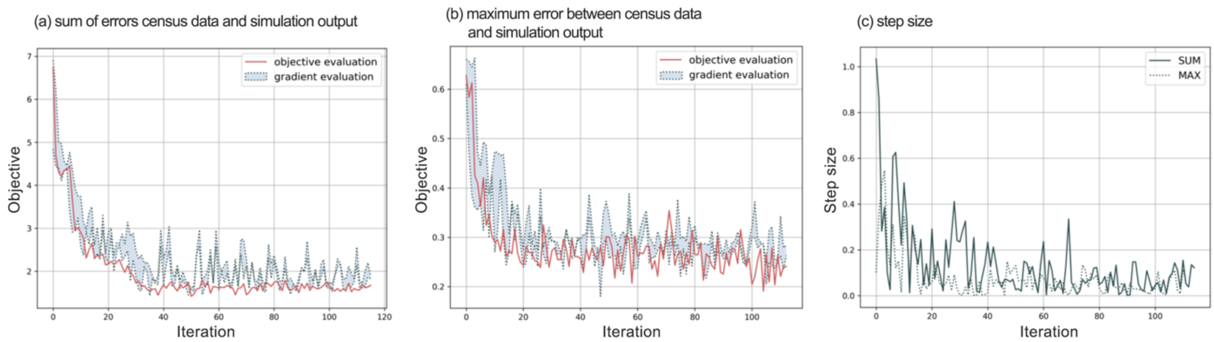
Table 2 Comparison between different calibration objective functions

	RMSN		**RMSPE**		**MPE**		**MANE**	
	value	change %	value	change %	value	change %	value	change %
Initial	0.463	-	6.039	-	-3.235	-	6.313	-
Sum	0.141	69.5	0.941	84.4	0.341	89.5	1.449	77.0
Max	0.153	67.0	1.022	83.1	0.064	98.0	1.440	77.2

However, the vibration of the objective function of maximum difference is much more severe than the sum of differences. The objective value still fluctuates and doesn't get converged after the 100th iteration as shown in Figure 10.

By using the sum objective function, the algorithm tries to find the gradient where general differences decrease. To the contrary, with the maximum objective function, the algorithm tries to minimize a specific mode at a certain distance until it is not the maximum difference anymore. Then the new maximum difference starts to be minimized. Ideally, when the maximum difference is minimized to a certain level, the general differences should also be limited within a small number. Especially the extremely large error should be avoided. However, in practice, the instability, as its major disadvantage, limits the effectiveness and accuracy of the algorithm.

Figure 10 Comparison between different calibration objective functions



7.2 Comparison between different step size rules

The basic step rule and GSM are introduced in Section 3.1 and Section3.2. As shown in Figure 11, the algorithm converges with fewer iterations with GSM. However, the additional simulations have to run in order to calculate the inner points of GSM, which results in a higher computational cost. Therefore, the efficiency of the algorithm is not improved significantly with GSM considering the running time or computational cost.

As the indicators of performance shown in Table 3, the results of calibration with two different step rules have no obvious differences.

Figure 11 Comparison between different step size rules

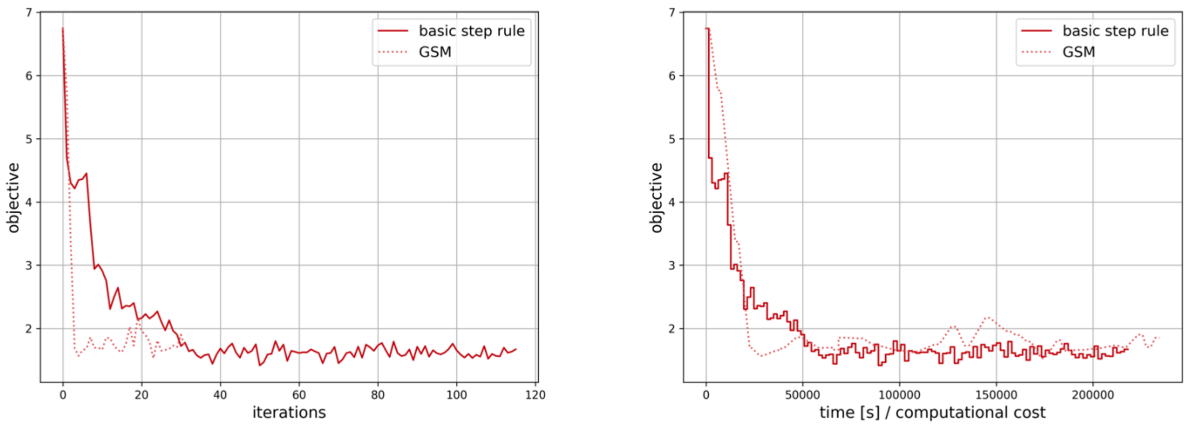


Table 3 Comparison between different step size rules

	RMSN		RMSPE		MPE		MANE	
	value	change %	value	change %	value	change %	value	change %
Initial	0.463	-	6.039	-	-3.235	-	6.313	-
Basic step	0.141	69.5	0.941	84.4	0.341	89.5	1.449	77.0
GSM	0.153	67.0	1.202	80.1	0.151	95.3	1.437	77.2

7.3 Sensitivity analysis

It is already proved by Spall (1992) [25] and Cipriani (2011) [10] that the parameters of SPSA are the determinants of algorithm efficiency. A good choice of the parameters helps SPSA to converge faster. Therefore, the further exploration of the parameters might be useful to improve the performance of the algorithm.

7.3.1 Parameter a

a is the essential parameter in Equation (1) and Equation (2), which determines the step size of algorithm directly. As clarified in Section 3.1, either too small or too large a has the negative impact on the algorithm efficiency. In the sensitivity analysis, a is set to be 0.2, 0.5, 0.7 respectively.

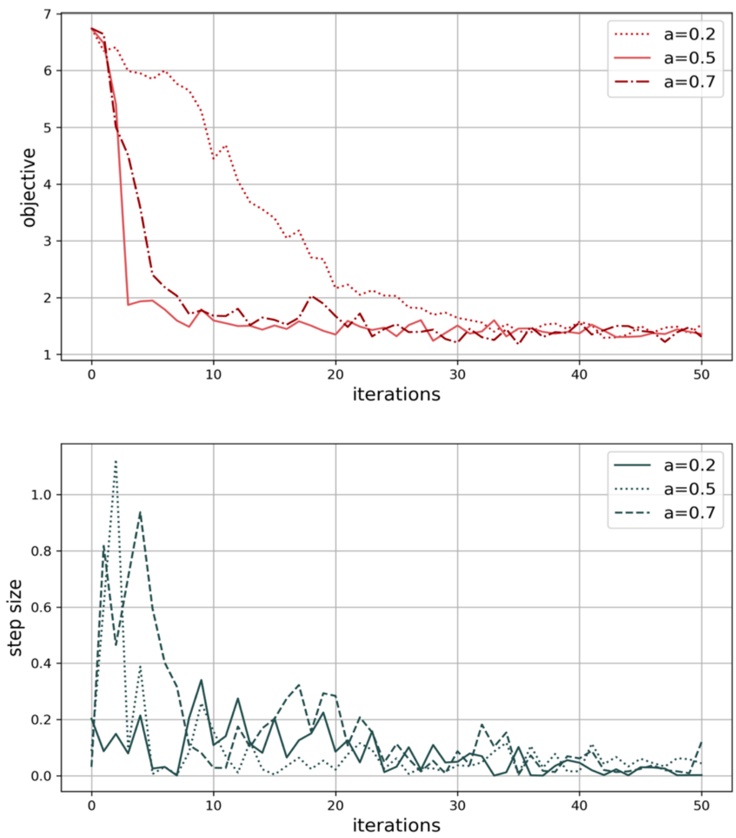
Table 4 Sensitivity analysis of parameter a

	RMSN		RMSPE		MPE		MANE	
	value	change %	value	change %	value	change %	value	change %
Initial	0.463	-	6.039	-	-3.235	-	6.313	-
$a = 0.2$	0.130	71.9	1.170	80.6	-0.135	95.8	1.367	78.3
$a = 0.5$	0.127	72.6	0.935	84.5	-0.030	99.1	1.228	80.5
$a = 0.7$	0.128	72.4	0.865	86.5	0.118	96.4	1.237	80.4

As shown in Figure 12, the algorithm with a of 0.2 converges much slower than 0.5 and 0.7. While it gets converged after 40 iterations, the algorithms with a of 0.5 and 0.7 are already stabilized after 30 iterations. Although compared with uncalibrated simulation, the results of $a = 0.2$ are acceptable (as shown in Table 4), there is still a gap compared with $a = 0.5$ or $a = 0.7$.

Obviously, the variables need more steps to move to the optimal point with smaller step size. That's why algorithm with smaller a is harder to get converged. However, the calibration with a of 0.7 is not the first one to get to the stable state. The objective value with a of 0.5 decreases faster than with a of 0.7. If a is too larger, the algorithm might take a large step far away from the optimal solution. Hence, it might get converged with more iterations.

Figure 12 Sensitivity analysis of parameter a



7.3.2 Parameter c

The value c plays a crucial role in SPSA algorithm since it defines the area, around which the objective function of current variables is evaluated to find a new gradient.

As illustrated in Figure 13, the algorithm converges faster with a higher value of c. However, due to the larger fluctuation the result of $c = 1.7$ is not as good as $c = 1.3$ or $c = 1.0$ (As shown in Table 5) In addition, the value of c has no direct correlation with step size according to the result of analysis, but has a positive correlation with the amplitude of gradient evaluation, which also affects the speed of convergence and the stability.

Figure 13 Sensitivity analysis of parameter c

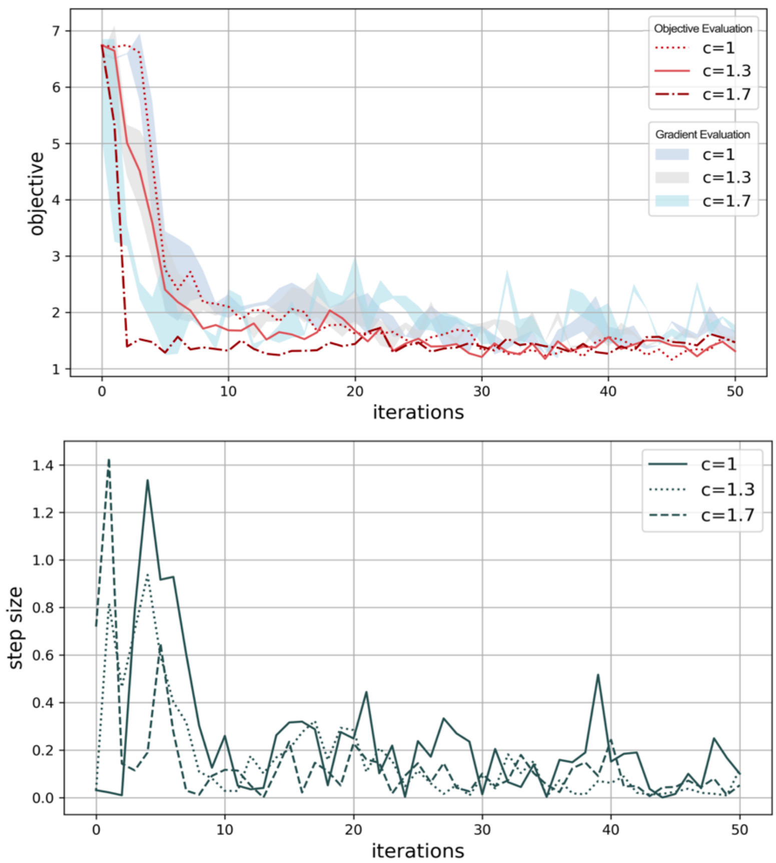


Table 5 Sensitivity analysis of parameter c

	RMSN		RMSPE		MPE		MANE	
	value	change %	value	change %	value	change %	value	change %
Initial	0.463	-	6.039	-	-3.235	-	6.313	-
c = 1.0	0.135	70.8	0.850	85.9	0.126	96.1	1.205	80.9
c = 1.3	0.128	72.4	0.865	85.7	0.118	96.4	1.237	80.4
c = 1.7	0.136	70.6	1.170	80.6	-0.136	95.8	1.367	78.3

7.4 Result of large scenario

The large scenario of 10% sample is used to test the robustness of the algorithm. Due to the considerable long running time, only 16 iterations have been finished in this thesis. But as shown in Figure 15, the algorithm tends to be converged after only six iterations. The indicators of performance (Table 6) also illustrates that the calibration result is already at a satisfactory level.

Figure 14 Calibration result of large scenario

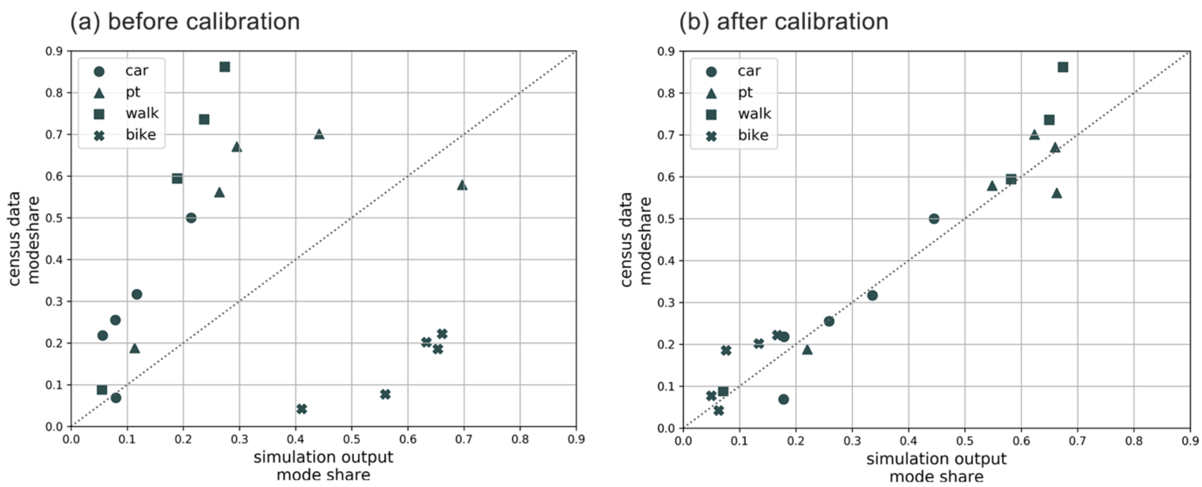
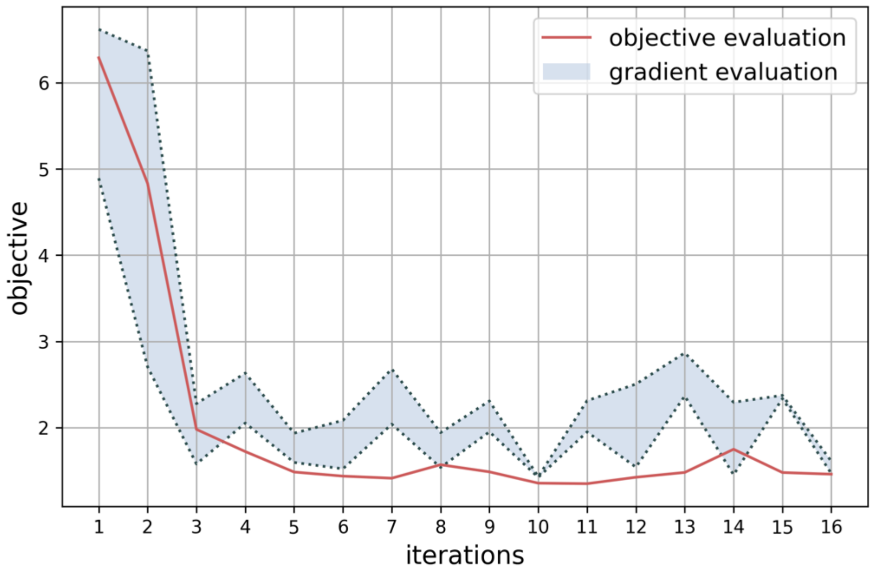


Figure 14 presents scatterplots of the same results Each point in these plots represents the mode share for one mode in one bin, with the census data plotted on the y-axis and the corresponding simulated value plotted on the x-axis. A diagonal line indicates the location of the points in the case of a perfect fit. Ideally, all points should lie on that line. As illustrated in Figure 14, the points are much closer to the line after calibration compared with the uncalibrated simulation.

Figure 15 Objective evaluation of large scenario



However, the gradient evaluation shown in Figure 15 still fluctuates heavily, though the objective evaluation seems to be stabilized. The calibration process indeed needs to run for more iterations. But at the same time, the parameter γ, which determines the reduction of c_k during iterations, might be chosen in a more suitable way.

The more visualizations of calibration result are attached in Appendix D Calibration result of large scenario. After calibration the differences of mode share distribution between census data and simulation output are reduced significantly.

Table 6 Calibration result of large scenario

	RMSN	RMSPE	MPE	MANE
Before	0.451	5.215	-2.893	5.924
After	0.119	0.862	0.070	1.072
Change %	73.6	83.5	97.6	81.9

8 Conclusion and Outlook

In this thesis, the SPSA algorithm has been implemented in MATSim for solving the calibration problem of an agent-based model. It is proved that SPSA is an efficient algorithm for calibrating multi-variables. However, the efficiency is largely depended on the choice of suitable parameters of SPSA algorithm. Therefore, testing and adjusting parameters are extremely crucial for solving the problem. The effects of different parameters on the algorithm are now more apparent by the sensitivity analysis. The better understanding of the meaning and effects of algorithm parameters is helpful to find the suitable parameters for the calibration problem. In addition to the parameter a and c, the sensitivity analysis of other parameters is also worth doing for the further step of this study.

The proposed combination of GSM and SPSA algorithm is also tested in this thesis. The algorithm indeed converges with much fewer iterations when GSM is applied. However, the computational cost of each iteration increases dramatically due to the considerable long running time of MATSim. In this case, the efficiency is not improved significantly by applying GSM in SPSA algorithm.

As the time limitation of the thesis, the number of calibration variables is only eight, which is still a long way from the ceiling of SPSA algorithm. SPSA should have the potential to calibrate over a hundred variables at the same time. Therefore, increasing the number of calibration variables is also the next step of this study.

In summary, SPSA algorithm is a suitable method for the calibration problem of an agent-based model due to its attractive efficiency and performance. So far, the study is just at the beginning stage. The further work, including the cross-validation of the result, increasing number of calibration variables needs to be done in the future. Additionally, some enhanced SPSA method, such as weighted SPSA, was developed, which also can be the method in the future to solve the more complicated calibration problem for MATSim.

Literature

[1] Antoniou, C., R. Balakrishna, Koutsopoulos, H. N. and M. A. Ben-Akiva (2011) Calibration methods for simulation-based dynamic traffic assignment systems, *International Journal of Modelling and Simulation*, **31**(3) 227-233.

[2] Antoniou, C., M. A. Ben-Akiva and H. N. Koutsopoulos (2007) Nonlinear Kalman filtering algorithms for on-line calibration of dynamic traffic assignment models, *IEEE Transactions on Intelligent Transportation Systems*, **8**(4), 661-670.

[3] Balakrishna, R., C. Antoniou, M. A. Ben-Akiva, H. N. Koutsopoulos and Y. Wen (2007) Calibration of microscopic traffic simulation models: Methods and application, *Transportation Research Record: Journal of the Transportation Research Board*, **1999**, 198-207.

[4] Balakrishna, R., M. A. Ben-Akiva and H. N. Koutsopoulos (2007) Offline calibration of dynamic traffic assignment: Simultaneous demand-and-supply estimation, *Transportation Research Record: Journal of the Transportation Research Board*, **2003**, 50-58.

[5] Balakrishna, R., H. N. Koutsopoulos and M. A. Ben-Akiva (2006) Simultaneous off-line demand and supply calibration of dynamic traffic assignment systems, Presentation at *the 85th Annual Meeting of The Transportation Research Board*, Washington, DC., January 2006.

[6] Balmer, M., K. Axhausen and K. Nagel (2006) Agent-based demand-modeling framework for large-scale microsimulations, *Transportation Research Record: Journal of the Transportation Research Board*, **1985**, 125-134.

[7] Ben-Akiva, M. A., S. Gao, Z. Wei, and Y. Wen (2012) A dynamic traffic assignment model for highly congested urban networks, *Transportation research part C: emerging technologies*, **24**, 62-82.

[8] Ben-Akiva, M. A., A. Davol, T. Toledo, H. N. Koutsopoulos, W. Burghout, I. Andréasson, ... and C. Lundin (2002) Calibration and evaluation of MITSIMLab in Stockholm, Presentation at *the 81st Transportation Research Board Meeting*, Washington, DC., January 2002.

[9] Beyeler, S. (2017) Practical Multi-Fidelity Bayesian Batch Optimization for Calibration of Dynamic Traffic Assignment Models, *Master thesis*, LAS, ETH Zürich.

[10] Cipriani, E., M. Florian, M. Mahut and M. Nigro (2011) A gradient approximation approach for adjusting temporal origin–destination matrices, *Transportation Research Part C: Emerging Technologies*, **19**(2), 270-282.

[11] Djukic, T., J. Van Lint and S. Hoogendoorn (2012) Application of principal component analysis to predict dynamic origin-destination matrices, *Transportation Research Record: Journal of the Transportation Research Board*, **2283**, 81-89.

[12] Flötteröd, G. (2009) Cadyts a free calibration tool for dynamic traffic simulations, Paper Presentation *9th Swiss Transport Research Conference*, September 2009.

[13]Flötteröd, G. (2017) A search acceleration method for optimization problems with transport simulation constraints, *Transportation Research Part B: Methodological*, **98**, 239-260.

[14]Flötteröd, G., M. Bierlaire and K. Nagel (2011) Bayesian demand calibration for dynamic traffic simulations, *Transportation Science*, **45** (4) 541-561.

[15]Flötteröd, G., Y. Chen and K. Nagel (2012) Behavioral calibration and analysis of a large-scale travel microsimulation, *Networks and Spatial Economics*, **12** (4) 481-502.

[16]Frederix, R., F. Viti, R. Corthout and C. Tampère (2011) New gradient approximation method for dynamic origin-destination matrix estimation on congested networks, *Transportation Research Record: Journal of the Transportation Research Board*, **2263**, 19-25.

[17]Hazelton, M. L. (2008) Statistical inference for time varying origin–destination matrices, *Transportation Research Part B: Methodological*, **42** (6) 542-552.

[18]Hägerstrand, T. (1970) What about people in Regional Science?, *Papers in Regional Science*, **24**(1) 6-21.

[19]Horni, A., K. Nagel and K. W. Axhausen (eds.), (2016). *The Multi-Agent Transport Simulation* MATSim, Ubiquity Press, London.

[20]Hollander, Y. and R. Liu (2008) The principles of calibrating traffic microsimulation models, *Transportation*, **35** (3) 347-362.

[21]Lee, J. B. and K. Ozbay (2009) New calibration methodology for microscopic traffic simulation using enhanced simultaneous perturbation stochastic approximation approach, *Transportation Research Record: Journal of the Transportation Research Board*, **2124**, 233-240.

[22]Lu, L., Xu, Y., C. Antoniou and M. A. Ben-Akiva (2015) An enhanced SPSA algorithm for the calibration of Dynamic Traffic Assignment models, *Transportation Research Part C: Emerging Technologies*, **51**, 149-166.

[23]Ma, J., H. Dong and H. Zhang (2007) Calibration of microsimulation with heuristic optimization methods, *Transportation Research Record: Journal of the Transportation Research Board*, **1999**, 208-217.

[24]Paz, A., V. Molano and C. Gaviria (2012) Calibration of corsim models considering all model parameters simultaneously, Paper Presentation *the 15th International IEEE Conference*, 1417-1422.

[25]Spall, J. C. (1992) Multivariate stochastic approximation using a simultaneous perturbation gradient approximation, *IEEE transactions on automatic control*, **37** (3) 332-341.

[26]Van der Zijpp, N. and C. Lindveld (2001) Estimation of origin-destination demand for dynamic assignment with simultaneous route and departure time choice, *Transportation Research Record: Journal of the Transportation Research Board*, **1771**, 75-82.

[27] Vaze, V., C. Antoniou, Y. Wen and M. A. Ben-Akiva (2009) Calibration of dynamic traffic assignment models with point-to-point traffic surveillance, *Transportation Research Record: Journal of the Transportation Research Board*, **2090**, 1-9.

Appendix A Literature review

First author	Year	Source	Purpose	Optimization Algorithm	Parameter	Objective	Assumption / Limitation / Comments	Simulation	Case study
Flötteröd G	2011	Flötteröd G, Bierlaire M, Nagel K. Bayesian demand calibration for dynamic traffic simulations[J]. Transportation Science, 2011, 45(4): 541-561.	To present a mathematically consistent and computationally efficient framework for the calibration of microsimulation-based travel demand models in the context of DTA	**Bayesian optimization** **Max. entropy** **Aggregate Path Flow Estimation** max. entropy	**Demand(Aggregate Path Flow Estimation)** din ithe number of trips on path i of OD pair n **Demand(Disaggregate Demand Calibration):** distribution Pn(i | x)	not guaranteed to be a global maximizer, however, because of a concave log-likelihood function, the above solution is the unique maximizer **max. entropy W(d|y)**	No assumptions about the deployed route choice and network loading model **Advantage**: transferability to a broad class of DTA microsimulations **Assumption:** 'proportional assignment', fixed route choice fraction in a interative fashion, $\Gamma = 0$	MATSim Activity-chain	Zurich
							Assumption1: $\pi(x)$ tight enough to allow for a linearization of Pn(i | x) around x0 = E{x|x~π(x)} **Assumption2**: macroscopic SUE model captures the averate conditions in the microsimulation and describe the deviations between the idealized and the real settings		
Flötteröd G	2012	Flötteröd G, Chen Y, Nagel K. Behavioral calibration and analysis of a large-scale travel microsimulation[J]. Networks and Spatial Economics, 2012, 12(4): 481-502.	To report on how ABDM can be implemented, calibrated and analyzed, using the metropolitan area of Zurich as an example.	**Bayesian optimization**	Utility function values		offsets are only used for choice and not for choice set generation, i.e., not for routing.	MATSim Activity-chain	Zurich
Paz A	2012	Paz A, Molano V, Gaviria C. Calibration of corsim models considering all model parameters simultaneously[C]//Intelligent Transportation Systems (ITSC), 2012 15th International IEEE Conference on. IEEE, 2012: 1417-1422.	To propse a calibration methodology for microscopic traffic flow simulation models that has the capability to simultaneously consider all model parameters and also to calibrate such time-dependent aspects of the model as link counts.	**SPSA**	all the parameters of the microscopic traffic flow simulation model	min root mean square (RMS)		CORSIM	
Spall J C	1992	Spall J C. Multivariate stochastic approximation using a simultaneous perturbation gradient approximation[J]. IEEE transactions on automatic control, 1992, 37(3): 332-341.	To present an SA algorithm that is based on a "simul- taneous perturbation'' gradient approximation instead of the standard finite difference approximation of Kiefer-Wolfowitz type procedures.	**SPSA**			Theory and numerical experience indicate that the algorithm presented here can be significantly more efficient than the standard finite difference-based algorithms in large-di- mensional problems.		

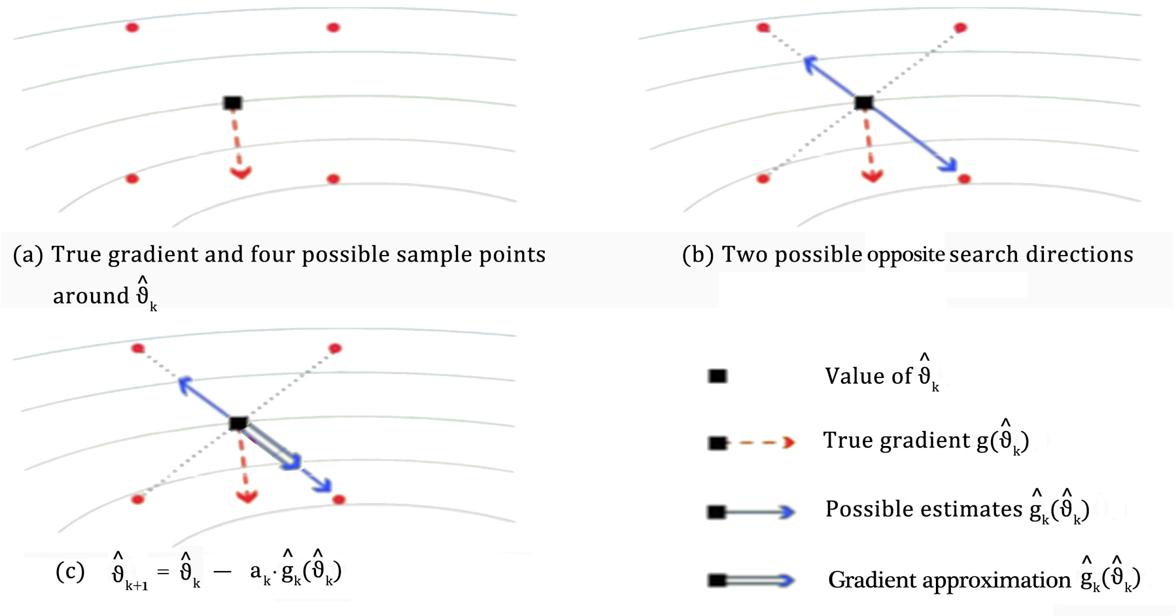
File name	Year	Source	Purpose	Optimization Algorithm	Parameter	Objective	Assumption / Limitation / Comments	Simulation	Case study
Balakrishna R	2007	Balakrishna R, Antoniou C, Ben-Akiva M, et al. Calibration of microscopic traffic simulation models: Methods and application[J]. Transportation Research Record: Journal of the Transportation Research Board, 2007 (1999): 198-207.	To present a methodology for simultaneously estimating all microscopic simulation model parameters by using general traffic measurements.	**SPSA** a seven-point Savirzky-Golay filter with degree 2 generated smoothed norms	time-dependent O-D matrices	Measures of Goodness of FitThe following statistics were used: • Normalized root-mean-square error (RMSN); • Root-mean-square percent error (RMSPE); • Mean percent error (MPE); • Theil's U-coefficient, with its bias, variance, and covariance components ; and • The GEH statistic, used in traffic engineering, traffic forecasting, and traffic modeling to compare sets of traffic volumes.	A appropriate solution algorithms must be able to: • Perform a global search by overcoming local hills and valleys, • Work without analytical derivatives (which would generally be unavailable), and • Converge in a reasonable time frame that does not grow rapidly with problem size.	MITSimLab OD pairs	New York
Lu L	2015	Lu L, Xu Y, Antoniou C, et al. An enhanced SPSA algorithm for the calibration of Dynamic Traffic Assignment models[J]. Transportation Research Part C: Emerging Technologies, 2015, 51: 149-166.	To present an enhances SPSA algorithm, called weighted SPSA, which incorporates the infromation of spatial and temporal correlation in a traffic network to limit the impact of nose and improve convergence and robustness. W-SPSA in this paper has been successfully applied to the calibration of DTA.	**SPSA with weight matrices** W-matrix can be calculated to incorporate both spatial and temporal correlations between each parameter and measurements in the DTA model	Demand on segments with sensors in every 15 min interval	Min. the difference between simulated value and realdata	**Advantage:** SPSA is a computationally efficient gradien approximation approach compared with FDSA. All parameters in the devision vector are perturbed at the same time. **Weighted** for a large-scale system has sparse correlations between parameters and measurements	DTA model DynaMIT OD pairs	Sigapore
Balakrishna R	2007	Balakrishna R, Ben-Akiva M, Koutsopoulos H. Offline calibration of dynamic traffic assignment: simultaneous demand-and-supply estimation[J]. Transportation Research Record: Journal of the Transportation Research Board, 2007 (2003): 50-58.	To present an offline DTA model calibration methodology, SPSA, for simultaneous estimation of all demand-and-supply inputs and parameters, with sensor data.	**SPSA**	O-D flows, capacities, route choice		explain SPSA Algorithm in details	DynaMIT	Los Angeles, California
Lee J B	2009	Lee J B, Ozbay K. New calibration methodology for microscopic traffic simulation using enhanced simultaneous perturbation stochastic approximation approach[J]. Transportation Research Record: Journal of the Transportation Research Board, 2009 (2124): 233-240.	To propose a new calibration methodology - the Bayesian ampling approach in conjunction with the application of enhanced SPSA - that enables the modeler to enhance calibration by considering statistical data distribution.	**Bayesian sampling approach** in conjunction with the **SPSA** approach	mean target headway and mean reaction time	Mean square variation (MSV) was used to evaluate the accuracy of the proposed E-SPS calibration approach.	Theis methodology enables the use of a wide range of traffic conditions from an observed data distribution. It also considers uncertainties.	PARAMICS	Freeway in Los Angeles, California

File name	Year	Source	Purpose	Optimization Algorithm	Parameter	Objective	Assumption / Limitation / Comments	Simulation	Case study
Ma J	2007	Ma J, Dong H, Zhang H. Calibration of microsimulation with heuristic optimization methods[J]. Transportation Research Record: Journal of the Transportation Research Board, 2007 (1999): 208-217.	In this study, another heuristics calibration algorithm is developed on the basis of the simultaneous perturbation stochastic approximation scheme and applied to calibration of several networks coded in Paramics.	**SPSA, GA, trial-and-error IA**	Glocal: Mean target headway, Mean reaction time, Driver aggressiveness, Driver awreness Local: Link headway factor, Link reaction factor, Ramp headway factor, Min ramp time, Ramp awareness distance, Sign-posting	the measurement of the extent between simulation results and real-world traffic measurements	The results indicate that the new heuristic algorithm can reach the same level of accuracy with considerably fewer iterations and computer processor time than other heuristic algorithms such as genetic algorithms and the trial-and-error iterative adjustment algorithm.	PARAMICS	Northern California
Vaze V	2009	Vaze V, Antoniou C, Wen Y, et al. Calibration of dynamic traffic assignment models with point-to-point traffic surveillance[J]. Transportation Research Record: Journal of the Transportation Research Board, 2009 (2090): 1-9.	To present a methodology for the joint calivration of demand and supply model parameters using travel time measurements obtained from these emerging traffic-sensing technologies. The calibration problem has been formulated as a stochastic optimization fromework.	**SPSA** **GA**	Demand: 482 O-D pairs times 8 intervals Supply: 5 parameters about speed-density relationship for 10 segment groups	minimizing the goodness- of-fit measure comparing the observed and fitted measurement values.	The results indicate that use of AVI data significantly improves calibration accuracy. For the small network, SPSA was found to be the most effec- tive algorithm, followed closely by GA. However, the GA ' s perfor-mance required additional computational effort because of a greater number of required function evaluations.	DynaMIT	New York
Antoniou C	2011	Antoniou C, Balakrishna R, Koutsopoulos H N, et al. Calibration methods for simulation-based dynamic traffic assignment systems[J]. International Journal of Modelling and Simulation, 2011, 31(3): 227-233.	To present a unified framework for off-line and on-line DTA calibration.	**off-line** (with a highly non-linear, noisy objective function, and capable of moving pastlocal optima towards a global solution): **SPSA** **on-line: LimEKF**	OD flows Capacities Traffic dynamics parameters Route choice parameters Travel times OD prediction parameters		successfully apply a DTA model to a real-world traffic network to replicagte the traffic fynamics accurately, a great number of parameters need to be calibtrated with local traffic surveillance data.	DTA model DynaMIT OD pair	Off-line: Los Angeles On-line: Southampton, UK
Antoniou C	2007	Antoniou C, Ben-Akiva M, Koutsopoulos H N. Nonlinear Kalman filtering algorithms for on-line calibration of dynamic traffic assignment models[J]. IEEE Transactions on Intelligent Transportation Systems, 2007, 8(4): 661-670.	To present an online calibration approach that jointly estimates demand and supply parameters of dynamic traffic assignment system. The approach is also empirically validated through and extensive application.	**Online** calibration problem is formulated as a **state-space** model. **Nolinear Kalman Filtering Algorithm(EKF, LimEKF, unscented Kalman filter)**	OD flows, speed-density relationship parameters, and segment capacities(can be easibly incorprate a different set of parameters)		On-line calibration deals with the dynamic adjustment of the off-line calibrated parameters utilizing real-time surveillance data a s they become available during the application of a DTA model LimEKF The LimEKF shows accuracy that is comparable to that of the best algorithm (EKF or UKF) but with vastly superior computational performance	DynaMIT DTA model OD pair	freeway network in Southampton
Ben-Akiva M	2002	Ben-Akiva M, Davol A, Toledo T, et al. Calibration and evaluation of MITSIMLab in Stockholm[C]//81st Transportation Research Board Meeting. 2002.	To describe results from a case stury to calibrate and evaluate the mifroscopic traffic simulation model MITSIMLab under congested traffic conditions.	Generalized least squeares (**GLS**) estimator	OD matrix and route choice model	traffic flow, point-to-point travel times, queue lengths	In the absence of detailed data, only aggregate data is used to calibrate the most critical parameters.	MITSIMLab OD pairs	Brunnsviken, Stockholm
Hazelton M L	2008	Hazelton M L. Statistical inference for time varying origin–destination matrices[J]. Transportation Research Part B: Methodological, 2008, 42(6): 542-552.	To describe the day-to-day evolution of OD matrices using a modest number of parameters.	**Bayesian inference** **Markov chain Monte Carlo methods**	time varying mean route flow vector		Views the OD estimation problem in a statistical perspective with a multivariate model for the link counts.	Statistical model	a section of the road network in Leicester

File name	Year	Source	Purpose	Optimization Algorithm	Parameter	Objective	Assumption / Limitation / Comments	Simulation	Case study
Frederix R	2011	Frederix R, Viti F, Corthout R, et al. New gradient approximation method for dynamic origin-destination matrix estimation on congested networks[J]. Transportation Research Record: Journal of the Transportation Research Board, 2011 (2263): 19-25.	To introduce the use of marginal computation (MaC) and use of kinematic wave theory principles to estimate the often non-linear relationships between OD flows and link flows under congested conditions.	**MaC algorithm Kinematic wave thery**	OD matrix		provide a practical O-D estimation methodol-ogy that could be applicable to **heavily congested** networks MaC is efficient techniques to calculate the sensitivity of the link flows to the O-D flows.	OD pairs	
Djukic T	2012	Djukic T, Van Lint J, Hoogendoorn S. Application of principal component analysis to predict dynamic origin-destination matrices[J]. Transportation Research Record: Journal of the Transportation Research Board, 2012 (2283): 81-89.	To explore the application of principal component anapysis (PCA) to reduce computational costs without a significant loss of accuracy.	**principal component analysis (PCA)**	dynamic O-D matrices		The results indicate three main patterns that ca be distinguished in dynamic O-D matrices: structural, structural deviation, and stochastic trend patterns.	OD pairs	
Ben-Akiva M E	2012	Ben-Akiva M E, Gao S, Wei Z, et al. A dynamic traffic assignment model for highly congested urban networks[J]. Transportation research part C: emerging technologies, 2012, 24: 62-82.	To present the synthesis of the solutions on both the demand and supply sides for the successful simulation and valibration in a complicated and highly congested traffic network	**Path-size Logit route choice model**	46832 time-dependent O-D flows, parameter for route choice, 19080 speed-density parameters	a weighted sum of dis- tances between time-dependent location-specific simulated measurements and field measurements (both counts and link travel times) and distances between calibrated variable values and their respective a priori values	The networks are characterized by large number of short links, complicated intersections and inter- changes, frequent on- and off-ramps and significant interferences from non-motorized traffic	DynaMIT-P	Beijing, China

Appendix B two-dimensional example of SPSA algorithm

Figure 16 two-dimensional example of SPSA



In the two-dimensional example, the optimization problem can be regarded as the problem of location choice of the variable. The variable moves on a plane, and tries to move to a point, where the objective value is minimum.

As illustrated in Figure 16, by calculating the objective value of two opposite sample points, the gradient approximation is calculated. The variable moves along the gradient by a suitable step.

Figure 17 two-dimensional calibration example

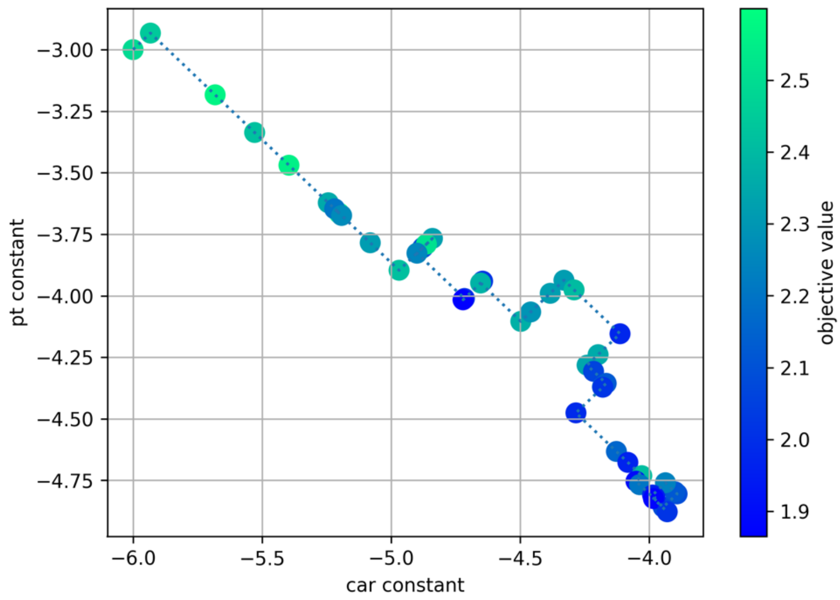


Figure 17 gives a concrete example of calibration with SPSA algorithm. The variables of pt constant and car constant move from the upper left of the plane to lower right long the gradient. The objective value is getting lower and lower long this direction.

Appendix C Parameters of SPSA algorithm for case study

	Step Rule	Objective function	a	α	c	γ
Sensitivity analysis of a	basic	sum	**0.2**	0.601	1.3	0.101
	basic	sum	**0.5**	0.601	1.3	0.101
	basic	sum	**0.7**	0.601	1.3	0.101
Sensitivity analysis of c	basic	sum	0.7	0.601	**1**	0.101
	basic	sum	0.7	0.601	**1.3**	0.101
	basic	sum	0.7	0.601	**1.7**	0.101
Objective function	basic	**sum**	1	0.601	2	0.101
	basic	**max**	4	0.601	2	0.101
Step rule	**basic**	sum	1	0.601	2	0.101
	GSM	sum	2	0.601	2	0.101
Large scenario	**basic**	**sum**	**0.5**	**0.601**	**1.3**	**0.101**

Appendix D Calibration result of large scenario

Figure 18   Comparison of mode share distribution between census data and simulation output of large scenario

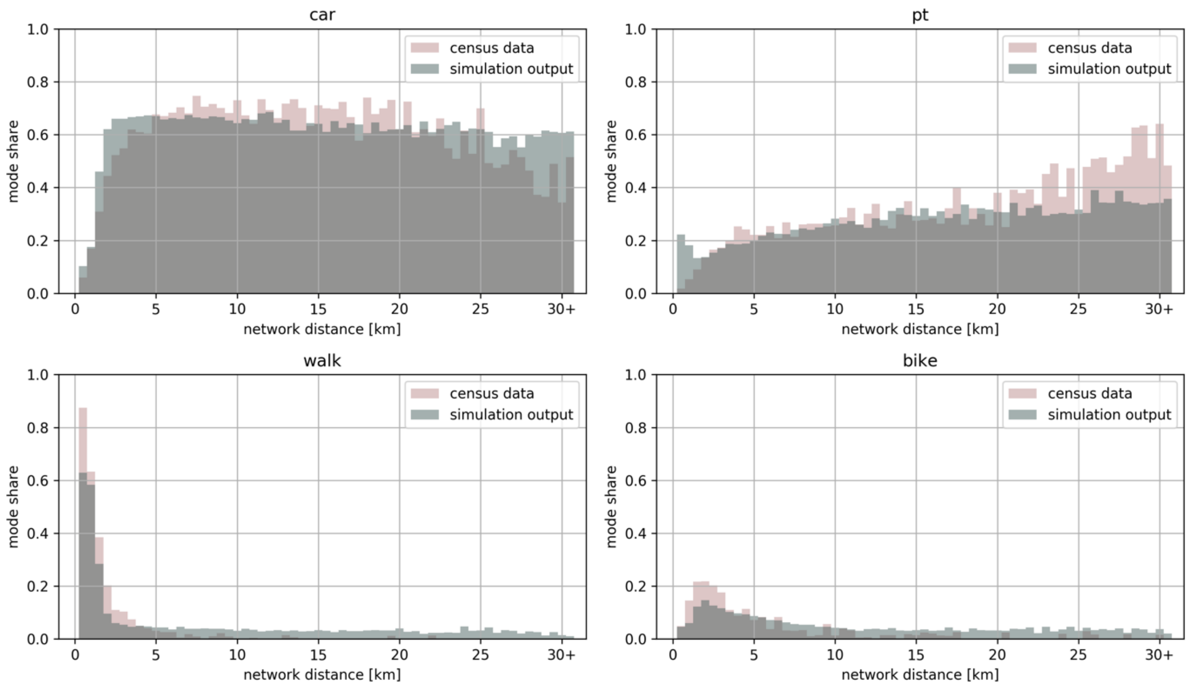


Figure 19 Comparison of mode share distribution with uneven bin size between census data and simulation output of large scenario before calibration

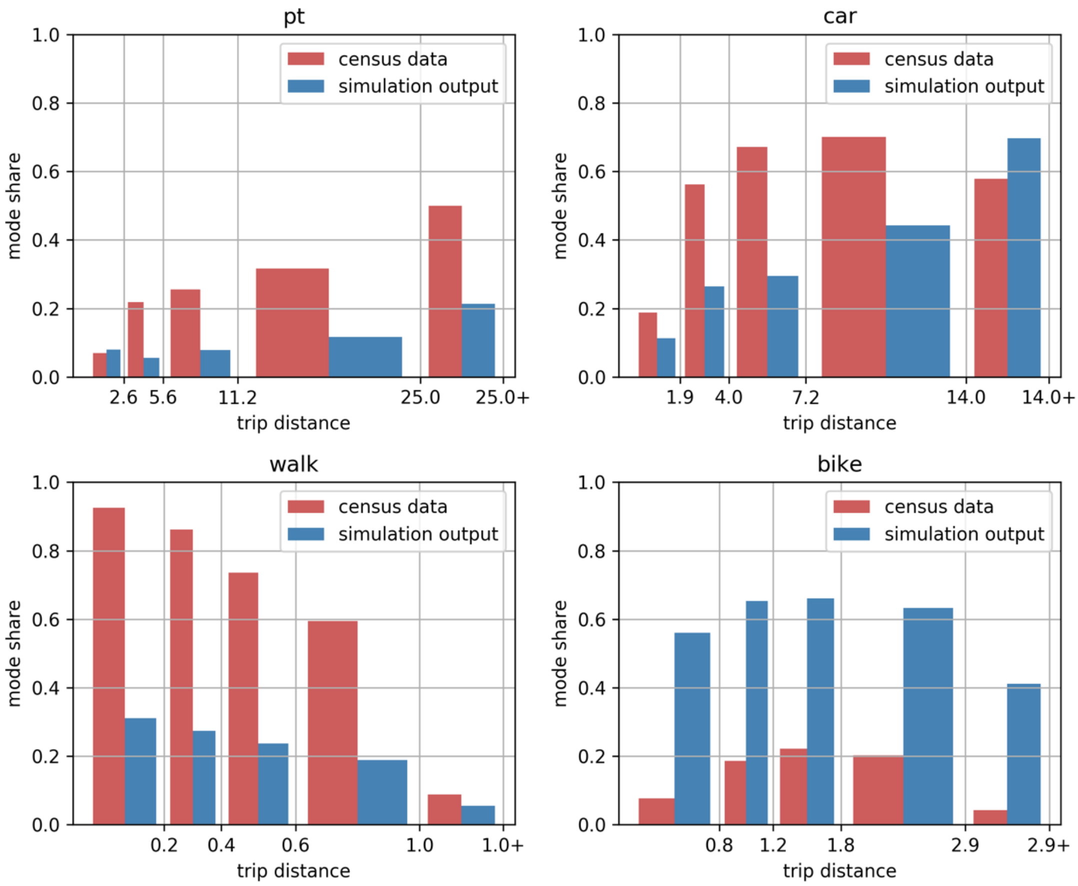


Figure 20 Comparison of mode share distribution with uneven bin size between census data and simulation output of large scenario after calibration