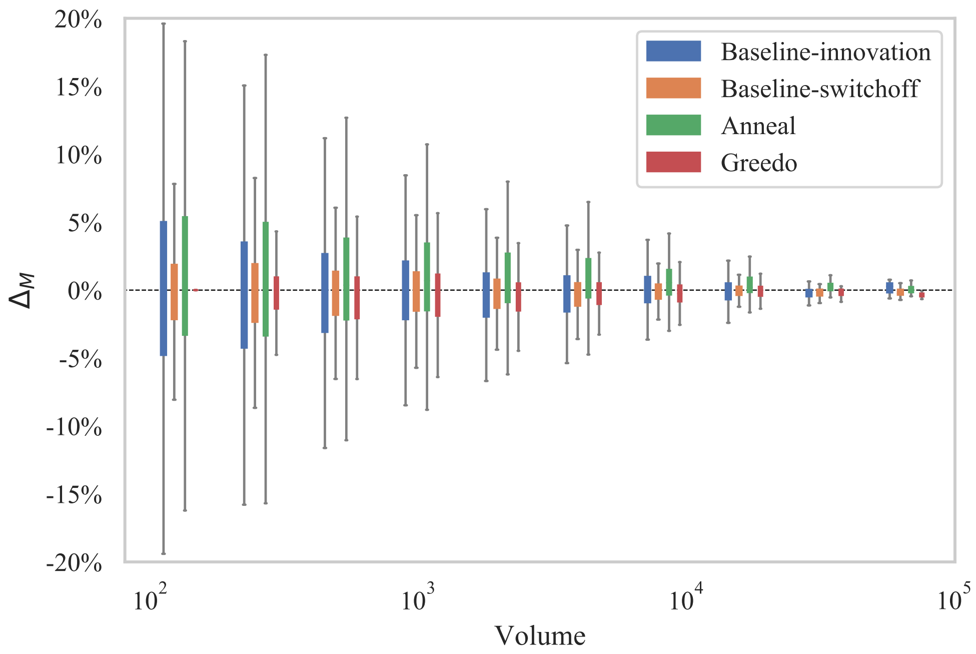


Convergence towards equilibrium in an agent-based transport simulation

Davi Guggisberg Bicudo

Master thesis **10XX**

Institute for Transport Planning and Systems **January 2020**

Institut für Verkehrsplanung und Transportsysteme
Institute for Transport Planning and Systems

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

ETH

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Eigenständigkeitserklärung

Die unterzeichnete Eigenständigkeitserklärung ist Bestandteil jeder während des Studiums verfassten Semester-, Bachelor- und Master-Arbeit oder anderen Abschlussarbeit (auch der jeweils elektronischen Version).

Die Dozentinnen und Dozenten können auch für andere bei ihnen verfasste schriftliche Arbeiten eine Eigenständigkeitserklärung verlangen.

Ich bestätige, die vorliegende Arbeit selbständig und in eigenen Worten verfasst zu haben. Davon ausgenommen sind sprachliche und inhaltliche Korrekturvorschläge durch die Betreuer und Betreuerinnen der Arbeit.

Titel der Arbeit (in Druckschrift):

CONVERGENCE TOWARDS EQUILIBRIUM IN AN AGENT-BASED TRANSPORT SIMULATION

Verfasst von (in Druckschrift):
Bei Gruppenarbeiten sind die Namen aller Verfasserinnen und Verfasser erforderlich.

Name(n):	**Vorname(n):**
GUGGISBERG BICUDO	DAVI

Ich bestätige mit meiner Unterschrift:
- Ich habe keine im Merkblatt „Zitier-Knigge" beschriebene Form des Plagiats begangen.
- Ich habe alle Methoden, Daten und Arbeitsabläufe wahrheitsgetreu dokumentiert.
- Ich habe keine Daten manipuliert.
- Ich habe alle Personen erwähnt, welche die Arbeit wesentlich unterstützt haben.

Ich nehme zur Kenntnis, dass die Arbeit mit elektronischen Hilfsmitteln auf Plagiate überprüft werden kann.

Ort, Datum	**Unterschrift(en)**
ZÜRICH, 13.12.2019	Bicudo

Bei Gruppenarbeiten sind die Namen aller Verfasserinnen und Verfasser erforderlich. Durch die Unterschriften bürgen sie gemeinsam für den gesamten Inhalt dieser schriftlichen Arbeit.

Acknowledgement

I learned so much during my master studies, concluding with this thesis, and could never thank enough for everyone's generous support.

First to Prof. Axhausen, who provides the great environment in the IVT and especially in the VPL group, where I could also work as a student assistant for two years under the guidance of Henrik Becker, whom I also thank. Thank you also to all the other professors and lecturers with whom I learned so much with the challenging work, and to my colleagues for sharing this experience.

I would like to thank the SBB for providing the model analyzed in this thesis and the generous computational resources for running it multiple times. This work was part of my internship there and I would like to acknowledge all my colleagues at the SBB who provided support and advice. Special thanks to Patrick Bützberger who leads the team, to Wolfgang Scherr for the challenging and very enriching topic and to Patrick Manser, my supervisor at the SBB who was very patient and supportive.

Thanks also to Gunnar Flötteröd with whom I could discuss the topic on many opportunities and who provided valuable explanations and ideas.

For this thesis, I would like to thank also Prof. Axhausen and Stefano Penazzi for the supervision and guidance from the IVT side.

Finally, I would like to thank my wife for her support in every possible way during this whole journey, to whom I also dedicate this work.

Master thesis

Convergence towards equilibrium in an agent-based transport simulation

Davi Guggisberg Bicudo
ETH Zürich, SBB
davi.guggisberg@sbb.ch

January 2020

Abstract

This thesis analyzes the convergence of a large-scale agent-based transport model based on the software MATSim. In MATSim, equilibrium is achieved iteratively by 1) the interaction of agents in a network simulation and 2) adjusting their daily mobility plans. Because such models implement computationally intensive iterative processes, equilibrium is often assumed, not explicitly verified. This work examines the existence and uniqueness of equilibrium in the MATSim model of the Swiss Federal Railways (SBB), as well as errors due to uncertainties. Four algorithms for customizing plans are tested and compared: a) fixed innovation rate, b) fixed innovation rate until global stability and then *switch-off* of innovation, c) incremental reduction of innovation rates and d) targeted innovation. The results of the experiments show that higher innovation rates lead to faster convergence, but also to a poorer equilibrium due to high variability, which is the case with option a). It was also identified that the number of iterations used so far is insufficient, with smaller volumes being more affected by early-termination errors. The analysis has also shown that the commonly used *switch-off* algorithm, option b), does not converge well nor leads to a unique equilibrium state. Based on this analysis, a recommendation is given for the alternative algorithms, options c) and d), as they present better equilibrium properties and thereby improve the interpretability of the model results. However, further development is needed as they also lead to longer computational times.

Keywords

MATSim, Convergence, Equilibrium, Uncertainty analysis, SIMBA MOBi, SBB

Preferred citation style

Guggisberg Bicudo, D. (2020) Convergence towards equilibrium in an agent-based transport simulation, *Master thesis*, Institute for Transport Planning and Systems, ETH Zurich, Zurich.

Masterarbeit

Convergence towards equilibrium in an agent-based transport simulation

Davi Guggisberg Bicudo
ETH Zürich, SBB
davi.guggisberg@sbb.ch

Januar 2020

Zusammenfassung

Diese Arbeit analysiert die Konvergenz eines agentenbasierten Verkehrsmodells, basierend auf der Software MATSim. In MATSim wird das Gleichgewicht iterativ erreicht, indem 1) Agenten in einer Netzsimulation interagieren und 2) ihre täglichen Mobilitätspläne anpassen. Weil solche Modelle rechenintensive iterative Prozesse verwenden, wird das Gleichgewicht häufig angenommen, aber nicht explizit verifiziert. Diese Arbeit untersucht die Existenz und die Eindeutigkeit des Gleichgewichts im MATSim-Modell der SBB, sowie Fehler aufgrund von Unsicherheiten. Es werden vier Algorithmen zur Anpassung von Plänen getestet und verglichen: a) fixe Innovationsrate, b) fixe Innovationsrate bis zur globalen Stabilität und dann *switch-off* der Innovation, c) schrittweise Reduktion der Innovationsraten und d) gezielte Innovation. Die Ergebnisse der Experimente zeigen, dass höhere Innovationsraten zu einer schnelleren Konvergenz, aber auch zu einem schlechteren Gleichgewicht aufgrund hoher Variabilität führen, was bei Option a) der Fall ist. Es wurde auch festgestellt, dass die bisher verwendete Anzahl Iterationen ungenügend sind. Die Analyse hat auch gezeigt, dass der häufig verwendete *switch-off*-Algorithmus, Option b), nicht gut konvergiert und keinen eindeutigen Gleichgewichtszustand garantiert. Eine Empfehlung für die alternativen Algorithmen, Optionen c) und d), wird gegeben, da sie bessere Gleichgewichtseigenschaften aufweisen und die Interpretierbarkeit verbessern. Weitere Entwicklung ist jedoch erforderlich, da sie zu längeren Rechenzeiten führen.

Schlüsselwörter

MATSim, Konvergenz, Gleichgewicht, Unsicherheitsanalyse, SIMBA MOBi, SBB

Bevorzugter Zitierstil

Guggisberg Bicudo, D. (2020) Convergence towards equilibrium in an agent-based transport simulation, *Masterarbeit*, Institut für Verkehrsplanung und Transportsysteme, ETH Zürich, Zürich.

Contents

List of Figures

List of Tables

1 Convergence and equilibrium in agent-based transport models

Transport can be intuitively modelled as a market. The *demand* is formed by individuals who wish to travel and the *supply* by the transport services such as public transport (PT) and the road network. This market is in equilibrium when the supply matches the demand and is characterized by the competitive price, in the case of transportation sometimes represented by the *generalized cost of travel*. Changes in the supply or demand lead to different equilibrium states.

Forecasting alternative scenarios for the transport market is the main task of transport models. These scenarios may consist of different demand patterns, such as population growth, changing preferences for individual mobility versus public transport or a higher willingness-to-pay. Policy evaluation concerns alternative supply scenarios, where possibilities may include the construction of a new train station, speed reduction at highways or the introduction of a fleet of autonomous taxis. Evaluating these scenarios require an appropriate model of the transport market.

Traditional transport models are of aggregate nature, i.e. they model traffic in terms of trip flows, or streams of travellers. Such models have a well-defined analytical formulation and the results are calculated in a deterministic way (Sheffi, 1985). Future challenges in transport planning, such as autonomous mobility and increasingly diverse mobility patterns, require more granularity. Agent-based models permit a wider range of possibilities but also comes at the costs of more computational time and the impossibility of an analytical formulation. Instead of continuous flows, a rule-based approach is employed, where the individual behavior and the rules of interaction between agents and the environment are modelled (Nagel and Flötteröd, 2012). The results of these models are stochastic, probabilistic in nature, which means that each time the model is calculated a slightly different result is expected.

This change from deterministic to probabilistic results requires a shift in paradigm in transport planning practice. Evaluating a scenario no longer provides a single value, but a *distribution*, which can, in turn, be represented by its mean and confidence interval.

It is the goal of this thesis to analyze these distributions and quantify the related uncertainty. For this purpose, the agent-based transport model developed at the Swiss Federal Railways (SBB), shall be investigated. The model is called SIMBA MOBi, and it consists of a large-scale model for all of Switzerland (Scherr et al., 2018, 2019b), modelling the daily travel pattern of approximately 8.5 million persons. It is one of the first models of its kind to be developed for corporate purposes, although other similar models have been developed before for research

(Horni et al., 2016).

The following sections of this chapter provide an introduction to the modelling framework as well as the main concepts used in this thesis. Chapter 2 expands the description of the modelling framework, providing details on MATSim's algorithm. Chapter 3 describes the theoretical basis for the analysis, the statistical techniques used and the experimental setup. The results of the experiments are analyzed and discussed in Chapter 4. At last, in Chapters 5 and 6, a broader interpretation of the results and their consequences are given, and a conclusion with recommendations for practice and future research is offered.

1.1 SIMBA MOBi

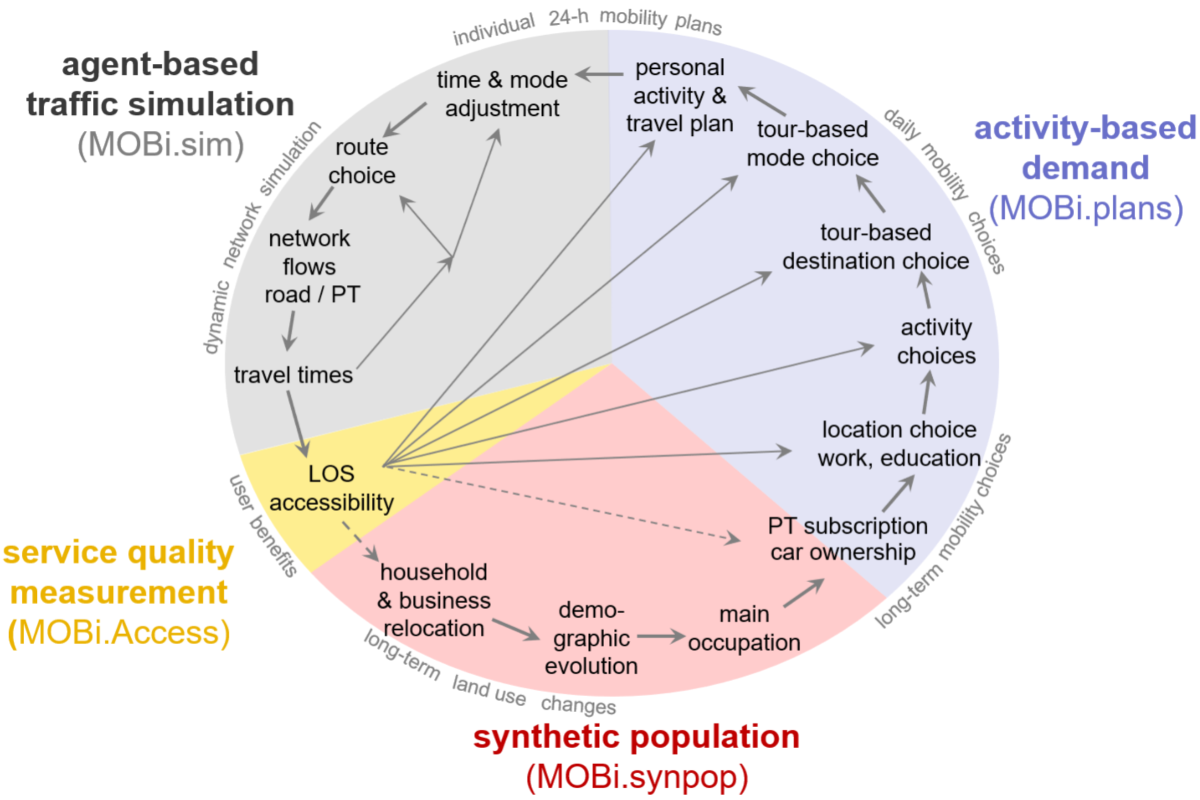
The different components, or sub-models, of SIMBA MOBi are displayed in Fig. 1 and briefly introduced as follows. In the first sub-model, *MOBi.synpop*, a synthetic population of agents representing the entire population of Switzerland, is created. Each agent has an array of synthetic personal attributes such as age and main occupation, as well as its household location. The second sub-model, *MOBi.plans*, is responsible for the agents' long and short-term choices related to travel behavior, such as mobility tool ownership and work or study locations. At this stage agents receive a complete daily schedule, or *plan*, including activity locations and which modes they use to travel between them. Refer to Scherr et al. (2019a) for a full description of *MOBi.plans*. In the third stage, *MOBi.sim*, agents interact in a common iterative simulated environment. At this stage some of these choices are updated and re-evaluated, allowing a gradual relaxation towards a state where agents are *mutually satisfied* with their daily schedules. The traffic conditions obtained in MOBi.sim are then used to calculate levels of service at the *MOBi.access* stage, which also closes the modelling loop by feeding this information to MOBi.synpop.

This thesis focuses on the analysis of the equilibrium of SIMBA MOBi's third stage, MOBi.sim, and its underlying framework MATSim, introduced in the following section.

1.2 MATSim

MATSim (Multi-Agent Transport Simulation) is a framework for microscopic modelling of travel behavior where individual agents represent utility-maximizing travellers (Horni et al., 2016). Utility, also known in MATSim as *score*, is obtained by executing planned activities

Figure 1: Architecture of SIMBA MOBi



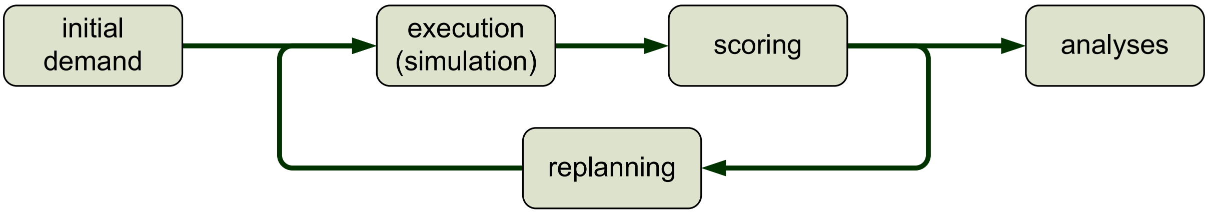
Source: Scherr et al. (2019b)

and lost by travelling. Travel is concurrently simulated for all agents in a common constrained environment - modelled as the PT service and the road network.

The agents (travel demand) affect the environment (travel supply) through congestion and are affected by it through longer travel times. An iterative process ensues, where agents concurrently adapt and re-evaluate their plans. Once agents no longer can improve their scores, this process is said to have reached the equilibrium. In this sense, equilibrium is not explicitly calculated, as in simpler market models, but expected to *emerge*, as further discussed in the next section.

MATSim's iterative process is summarized in Fig. 2. A run of this iterative process repeats the loop often for hundreds of iterations and is sometimes named simply a *simulation*, which is not to be confused with the mobility simulation that occurs every iteration in the execution step.

Figure 2: The MATSim loop



Source: Horni et al. (2016)

1.3 Equilibrium

Transport models replicate traffic conditions by reproducing demand and supply models that are mutually consistent, i.e. are in equilibrium. This consists of a *user equilibrium* (UE), or Nash equilibrium from the theory of non-cooperative games (Nash, 1951), where no player can individually improve their outcome by unilaterally changing their strategy. For transport models, according to Nagel and Flötteröd (2012), this type of equilibrium can be most intuitively formulated as a fixed point; in their words: "find a demand pattern that generates network conditions that in turn cause the same demand pattern to re-appear". From this definition, Nagel and Flötteröd (2012) gradually describe the conceptual development of the solutions to the traffic assignment problem, from UE to *stochastic user equilibrium* (SUE), and from the traditional static forms to the fully dynamic agent-based models.

This thesis analyzes the *convergence towards equilibrium* of an *agent-based SUE* model. Meister (2011) defines the agent-based SUE as "(...) a system state where the number of agents which perceive that they can improve their state is minimized, given a dynamic environment where a constant share of all agents change their plans".

Richiardi et al. (2006) propose that equilibrium in agent-based models can be analyzed at the *micro-level*, where, in the iterative process, agents no longer change their strategies, and at the *macro-level*, where the state of relevant aggregate statistics are stationary (or in *steady-state*). Meister (2011) proposed a method to measure micro-level equilibrium, where he observes the global share of agents that no longer can produce better solutions. This thesis focuses on the analysis of equilibrium at an aggregate, macro-level. In this sense, although equilibrium is a property of the choices of the agents, it shall be measured indirectly, by evaluating aggregate statistics throughout the iterative process, i.e. time-series of the statistics. Convergence is thus in this thesis defined as the approach towards equilibrium of the agent-based model, measured by the approach towards stationarity of its relevant statistics.

A good agent-based transport model is one that, under equilibrium, represents well the desired properties of the (real-life) equilibrium of transport markets. Equilibrium is hence a necessary but not sufficient condition, since the model may reach an equilibrium that does *not* represent the transport market well. In this sense, a good algorithm for such models is one that reaches a *good* equilibrium for the lowest possible computational cost.

1.4 Uncertainty analysis

Transport models are subject to a multitude of sources of uncertainty, intrinsic to their goal of modelling complex human behavior and interactions. Rasouli and Timmermans (2012) divide the sources of uncertainty into input uncertainty, concerning errors in input data such as sampling bias and survey design issues, and model uncertainty, which is further divided into estimation uncertainty and specification uncertainty. The former stems from errors in calibrating a model's parameters while the latter is related to the model's assumptions, equations, and algorithms (Petrik et al., 2018). This thesis focuses on analyzing MOBi.sim's specification uncertainty.

One particular type of uncertainty which has been extensively investigated in transport modelling literature is stochastic variability, usually measured by running the model several times with different sequences of random numbers and evaluating the variance between results. This type of analysis allows estimating the potential error caused by running the model a single time instead of multiple times, sometimes also evaluating the variance between iterations. These two analyses can be respectively named *between-run variability* and *within-run variability*. Horni et al. (2011); Paulsen et al. (2018) performed this type of analysis with MATSim, by respectively running thirty times a scenario for Zurich and a hundred times a scenario for Santiago. In their work, they also discuss the sources of randomness causing the output variability. This type of analysis may also be done for validation purposes, as in Chakirov and Fourie (2014); Hemdan et al. (2019), where the authors ran their analysis four times each to ensure their results were not excessively affected by random variability.

The transport modelling literature provides several examples of uncertainty analysis and Rasouli and Timmermans (2012) provide a comprehensive review. In their paper, most of the studies concerned models based on random utility theory, where discrete choice models are used to estimate choice dimensions such as mode, destination, trip frequency, activity types and time of the day. Simulation in this kind of model often relates to the Monte Carlo draws used to estimate choice distributions. MOBi.plans is an example from this category, in some cases named activity-based models of travel demand, when demand is modelled from the perspective of activity-based decisions. In some cases, this demand is then fed into a traffic assignment model to obtain network flows, as done by Gibb and Bowman (2007) who performed an uncertainty analysis of this type of setup.

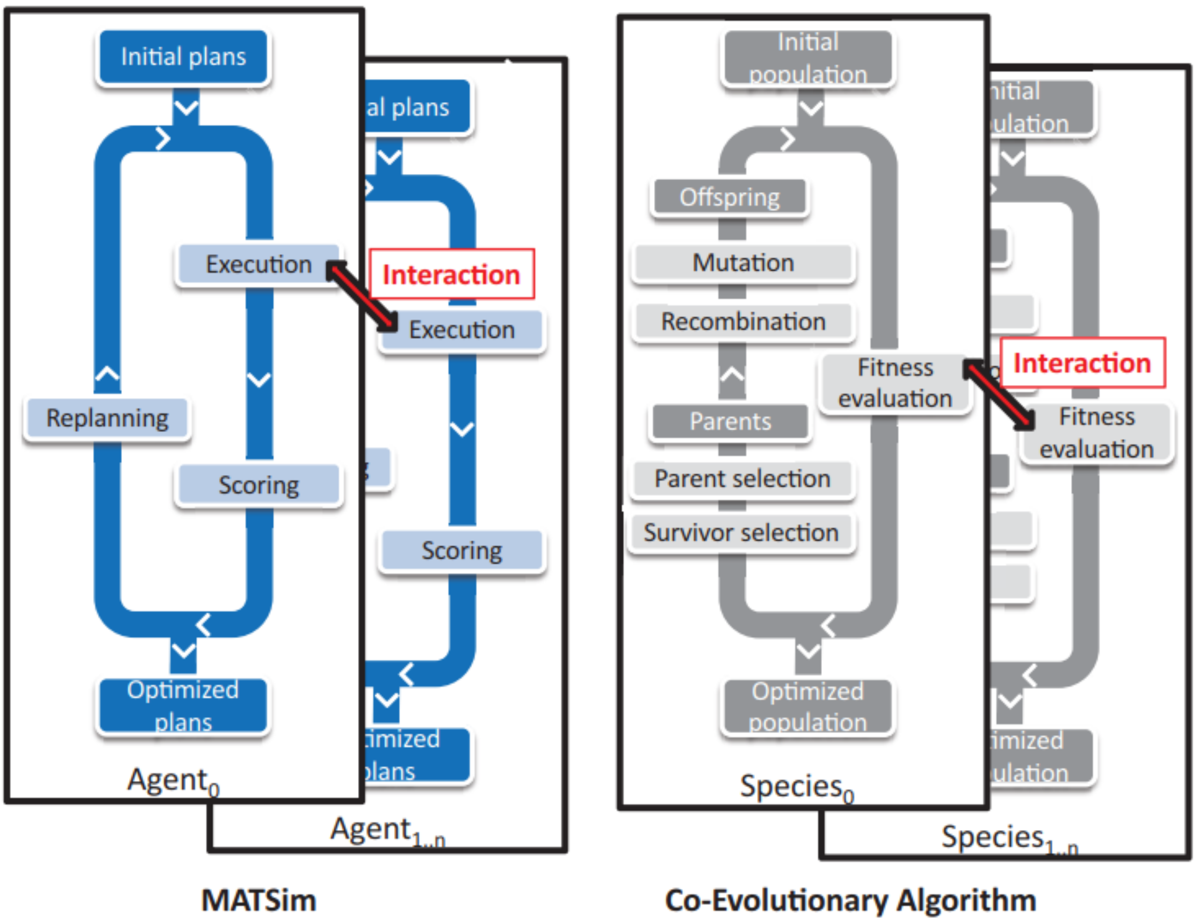
In the context of agent-based transport models, such as MATSim and TRANSIMS (Smith et al., 1995), simulation refers rather to the *physical* movement of agents in the virtual shared environment.

Uncertainty analysis commonly involves some kind sensitivity analysis (SA), performed by Zhuge et al. (2019) and Llorca and Moeckel (2019). Richiardi et al. (2006) suggest several types of SA that can be done to evaluate agent-based models, whereas ten Broeke (2017) describes different methodologies that can be used to perform SA for agent-based models. SA can, for example, be done by testing different algorithms to see whether they reach comparable equilibrium states, as done by Nagel et al. (2000) with different traffic assignment methods in TRANSIMS.

2 Innovation within MATSim

MATSim's learning process is closely related to that of co-evolutionary algorithms, as seen in Fig. 3. The parallels to natural evolution are particularly helpful for intuitively understanding the learning process. Within this framework, each agent represents a *species* and the plans they store in memory a *population*. These populations of plans share a common *physical environment*, the constrained network of PT services and roads. The mechanism that produces the *offspring*, that is, the plans for further iterations, is called Replanning in MATSim. The *fittest* plans survive the iterative *co-evolution* of the different *species* (agents).

Figure 3: MATSim's learning and the co-evolutionary algorithm



Source: Horni et al. (2016)

At the beginning of the Replanning phase, a fixed share of the agents is randomly selected for innovation, which means one of their plans, selected also randomly, will be copied and mutated. This newly produced plan is then executed (evaluated) in the next iteration. All other agents simply select one of the plans in their memory for the next iteration, for instance by picking the one with the best score. Once an agent's memory is full (controlled by a global parameter), one of the plans is discarded, usually the one with the worst score.

This process leads to a gradual improvement of the plans towards an SUE, and is summarized in Algorithm 1.

Algorithm 1 Co-evolutionary, population-based search. Adapted from Nagel and Flötteröd (2012).

1. **Initiation:** Generate at least one plan for every agent.
2. **Iterations:** Repeat the following many times.
   a) **Plan selection:** Select one plan for every agent.
   b) **Plan scoring:** Obtain a score for every agent's selected plan by executing all selected plans simultaneously in a mobility simulator and attach a performance measure (score) to each executed plan.
   c) **Plan innovation:** Generate new plans for some of the agents.

Algorithm 1 does not, however, specify a stopping criterion. The simplest way to evaluate whether the simulation has reached a stable state is by visually checking the development of some global statistics. This can be achieved by looking at some plots MATSim automatically generates and updates every iteration, showing the development of average scores, average travel distances, and mode shares. Once the slope of the curve is small enough, the simulation can be said to have converged (according to this visual criterion).

2.1 Plan innovation

The way new plans are generated is controlled by the so-called innovation strategies, or modules, which can either mutate the plan in a random or in a best-response manner. The former will randomly change plans in some aspect while the latter will try to calculate the best option for a particular choice dimension.

Commonly used innovation modules in MATSim are:

- *ReRoute* (best-response): re-calculates the shortest paths between each activity pair in a plan, based on the travel times from the previous iteration.
- *SubtourModeChoice* (random innovation): searches a plan for sub-tours, defined as a sequence of trips starting and ending at the same activity such as home-work-shop-home or work-leisure-work, and randomly assigns different modes of travel.

- *TimeAllocationMutator* (random innovation): randomly changes activities' starting times and durations, within a specified range.

TimeAllocationMutator is often combined with ReRoute as a strategy that innovates time while ensuring reasonable routes at the new times, which is particularly important in case of PT routes in order not to miss the scheduled services.

Each module has a configured replanning rate, which indicates the share of agents that shall be selected for innovation for a given module. The sum of the replanning rates of all modules can be named global innovation rate, and commonly ranges between 10-30%. The configuration for the model studied in this thesis is presented in Section 3.4.1.

2.2 Plan choice

As mentioned before, agents that are not selected for replanning simply choose an existing plan, from their experience (their memory), for the next iteration. For this, there are also different possibilities, implemented in the so-called choice modules.

The simplest one is to choose always the best plan in memory (*BestScore* strategy). Since this option has a high risk of leading to a local minimum, it is often combined with the random plan choice strategy (*SelectRandom*), where a share of the agents will select a plan at random. A more elaborate way of doing this is by making agents choose plans in a probabilistic manner. In MATSim, this is the *SelectExpBeta* strategy and follows the logit model,

$$P_i = \frac{e^{\beta S_i}}{\sum_j e^{\beta S_j}} \tag{1}$$

where S_i stands for the score of plan i and β models the agent's sensitivity to higher scores.

A modified version of Eq. (1) is implemented in the *ChangeExpBeta* strategy, where instead of choosing according to the plan probabilities, agents have to decide whether or not to switch to a randomly selected alternative plan. This switching process *converges* to the probabilities of Eq. (1), as explained in Nagel and Flötteröd (2012). *ChangeExpBeta* is the choice strategy used in SIMBA MOBi.

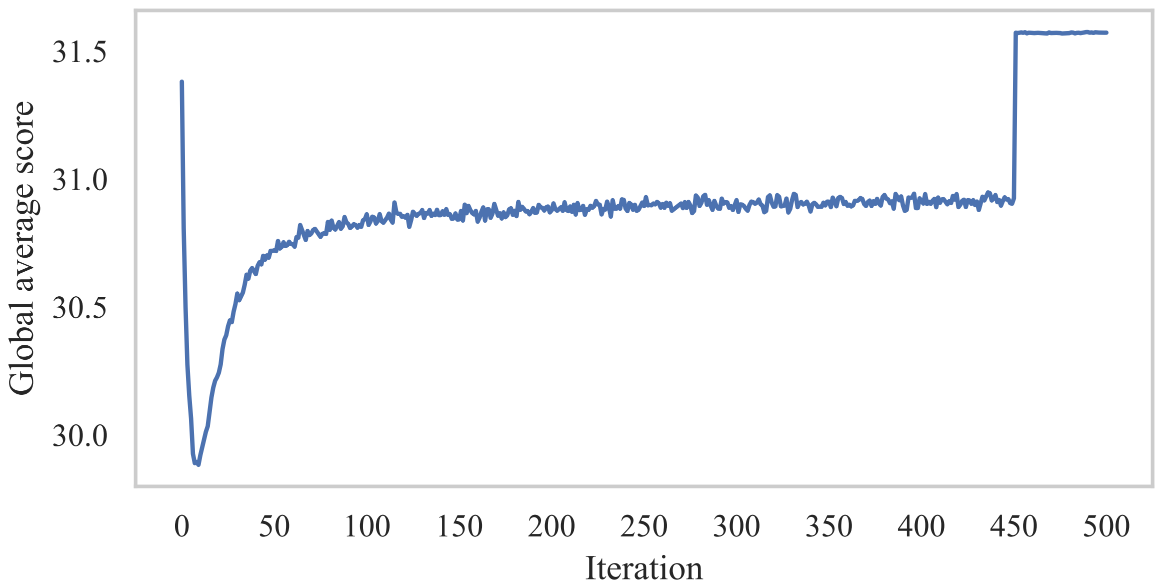
2.3 Global convergence and MATSim's behavioral model

From a discrete choice theory perspective, the plan innovation process can be seen as a sampling mechanism (Flötteröd and Kickhöfer, 2016), allowing agents to gradually update their choice sets. New plans are always evaluated, and since random innovation often generates bad solutions, this process introduces undesired randomness into the system.

If the global innovation rate is small enough, the effect may be neglected. Another way of dealing with this excess randomness is by eventually switching innovation off after the system becomes stable (i.e. skip step **c)** in Algorithm 1). The iterative process continues for some additional iterations for relaxation of the plan choices. It is inevitable, however, that the plans contained in the agents' memories were *learned* under quite different conditions, i.e. where a substantial share of the agents behaved randomly.

This approach generates a disturbance in the system, typically characterized by a *jump* in the global statistics, as seen in Fig. 4 at iteration 450. The system's development is divided into two phases, before and after stopping innovation. The first phase may be called *innovation phase* and the second *choice-only phase*.

Figure 4: MATSim's typical global average score curve (global innovation rate of 20%)



From the definition of equilibrium as a fix-point provided in Section 1.3, the results of future iterations must depend solely on the results of recent past iterations. From the perspective of the agents, it means that the environment is predictable enough so that they can decide on what to do next based on what they experienced in the recent past. The jump seen in Fig. 4 breaks this definition, since the two states are quite different.

In the first phase there is a large share of agents taking random decisions, often far from optimal (e.g. walking several kilometers instead of taking the train or driving). This generates unrealistic network conditions, causing agents to be learning sub-optimal solutions. Agents then start a new phase where everyone behaves more rationally (no one is testing random decisions), which they have never experienced before. In this new phase agents are no longer allowed to try new solutions, hence the flatness of the score curve. The reason why the jump in Fig. 4 is positive is that the agents which were getting low scores due to their execution of bad plans no longer are forced to do this and immediately switch to their better plans in memory, thus increasing the global average.

2.4 Alternative innovation algorithms

This section introduces two alternative innovation algorithms for MATSim, which shall also be analyzed for comparison with the traditional option described in the previous sections.

2.4.1 Simple annealing

The simplest way to achieve lower randomness and avoid the arbitrary shut-off point is to adopt a progressive reduction of the innovation rate. This method has been previously proposed in the context of agent-based transport models and applied by Chakirov (2016); Rickert (1998).

This reduction schedule may be arbitrarily defined or follow a function of the iteration number, e.g. linear or exponential decay. The method of successive averages (MSA) has been applied in many contexts in transport modelling and offers an interesting possibility since it converges *almost surely* to a minimum (Sheffi, 1985). Letting R_I be the innovation rate at iteration I, the MSA can be defined as

$$R_I = I^{-\gamma} \tag{2}$$

where the factor $\gamma = [0.5, 1.0]$ controls the rate of decay.

2.4.2 Greedo - targeted innovation

Greedo is a recently developed optimization-based replanning module for MATSim (Flötteröd, 2019). It works by running an inner loop in the replanning phase, as in (Fourie et al., 2013), where agents iteratively test multiple alternatives without actually evaluating their new plan in the execution phase. This constitutes the search for a *better-response*, in contrast to random innovation or best-response. Greedo then chooses which agents are allowed to execute their new plans in the mobility simulation. This decision is based on the expected utility gain and the amount of disturbance the agent is expected to cause in the network. Greedo's optimization function is constituted of these two terms, global utility maximization and minimization of network variability. The intuition is that by keeping network variability as low as possible, agents have a more predictable environment and are thus able to improve faster and reach a higher score.

In Greedo, agents have only one plan in memory, which removes the need for any plan choice module. With this change, MATSim's iterative process becomes an agent-based UE instead of SUE. The final outputs may still be stochastic due to other sources of randomness, particularly in the mobility simulation, but the agents' learning process leads to deterministic results since at equilibrium switching between plans no longer occurs (Flötteröd, 2016).

3 Analysis of convergence

This chapter provides the theoretical background for the analysis performed in this thesis. The first section frames the problem of analyzing the equilibrium properties of an agent-based model; Section 3.2 defines what are the measures of interest; then Section 3.3 gives details on the actual methods applied; and finally, Section 3.4 explains the experimental setup.

3.1 Equilibrium analysis

The emergent properties of an agent-based model, are defined by its equilibrium state and can be represented by its set of relevant statistics. Letting Y_i be one such statistic at iteration i, Richiardi et al. (2006) have formally specified its equilibrium state Y_e as:

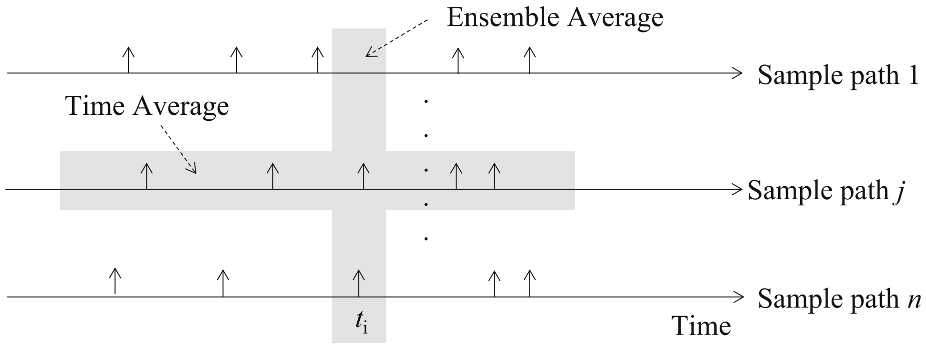
$$Y_e = \lim_{i \to \infty} Y_i \equiv g(x_1, ..., x_n; \alpha_i, ..., \alpha_n) \tag{3}$$

where x_n is agent n's state and α_n its parameters. g is the function to calculate the statistic and of intractable algebraic form since it is a transformation of the function defining the agents' mutual interactions. The agent-based model used in this thesis is stochastic, so x_n becomes random variables, and consequently also Y. Note that in this formulation x and Y_e are independent of the iteration i, meaning the system becomes *stationary*, or in *steady-state*, as $i \to \infty$. They are only conditional on g and the parameters α, not on the agents' state at the beginning of the simulation (or at any other finite point).

The consequence is that every time the model is run, the same results are expected, a property called *ergodicity*. If, in practice, it is verified that Y_e is dependent on the initial state, usually defined in computer simulations by the random seed, the model is said to be non-ergodic. If, for example, the statistic of interest Y_e would be the mean, then a stochastic process could be said to be mean-ergodic if the average over many iterations approaches the ensemble average (Park, 2018). Stationarity is thus a necessary condition for ergodicity. Fig. 5 illustrates mean-ergodicity, where each sample path in the present case is a simulation run. This thesis does not analyze ergodicity of higher statistical moments.

If a process is stationary and ergodic, Y_e can be consistently estimated by a single sufficiently long array of Y_i realizations. This means that, once the agent-based simulation has reached stationarity with respect to the statistic Y and ergodicity has been verified, the population parameter can be consistently estimated by the sample statistic (Hayashi, 2000). If, on the other hand, the model is

Figure 5: Illustration of the concept of the mean-ergodicity



Adapted from Park (2018)

non-ergodic, multiple simulation runs are required for estimating Y_e (Grazzini, 2012).

3.2 Aggregation of discrete random variables

Choices in transport behavior are usually discrete and the number of options finite. Some dimensions have a small set of possibilities, such as mode choice, while others a very large such as route choice, but both are finite. Time choices are an exception and can be continuous, and although in practice individuals do not optimize their schedules until the very last second, this is left open to MATSim agents.

The interest of agent-based models is not however on the choices of individual agents but on the *emergent* properties of their interactions. In other words, the interest lies in the individuals' *aggregate* behavior. Aggregates can be set up to answer different questions, such as how many persons take PT to work, or how many agents travel by car through a certain highway section, or what is the average age of agents travelling during the night. Agent-based models' microsimulation nature allows a large variety of possible aggregate analysis. This thesis will focus its analysis on some of the most obvious ones:

- **global aggregates**
  - average travel distance
  - average agent scores
  - modal split (share of trips with each of the available modes)

- **local aggregates**
  - daily PT stop volumes (sum of boardings and alightings)
  - daily link volumes

Scores and travel distances are attributes that every single agent has, and the related global aggregates are a simple average over all agents. The other aggregates, including modal split, refer to parts of the system and have a zero-sum nature. Part of the reason is that agents in MATSim cannot drop or add new activities, meaning the sum of all trips is always the same. Even if they could, the day still has a maximum of 24 hours to be split between activities and travel. This makes these aggregates more strongly correlated, since an increase of trips in an area or with a particular mode, mandatorily means a decrease somewhere else.

To calculate these aggregates, consider the set of individuals $a_i = a_1, \ldots, a_n$, the set of options in their choice sets as $o_j = o_1, \ldots, o_m$ and x_{ij} as the binary choice of agent a_i for option o_j. The choice probability of x_{ij} is defined by the Bernoulli random variable X_{ij}, which equals 1 if the given option is chosen and 0 otherwise, and its expected value $E(X_{ij}) = p_{ij}$. Matrix 4 presents the array of choice probabilities p_{ij} of all individuals a_i choosing a given option o_j.

$$\begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{m1} & p_{m2} & \cdots & p_{mn} \end{bmatrix} \tag{4}$$

Letting the aggregate's sum of all individuals' choice for o_j be given by the random variable O_j, its expected value is defined as

$$E[O_j] = \sum_{i}^{n} p_{ij}, \tag{5}$$

which can denote for example the expected value for the volume at a given link.

In agent-based models Matrix (4) is of intractable form, but its marginals, given by Eq. (5) can be estimated from the simulation's output. Each iteration provides a single realization of O_j, while $E[O_j]$, along with its confidence interval (CI), can be estimated from O_j's sampling distribution, i.e. multiple realizations of O_j.

3.3 Statistical methods and tests

For the tests and analysis performed in this thesis, non-parametric methods are used, since they require fewer assumptions on the underlying data, important for agent-based models due to their lack of analytical formulation (Grazzini, 2012). One particular limitation of the local aggregates is their discrete nature, especially for smaller ones. This becomes more pronounced in models that simulate a sample of the total population, where e.g. each additional agent in a 10% model represents a step of 10 in aggregates. Two issues related to this discretization are here solved by adding white-noise to the aggregates. The first relates to the non-parametric tests for ergodicity and stationarity presented here, which are based on the ordering of the data and have issues handling ties, while the second issue prevented is the case of zero-divisions when calculating ratios.

3.3.1 Variability

Output variability is measured *within-run* (between iterations) and *between-run* (same iteration, multiple runs) with a simple relative standard deviation measure, or coefficient of variation (CV), defined as the sample standard deviation divided by the sample mean, which can be conveniently represented as a percentage. Given this, the following measures of variability are proposed:

- CV^I: variability over a set of iterations I
- CV^N: variability over a set of runs N
- $\overline{CV_a^I}$ and $\overline{CV_a^N}$: average CV^I and CV^N over a group of local aggregates a

In a well-behaved simulation and under stationary conditions, the within-run variability is expected to be reflected in the between-run variability. Given this, a variability-ratio measure is proposed. With CV_n^I representing the variability of run n over I and CV_i^N representing ensemble variability at iteration i, the variability ratio is given by

$$VR = \frac{\sum_N CV_n^I}{\sum_I CV_i^N}, \tag{6}$$

which can also be calculated as VR_a for local aggregates by replacing CV^I and CV^N by $\overline{CV_a^I}$ and $\overline{CV_a^N}$ respectively. Intuitively, VR values close to one denote the expected representativeness between internal and ensemble variabilities.

3.3.2 Stationarity

For testing stationarity, the Mann-Kendall trend test is used (Kendall, 1938; Mann, 1945; ten Broeke, 2017). The test's statistic, denoted as τ, offers a measure of the degree of correlation between two random variables and ranges between -1 and 1, where a value of 0 means the variables are not correlated. In the case of a time-series, it measures the auto-correlation, and a value of τ close to 0 means stationarity. τ's asymptotic behavior follows a normal distribution and offers the basis for the two-tailed hypothesis test for trend, where the null-hypothesis is of stationarity, i.e. that τ is equal to 0.

The trend test is negatively affected by dependency of observations that are close in time, i.e. it is highly sensitive to auto-correlation (ten Broeke, 2017), which can be mitigated by applying it to time-windows across iterations instead of single observations, as demonstrated by ten Broeke (2017) and also in a similar case by Grazzini (2012).

The share of aggregates that achieve stationarity can be used as a stopping criterion for the iterative process.

3.3.3 Ergodicity

For testing ergodicity, the Runs test is applied (Wald and Wolfowitz, 1940; Grazzini, 2012; ten Broeke, 2017), which was developed to test whether two samples stem from the same population. In the case of testing ergodicity, one sample consists of N equally-sized time-series generated by different simulation runs and the other sample consists of a single long time-series split into N smaller sub-series of the same length as those from the first sample. If the model is ergodic, the test will confirm that the two samples come from the same population, which is also the test's null hypothesis. Stationarity is a necessary condition for ergodicity (Park, 2018; Grazzini, 2012), and must be tested in advance.

3.3.4 Early stop error

The early stop error refers to the error caused by stopping the iterative process before it has converged to equilibrium. Since the simulation outputs are stochastic and thus samples from a probabilistic distribution, the output from multiple iterations must be considered to estimate the expected value, as discussed previously in Section 3.2.

The calculation of the magnitude of the early-stop error can be given by

$$\Delta_M = \frac{1}{M}\left(\sum_{i=N-M}^{N} x_i - \sum_{j=I-M}^{I} x_j\right) \tag{7}$$

where N is the final iteration, I is the iteration where the simulation would be hypothetically stopped and M is the number of iterations used to calculate the expected value, i and j are iteration indexes and x is the aggregate evaluated. Under stationarity, large M produces a better estimation of the mean, but since this cannot be assumed it is better to use a small enough value to avoid biases due to trends, which are to be expected in this case.

Eq. (7) can also be used as a stopping criterion. By progressing N and I, it can measure the average gradual changes in the aggregates, which can be interpreted as an estimator of the global slope of the trend towards convergence. This would be equivalent to the stopping criterion proposed by Sheffi (1985) for SUE traffic assignment using MSA. Fixing M in Eq. (7) and let Δ_a^n be the deviation in the mean for aggregate a in iteration n, the stopping criterion can be formalized as

$$\frac{\sqrt{\sum_a \left(\Delta_a^{n+1} - \Delta_a^n\right)^2}}{\sum_a \Delta_a^n} \leq k \tag{8}$$

where k is an arbitrary target value for stopping the simulation. This criterion is less strict than the one proposed in Section 3.3.2, since it only requires the trend to be small enough instead of waiting for it to come to a halt (become stationary).

3.4 Experimental setup

This section describes MOBi.sim, the model used in the analysis, as well as the setup used for the different tests and comparisons.

3.4.1 The MOBi.sim model

As described in the first chapter, MOBi.sim is a MATSim model that takes an input demand, consisting of a synthetic population of agents, each having an initial day schedule, and mutually relaxes these plans in an iterative process towards an SUE. MOBi.sim is mostly implemented as a regular MATSim model, with the exceptions of a more elaborate scoring function (allowing more granularity of behavioral parameters), and a time penalty for car trips in dense urban areas. More details of MOBi.sim are offered in Scherr et al. (2018, 2019b).

The innovation modules used for the main subpopulation and their respective replanning rates are:

1. *ChangeExpBeta*: 70%
2. *ReRoute*: 5%
3. *SubtourModeChoice*: 15%
4. *SBBTimeMutation_ReRoute*: 10%

SBBTimeMutation_ReRoute is a module that combines random time mutation and re-route, as described in Section 2.1, but customized for randomly varying time based on the original plan instead of the current values. The allowed range of variation is ± 30 minutes.

MOBi.sim uses the innovation switch-off approach described in Section 2.3, and usually the last 10% of the iterations are executed without innovation. Another important aspect of MOBi.sim for this analysis is that PT vehicles are not capacitated, hence constraints for PT are only imposed by the timetable.

The model is large-scale, covering the entire population of Switzerland, with about 8.5 million agents. The computational effort to run the entire simulation is prohibitively high, and most simulations are run with a 25% or 10% sample of the population.

A large machine (96 cores, 750 GB of RAM) takes about three days to simulate 500 iterations of the 10% scenario. To reduce the computational effort the model is further reduced with the use of a spatial cutter tool recently developed for MOBi. This tool functions by removing agents without trips inside the chosen area and fixating the mode choice for trips crossing the scenario's boundaries. The chosen region was of Luzern and its neighboring cantons, as displayed in Fig. 6. The final model has a total of about 136'000 nodes, 323'000 links and 95'000 agents (10% sample), and its computational time for 500 iterations is of 12 hours.

Figure 6: Geographical boundaries of the cut model

3.4.2 Simulations

Chapter 2 presented the innovation algorithms investigated in this thesis. The traditional one shall be named *Baseline* and be further split into two alternatives, here named *Baseline-innovation* and *Baseline-switchoff*. The former shall represent the innovation algorithm ran with constant replanning rates throughout the iterative process, and the latter the case when the rate is kept constant but innovation is switched off towards the end of the process.

The process of switching innovation off produces a discontinuity in the system, as shown in Fig. 4, which brings difficulties to statistical analysis as well as opens the question of what else could have happened if the stopping point would have been later or earlier. To solve these two issues, the following is proposed:

1. Run a *Baseline-innovation* simulation for N iterations and write the intermediate states of the population of plans P_I at a rate r, where $I = (I_0, I_{1*r}, I_{2*r}, I_{3*r}, \ldots, I_N)$ is the iterations where the innovation would have been hypothetically switched off. $r = 10$ is used.

2. Run one simulation with each of the populations P_I without innovation for a few more iterations, i.e. as if the innovation had stopped at this point.
3. Measure the state of every variable of interest in each of these simulations and build continuous time-series.

Since the state is frozen quite rapidly after innovation is turned off and to limit computational time, the number of iterations ran with each simulation of step **2)** is set at 10. The process described above represents a pseudo-innovation algorithm, since it requires running multiple separate simulations, and is thus highly unsuited for practical applications, but it allows direct comparison with the other innovation algorithms.

The other two innovation algorithms evaluated are named *Anneal*, the algorithm using the gradual reduction of the replanning rates described in Section 2.4.1, and *Greedo*, the optimization-based innovation algorithm described in Section 2.4.2

Each innovation algorithm is run once for 2000 iterations, instead of the typical 500 iterations, for evaluating convergence. Both Anneal and Greedo use an MSA schedule, which is set to start at 100% at iteration 0 and decrease following the schedule defined in Eq. (2) with $\gamma = 0.75$, reaching approximately 1% at iteration 500. At this point the MSA schedule is frozen and this rate is kept constant. After freezing innovation rates, Anneal becomes similar to Baseline-innovation, but with a much smaller innovation rate.

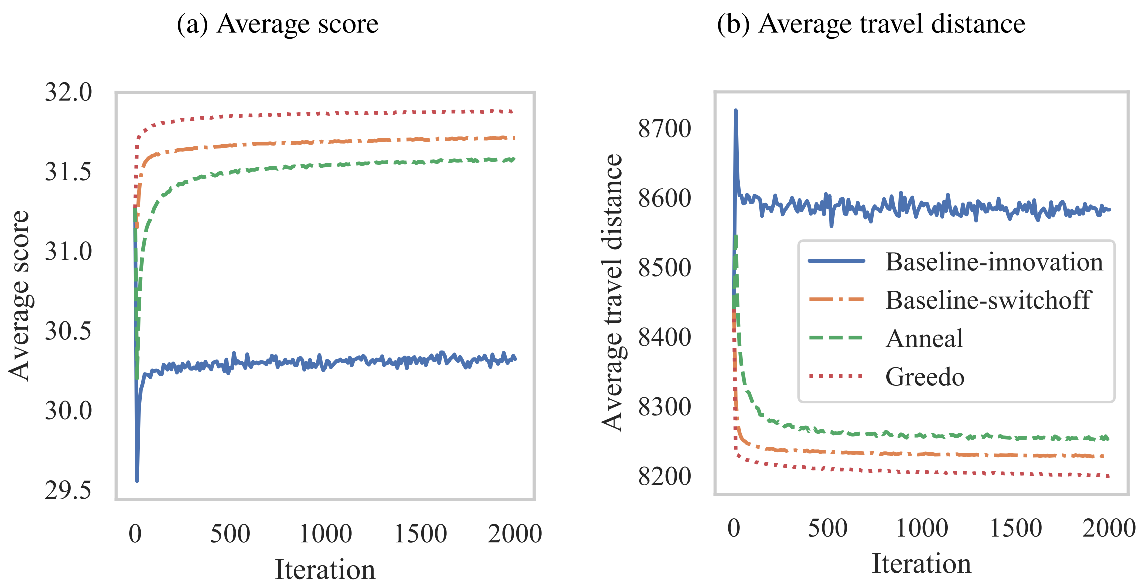
Other properties require ensemble runs, so additional 10 runs of Baseline are run for 450 iterations with the last 50 iterations being without innovation. Anneal and Greedo are also run for another 10 times each, but up to 2000 iterations for evaluating also ergodicity, which won't be evaluated with Baseline due to the difficulty of the requirement of multiple simulations as described above and because of Baseline-innovation's similarity to Anneal once the innovation rate becomes constant.

4 Results

4.1 Global statistics

In MATSim, arguably the easiest way to evaluate the simulation's convergence is by looking at the automatically generated global statistics. The most aggregated ones are the average scores and travel distances, shown in Fig. 7. Although the score function is influenced by more than only travel distances, the two curves, quite remarkably, look vertically mirrored.

Figure 7: Global aggregate statistics

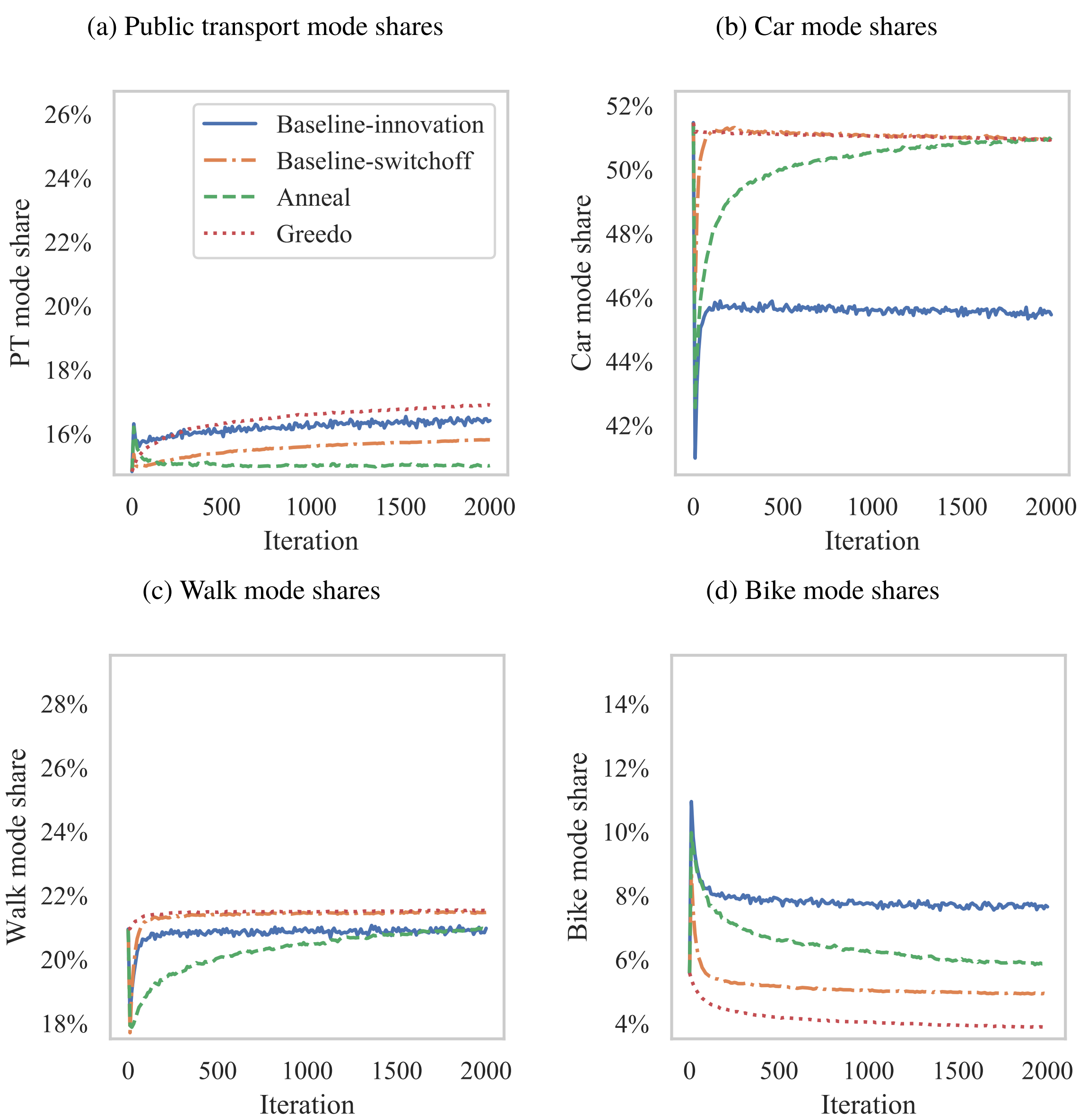


From Fig. 7(a), it might seem that all simulations reached stable results concerning the scores, although at a closer look it appears that a small trend still exists. In absolute terms, Greedo reached the highest score, which signifies that the agents are, in average, more satisfied with their solutions, which, among others, is caused by the smaller average travel distances seen in Fig. 7(b).

Also important to evaluate convergence is the modal split, i.e. the share of trip legs performed with each of the simulated modes. Fig. 8 shows the mode shares in separate subplots and compared among the different innovation algorithms, with the scaling kept fix but the positioning of the Y-axis varied.

As may be seen, none of the algorithms was able to achieve stable mode shares at all modes, indicating an overall lack of convergence. In terms of trend, Baseline-innovation seems to be the most stable. Anneal was the only one to reach stable results at PT as well as the only one not to reach it for car mode.

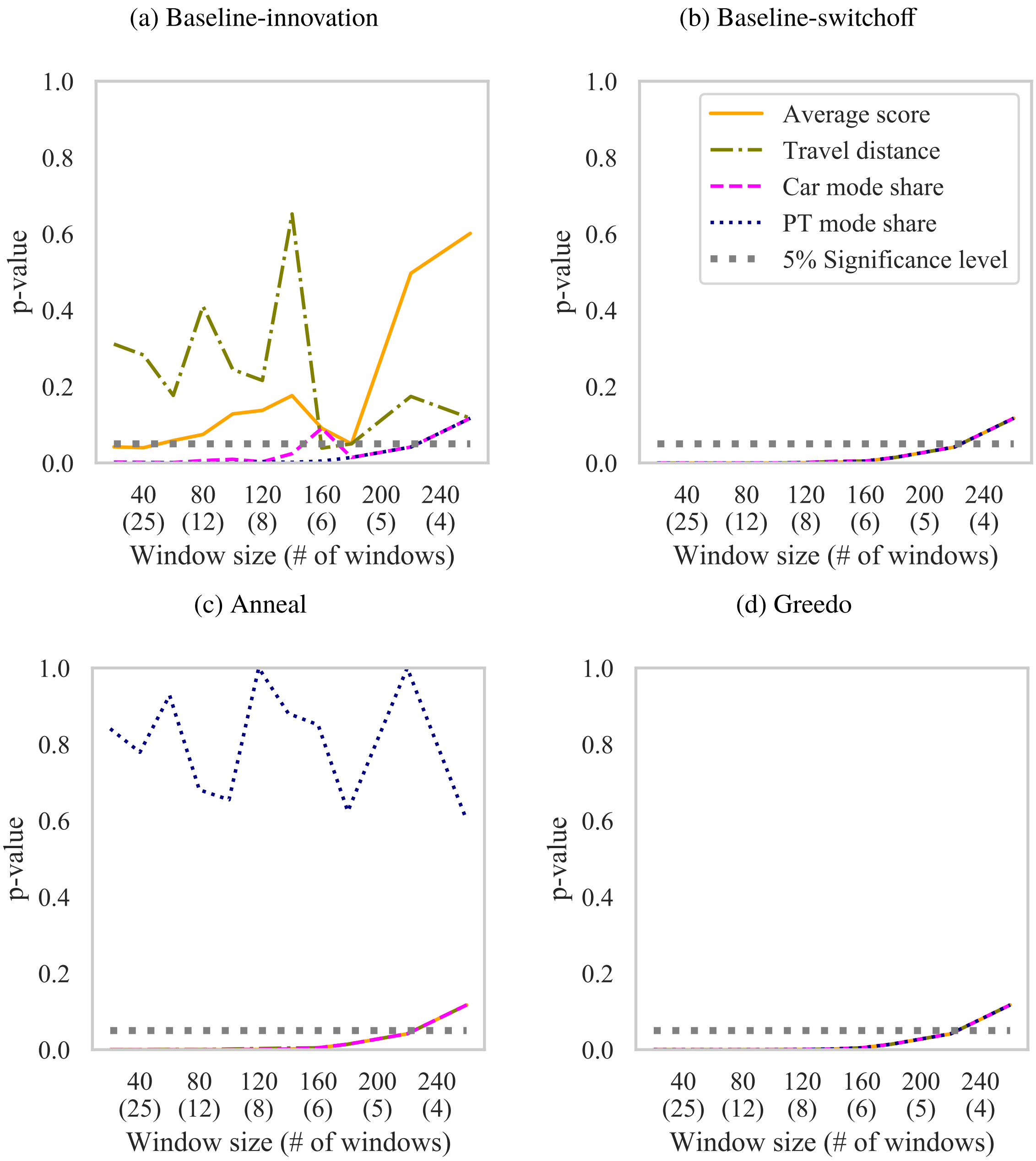
Figure 8: Mode shares (*ride* mode omitted for brevity; please note the different ranges in the Y-axis)



4.1.1 Stationarity

Stationarity tests allow for a more robust assessment of the actual stability of the results. Fig. 9 displays the resulting p-values of the tests on the last 1000 iterations, with varying window

Figure 9: p-values of stationarity tests of global statistics with different window sizes (shares for *bike*, *ride* and *walk* modes omitted here but available in the appendix.)



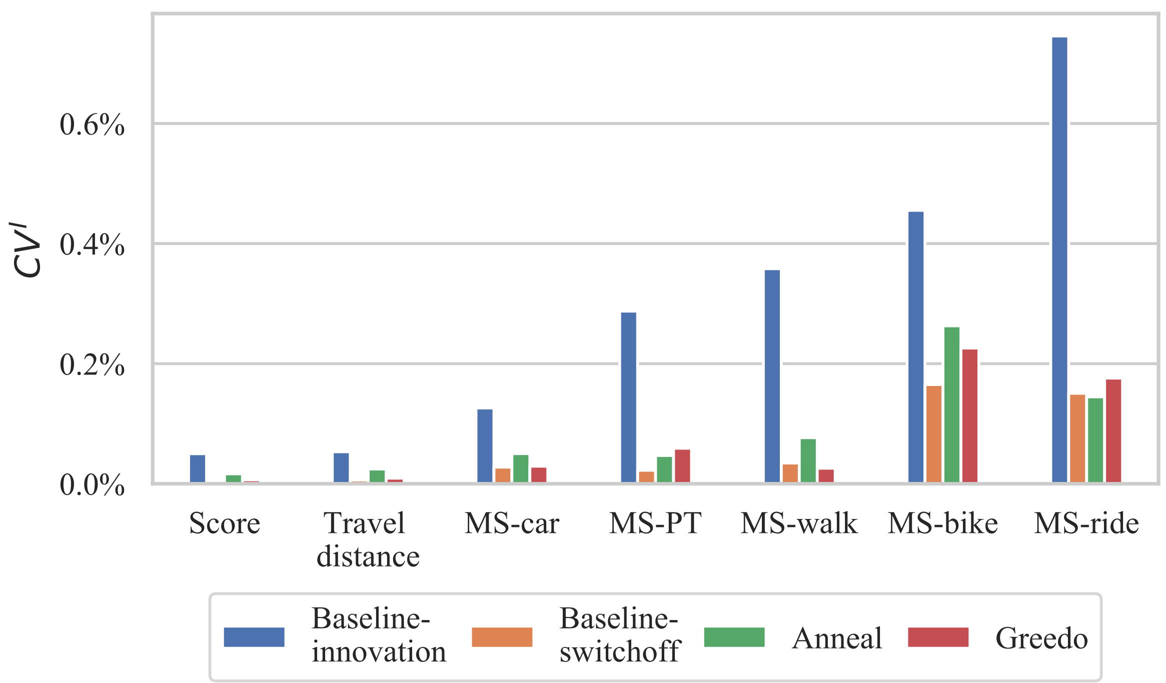
lengths (and number of windows respectively). The analysis with varying window lengths increases confidence in the results Grazzini (2012); ten Broeke (2017).

It may be seen that the algorithm coming closest to stationarity of scores and travel distances is Baseline-innovation, although the same cannot be said for mode shares. It is hard to say that any of the simulations converged after 1000 iterations since the majority of the global statistics had the stationarity hypothesis rejected. The effects of this issue on local aggregates shall be investigated later in this chapter.

4.1.2 Variability

Also of interest is how stable the results are in terms of within-run variability, shown in Fig. 10. For this analysis, only the last 50 iterations were used to avoid biases in the standard deviation due to the existing trends. It is hence assumed that the trend on a span of 50 iterations is negligible.

Figure 10: Variability in global statistics



As it may be seen in the Y-axis, variability at the global scale is very small and even negligible for reporting. However, since these are averages over the entire population, small differences at this scale have large consequences at the local scale, as discussed in Section 4.2.

In relation to the other algorithms, the higher variability of Baseline-innovation reflects its high innovation rate. This causes the algorithm to converge faster, as seen in the previous section, but also leads to a worse equilibrium, as seen by the scores in Fig. 7(a). The reason why modes *bike* and *ride* have in general higher variability could be explained by them being unconstrained, since these modes, along with *walk* are teleported rather than network-simulated.

4.2 Local aggregates

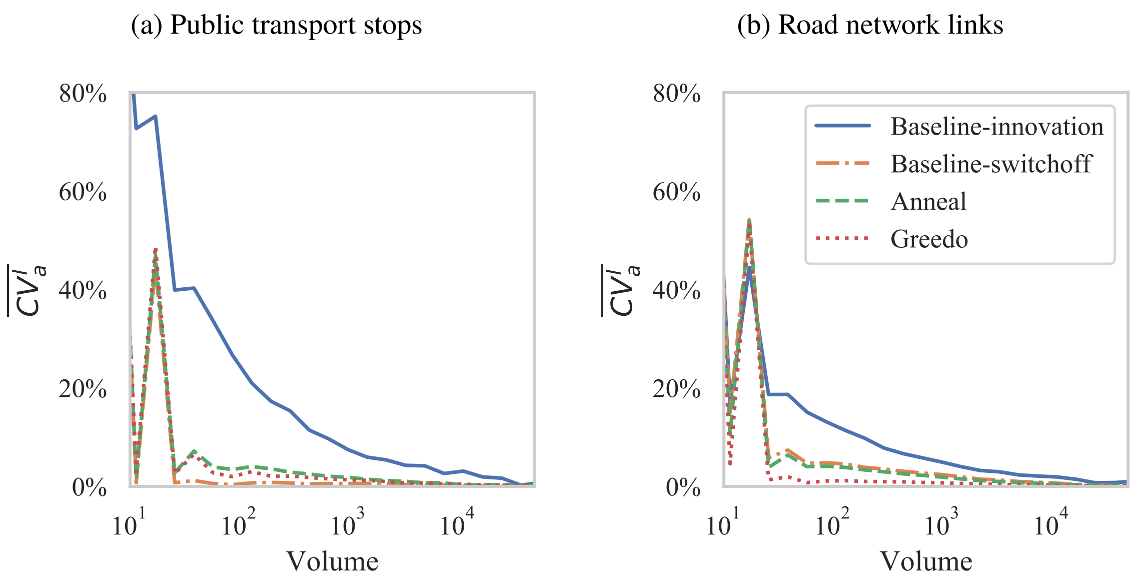
This section evaluates the results of local aggregates, namely link and PT stop volumes. Volumes refer to the total road traffic of a day in the case of links and to the sum of boardings and alightings for an entire day in the case of public transport stops. Of particular interest is the evaluation of how the non-convergence detected in the global statistics, in Section 4.1, affects the local aggregates.

4.2.1 Variability analysis

This section analyzes the variability of the local aggregates and partially reproduces the work done by Horni et al. (2011); Paulsen et al. (2018).

The variability results for the last 50 iterations are displayed in Fig. 11. The plots confirm previous results from the literature, in that smaller aggregates display higher variability.

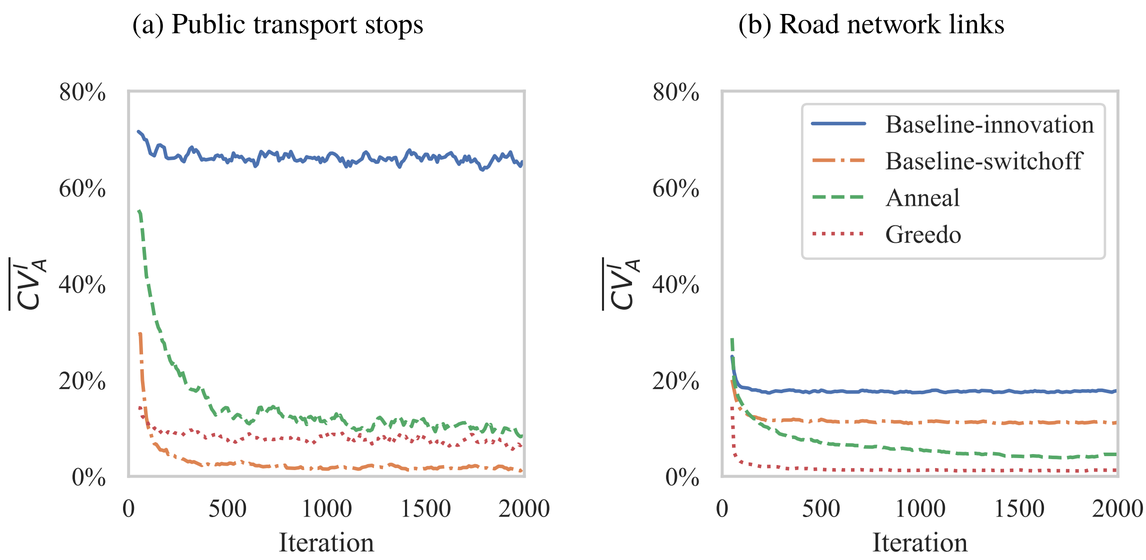
Figure 11: Within-run variability of local aggregates



As expected due to its high innovation rates, Baseline-innovation has the highest variability. Greedo has the lowest variability in regards to links, whereas Baseline-switchoff displays the lowest results for stops.

Of particular interest to the convergence analysis is whether variability improves with the

Figure 12: Progression of within-run variability



progression of the simulation. Fig. 12 displays the results of this analysis, where $\overline{CV_a^I}$ is progressively measured over windows of 50 iterations. A slight downward trend can be seen in some of the curves, meaning that although variability diminishes as the simulation progresses, a minimum is eventually reached.

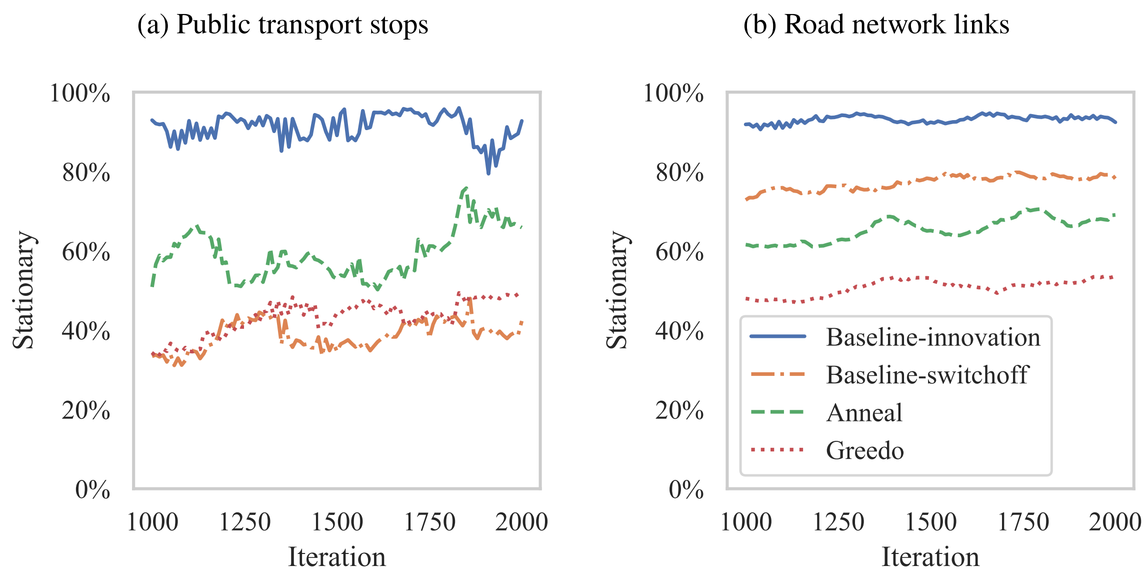
Baseline-innovation has the highest results, although with apparently little or no trend, similarly as what was observed in Section 4.1.2. Greedo has very low variability in respect to links and reaches this point also very quickly, which is related to the fact that minimizing variability is part of its objective.

Figs. 11 and 12 show that links are overall more stable than PT stops. This might have to do with the fact that PT in MOBI.sim is not constrained, and also explain why Greedo is less successful with PT variability - unconstrained PT leads to fewer consequences on the system variability and are thus less limited by Greedo.

4.2.2 Stationarity

Fig. 13 displays the results of the stationarity analysis, where the Y-axis represents the share of *stationary volume*, defined as the total volume of stationary aggregates divided by the total volume of all aggregates, hence a weighted stationarity measure.

Figure 13: Progression of stationary (weighted by aggregate size)

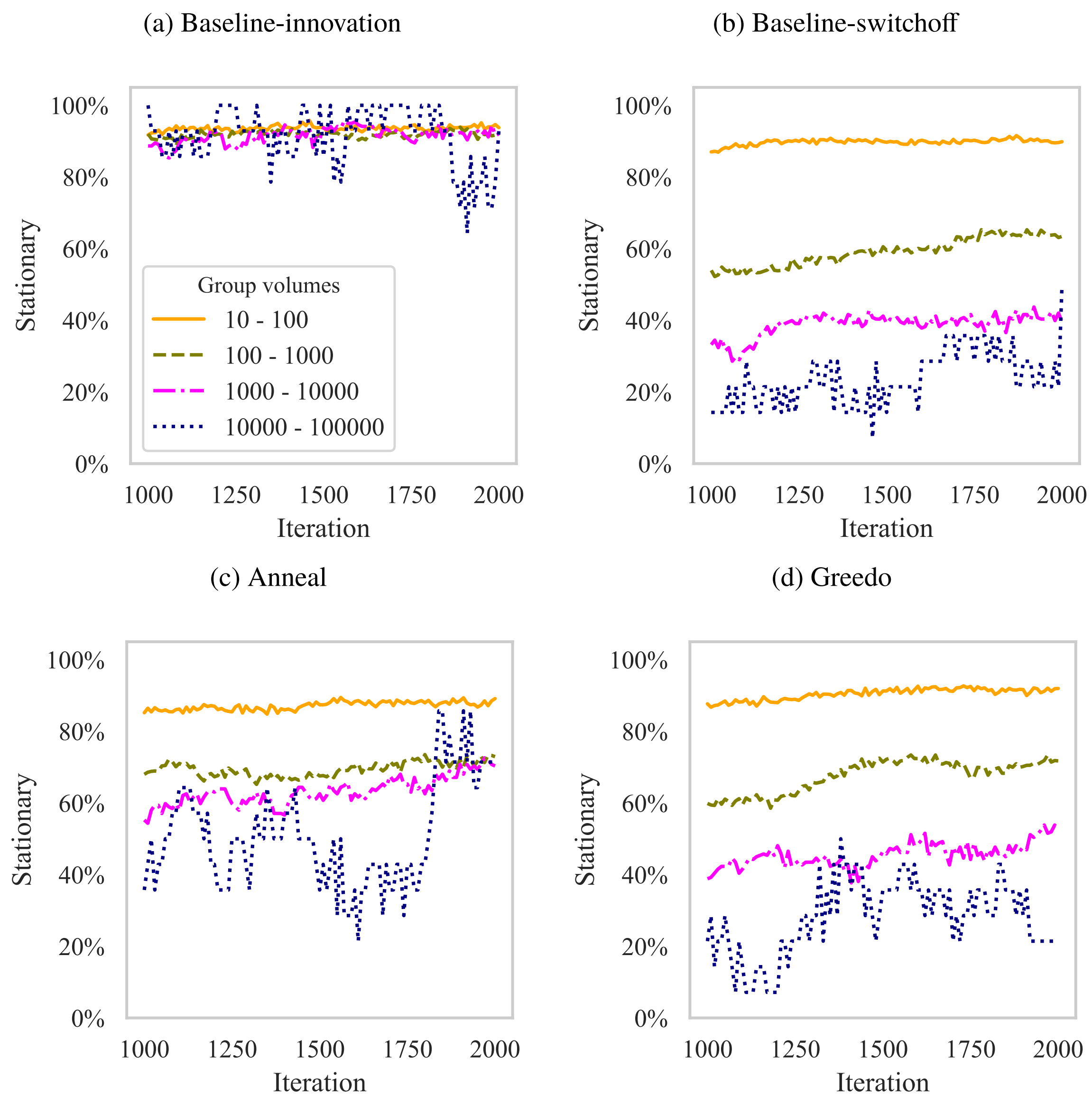


Here Greedo achieves worse results, contrasting to those of the previous section, indicating that although outputs are more stable, they are likely to have a long-term trend. Baseline-innovation on the other hand achieves highest stationarity shares, similarly as with global aggregates in Section 4.1.1 and likely for the same reasons.

Fig. 14 displays the results of the stationarity tests for PT stops grouped by volumes (logarithmic bins). The same plots for links are omitted here for brevity but available in the appendix. The plots indicate that larger aggregates take longer to converge to a stable state. This goes in the same direction as the results from Llorca and Moeckel (2019), where they showed that simulations with more agents take longer to converge.

The fact that a share of the local aggregates achieves stationarity despite lack of global convergence shows that parts of the system may reach a stable state while other parts are still changing. There is no guarantee however that these aggregates will remain stationary, due to the high correlation with the rest of the system, and in particular because of the zero-sum nature of the aggregates, as discussed in Section 3.2.

Figure 14: Share of stationary per aggregate volume for PT stops



4.2.3 Early stop error

This section analyzes the magnitude of the errors caused by an early stop in the simulation. Δ_M, from Eq. (7), is calculated for every local aggregate and displayed in Fig. 15, grouped by volumes in (logarithmic) bins. The number of aggregates within each bin, averaged between the four algorithms, is overlaid in gray below the boxplots. The deviation is calculated between iterations 500 and 2000, setting M (number of iterations to average) to 50.

Each data-point in the graph is the Δ_M of an aggregate. The range of the boxplots show the range of the distribution of the Δ_M for a particular volume category. Of particular interest is the middle of the boxplots, which represent the median and display the existence of a trend in the group in case it lays above or below the dashed 0% line. A positive Δ_M means the aggregate increased its volume over the course of 1500 iterations. In an ideal case, the boxplots should be as tight as possible, indicating small variability, and the medians as close as possible to 0%, indicating minimal early-stop error in the group.

In fact what can be seen from the plot is that an overall trend is present, as indicated by the medians deviating from 0%. This trend is positive for stops and negative for links, and clearer for larger stops since most of the boxplots lay entirely above the dashed line. Fig. 15 shows also the variability evaluated in Section 4.2.1, from the large and decreasing ranges of the boxplots - smaller aggregates have higher variability. The aggregates are here averaged over M, which is supposed to mitigate variability. Long-term variability due to trends in larger aggregates could be the reason why averaging was not able to fully mitigate the large ranges in smaller aggregates. For example, a stationary small PT stop may still be affected later due to a trend on a large upstream stop.

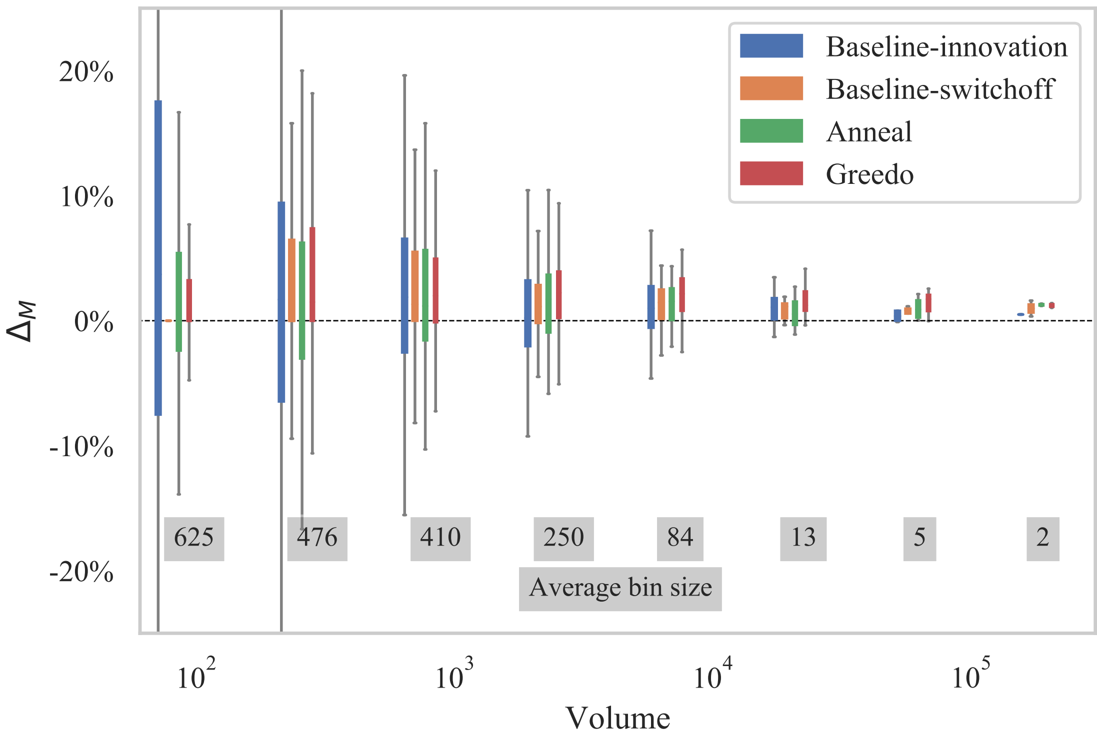
Even though large aggregates show a persistent trend, the magnitude of the errors is very small, about 1-2%. On the other hand, although small aggregates are more likely stationary, as seen in the previous section, they are still subject to large early stop errors.

The results of the stopping criterion proposed in Section 3.3.4 are shown in Fig. 16. This stopping criterion measures the gradual reduction of Δ_M over the simulation. Intuitively the plots may be interpreted as the progression of the average trend present in the aggregates. Note in the vertical axis of each plot that the results are different by an order of magnitude, likely due to the smaller variability in links.

The comparison among the innovation algorithms shows that Greedo achieves a higher stability and faster than the others, especially for links, while Baseline-innovation achieves worse results. By visual inspection, it seems that Baseline-innovation has the smallest trend among the curves but also the highest k values. This reflects what was seen in Fig. 15, where the algorithm displayed smaller Δ_M on average, representing smaller trend, but wider distribution, representing the high variability. Differently from Δ_M, k is always positive so it is not expected that the Baseline-innovation values for k would reduce any further, in contrast to the other algorithms. This is another way to measure the level of precision an algorithm is able to achieve, with the highest precision being achieved by Greedo for links.

Figure 15: Distribution of early stop errors (boxplots for small aggregates and some of the whiskers are concealed in order to improve the visibility of the other boxplots)

(a) Public transport stops



(b) Road network links

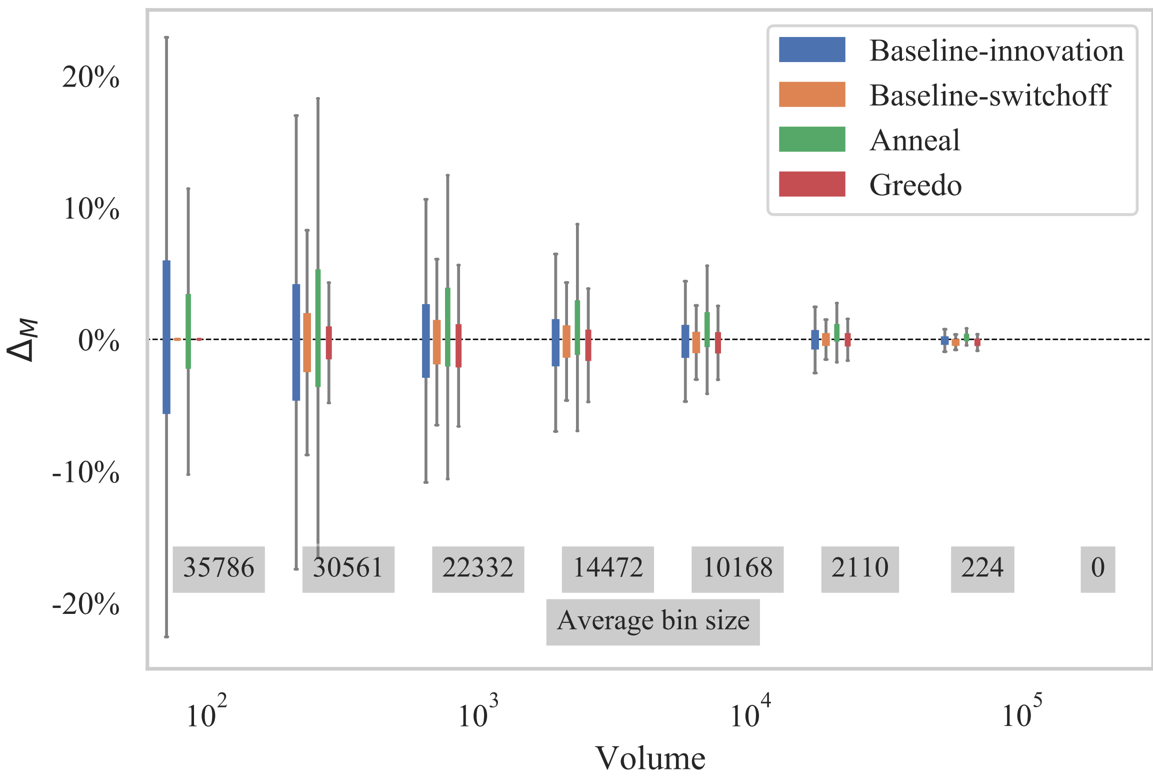
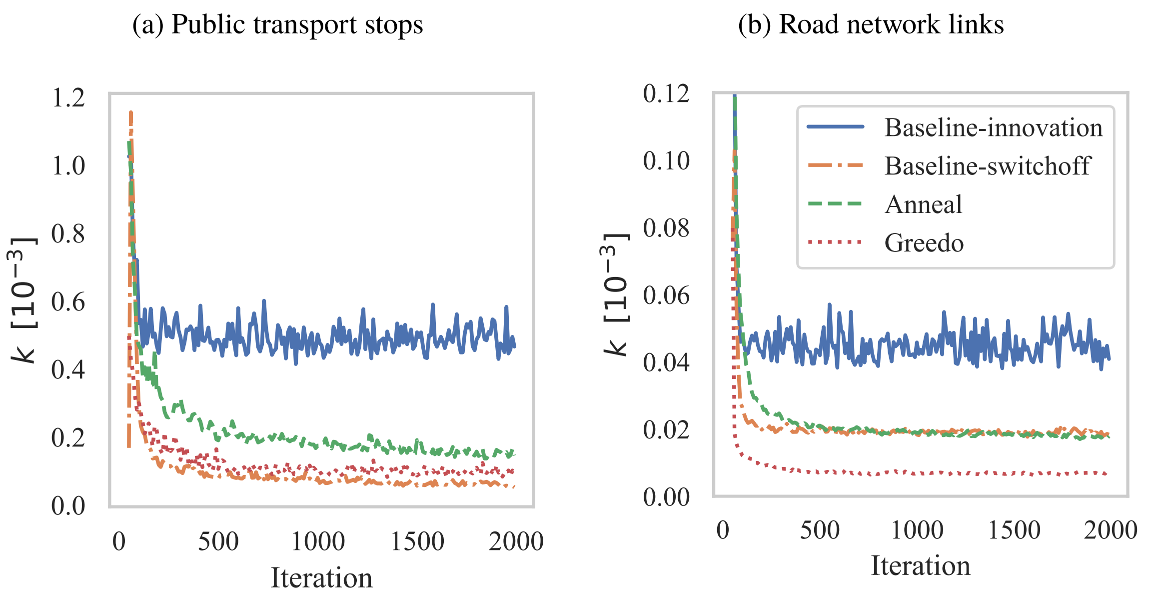


Figure 16: Stopping criterion



4.3 Ensemble analysis

This section looks at an aspect of convergence that cannot be measured in a single run - how well the population parameters can be estimated by the sample statistics. In practical terms, it is of interest to identify whether the outputs of a single run are a good approximation of the ensemble run.

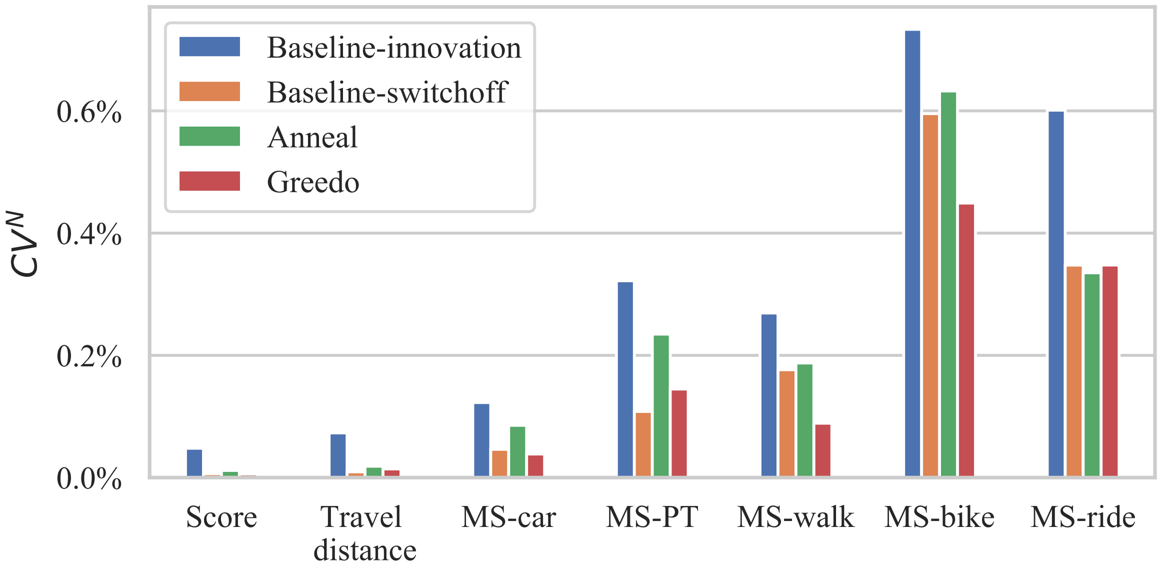
4.3.1 Variability

For evaluating ensemble variability, different iterations were used in contrast to the analysis of single simulation runs, as described in Section 3.4.2, so the results are not directly comparable. For these calculations, between-run statistics were calculated for each of the last 50 iterations and then averaged. This section also partially reproduces the work of Horni et al. (2011); Paulsen et al. (2018).

Fig. 17 compares the between-run variability CV^N among the different algorithms. All results are below 1%, showing that at this level of aggregation, variability is negligible.

Fig. 18 displays the VR, or variance ratio, for each innovation framework. Values close to one

Figure 17: Between-run variability of global statistics



mean that the variability of the ensemble is well captured in individual runs. Here, Baseline-innovation has values close to one for all global aggregates, and since Anneal which is very similar has not, this is probably because it achieved convergence. Baseline-switchoff, on the other hand, has the lowest values of VR, hence where the sample variability is least representative of the population. The consequence is although minimized by the overall low CV^N values seen in Fig. 17.

The same analysis as above can be replicated with local aggregates. Firstly, the $\overline{CV_a^N}$ values are shown in Fig. 19, where the magnitude of the errors caused by running a single simulation can be seen. Once again, the size of the aggregate plays a critical role, where large aggregates have CV^N comparable to that of global statistics but smaller ones can have substantial deviations.

Finally, the VR may also be calculated for individual aggregates, as displayed in Fig. 20. Similarly as before, the variability in Baseline-innovation is highly representative, Baseline-switchoff is little representative and Anneal and Greedo are in the middle. The fact that the other three innovation algorithms have not reached VR_a values close to one may be caused by their lack of convergence.

Figure 18: Variability ratio

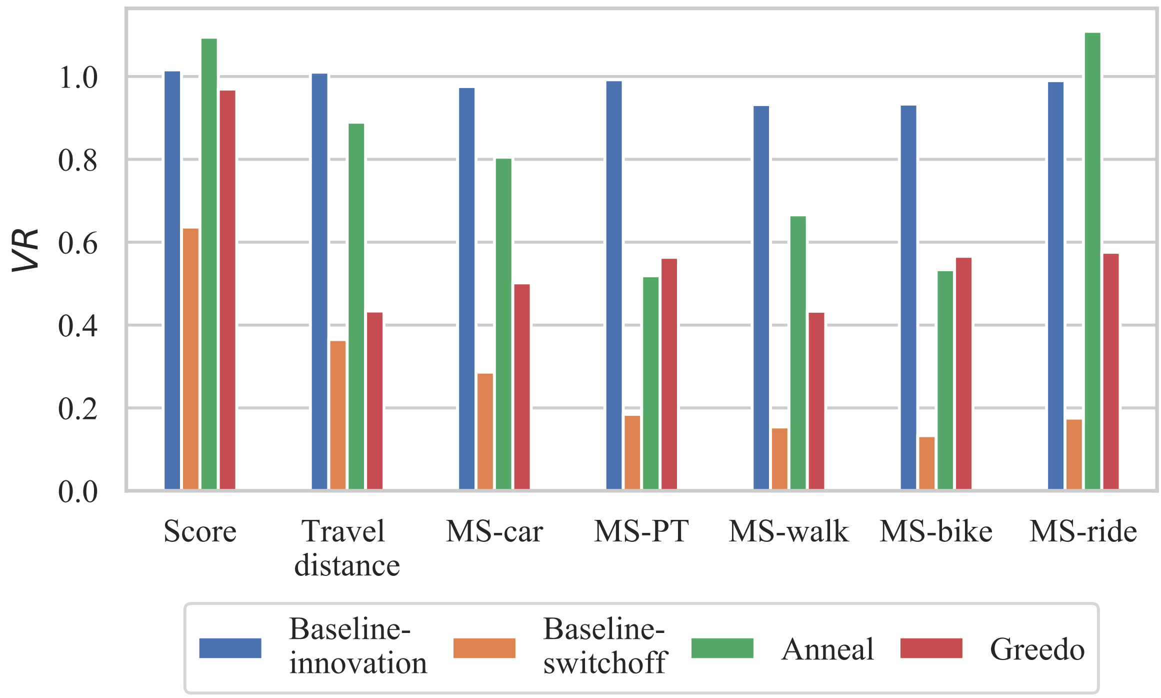


Figure 19: Average between-run variability per volume

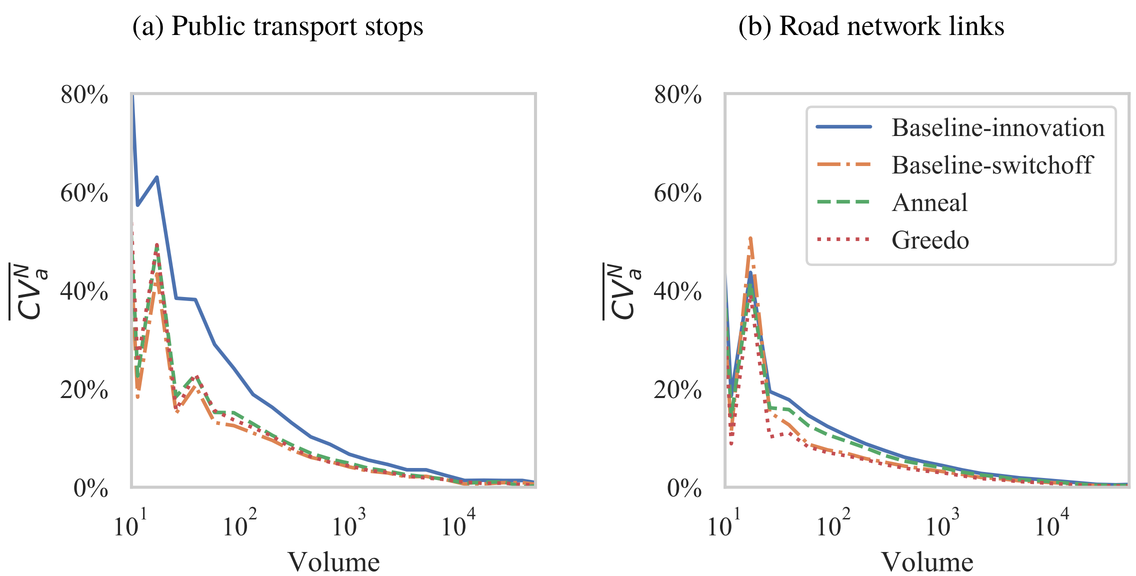
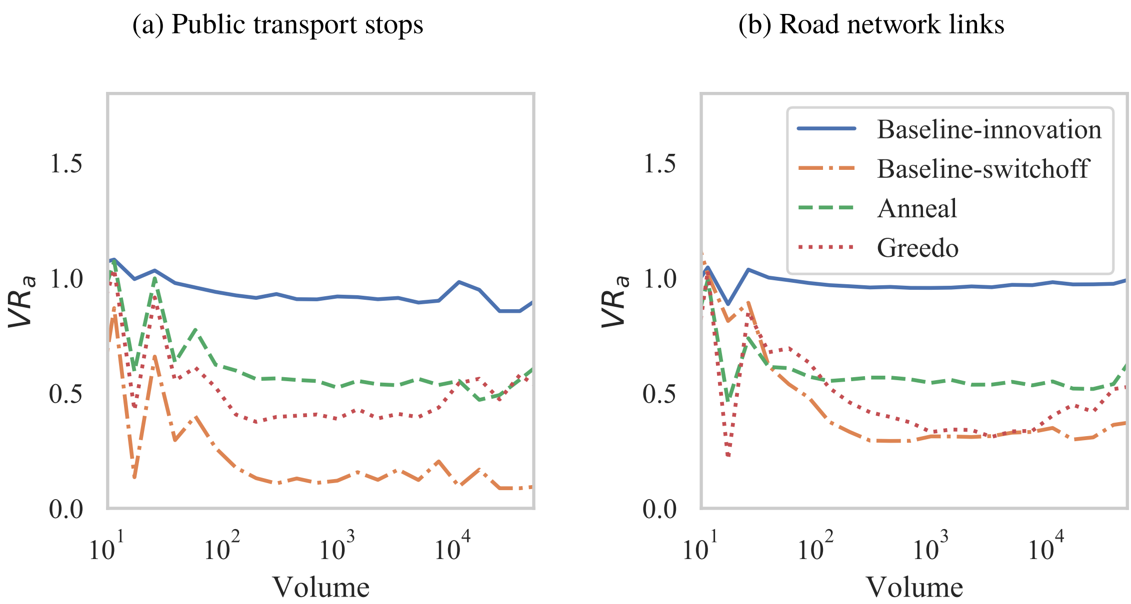


Figure 20: Variability ratio of local aggregates



4.3.2 Ergodicity analysis

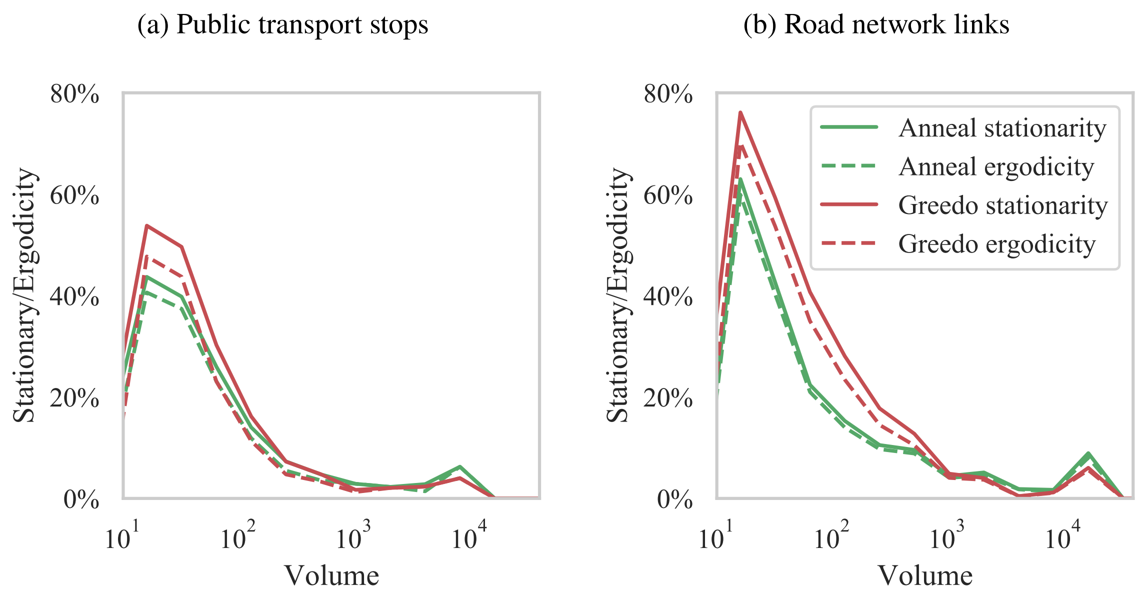
Since stationarity is a necessary condition for ergodicity, the analysis cannot be done for global statistics since they have not passed this test (Section 4.1.1). Local aggregates that achieved stationarity over all runs can, however, be tested, but with the cautionary note that, despite their ensemble stationarity, it is still possible that these aggregates are not yet at their equilibrium state.

For this test, the ideal case would be to have at least 30 runs in the ensemble run, to have a better approximation of the test statistic to the normal distribution. For practical reasons, this was not possible, so each run shall contribute with three samples to make up to 30. This is not ideal, but since the null hypothesis is of ergodicity, artificially increasing the sample size will only make the test more likely to reject the null hypothesis, thus providing at least a lower bound for ergodicity among the samples.

With a window size of 30, each run then contributes with 90 iterations to build the test sample. The base sample is an additional run, where 900 iterations are used to build the 30 windows of 30 iterations.

Fig. 21 displays the results of the ergodicity test for Anneal and Greedo. It can be seen that almost all aggregates that pass the stationarity test also pass the ergodicity test, which means the

Figure 21: Stationarity and ergodicity of local aggregates



procedure of increasing samples described above probably did not have a substantial impact since more samples in the ensemble run would only make the ergodicity curve even closer to the stationarity curve.

The reason why a smaller share of aggregates is stationary in Fig. 21 in comparison to Fig. 13 is that here only aggregates stationary across all runs are considered. The large majority of these also passed the ergodicity test, which makes them more likely to be at their equilibrium state since they also all converged to the same distribution (despite incomplete global convergence).

4.4 Results overview

The results from this chapter have shown the different aspects of equilibrium and search thereof for each of the innovation algorithms investigated.

As expected, Baseline-innovation, with its high replanning rate, converges faster, but also has very high variability - which decreases confidence in local scale outputs, in the behavioral model due to its high share of agents performing random innovation and thus in the overall quality of the achieved equilibrium. Baseline-switchoff, a pseudo-algorithm constructed here to analyze the hypothetical effect of switching innovation off at different iterations, in principle should benefit from Baseline-innovation's fast convergence rate and mitigate the variability problem at the same time, but what was seen is that switching-off at different iterations has a long term trend effect, i.e. it looses much of the fast convergence advantage, likely due to the heterogeneity between the two states (before/after switch-off). In terms of variability, even though it may greatly decrease the within-run variability, some of it still shows up as between-run variability, breaking the representativeness of a single run, as seen in Figs. 18 and 20.

The two alternative algorithms tested, Anneal and Greedo, offer incomplete solutions to the issues of Baseline, but solve fundamental problems with its approach and are thus promising alternatives. Anneal is a solution in the middle since it reaches much lower variabilities and produce more representative results, but at a cost of even slower convergence rates. Greedo is a somewhat different approach, but the final results are still comparable to the other algorithms, which is a positive result for the iterative agent-based transport modelling approach in general. The algorithm achieves high scores much faster than the others, but a long-term trend remains that delays convergence. This might be due to its UE design, where agents are *more rational* and hence able to optimize routes and departure times in greater detail.

Table 1 provides a qualitative and relative comparison between the studied algorithms, showing which algorithm performed best, worse and somewhat in-between at each analysis. The arrows in the table offer a row-wise ranking between the algorithms, i.e. upwards arrow means that the algorithm performed best in comparison to the others in this analysis, downwards arrow that it performed worst and straight arrow that it obtained results somewhat in the middle.

Table 1: Qualitative summary of results (relative comparison between algorithms)

Statistic	Baseline-innovation	Baseline-switchoff	Anneal	Greedo
Score	↘	→	→	↗
Score (conv. speed)	↗	→	↘	↗
Travel distance	↘	↗	→	↗
Mode shares	↗	↗	↘	↗
Mode shares (conv. speed)	→	→	↗	↗
Stationarity (global stat.)	↗	↘	→	↘
Stationarity (stops)	↗	↘	→	↘
Stationarity (links)	↗	→	→	↘
WR Variability (global stat.)	↘	↗	→	↗
WR Variability (stops)	↘	↗	→	↗
WR Variability (links)	↘	→	↗	↗
BR Variability (global stat.)	↘	↗	→	↗
BR Variability (stops)	↘	↗	↗	↗
BR Variability (links)	↘	→	→	↗
Variability ratio VR (global stat.)	↗	↘	→	↘
Variability ratio VR (stops)	↗	↘	→	→
Variability ratio VR (links)	↗	↘	→	↘
Ergodicity (stops)	X	X	↗	↗
Ergodicity (links)	X	X	↗	↗

5 Discussion

The results from the experiments conducted in this thesis show that, in general, overall convergence takes much longer than the typical 500-1000 iterations. Even though the score curves become considerably stable after about 200 iterations, a persisting trend shows the presence of ongoing internal processes that cannot be neglected if local aggregates are to be analyzed and not only global travel distances. This issue is also present in mode shares (Fig. 8), and although the deviations might be relatively small (Fig. 15) for large aggregates, smaller aggregates are still strongly affected. The lack of stationarity also hinders the possibility of calculating more accurate and interpretable outputs, consisting of the mean and confidence intervals. This is true even for smaller aggregates that achieve stationarity, since they are affected in the long-term by larger aggregates, as discussed in Section 3.3.4.

In terms of equilibrium uniqueness, or ergodicity, although the incomplete convergence hinders the analysis, it seems that the iterative traffic assignment and learning of agents in MATSim can achieve this desirable property with different learning algorithms, confirming previous results from Nagel et al. (2000); Rickert (1998). The switch-off procedure described in Section 2.3 seems to break ergodicity. Paulsen et al. (2018) in their variability analysis with the Santiago scenario, which also uses this approach, noticed the same issue. The ensemble variability becomes however only substantial for smaller aggregates, which minimizes the issue depending on the analysis intended. Also in terms of variability, the results confirmed previous findings from Paulsen et al. (2018); Horni et al. (2011), that variability decreases with larger aggregate sizes.

In practical terms, each innovation algorithm tested has its advantages and disadvantages, which are summarized in Table 2.

Table 2: Pros and cons analysis of the evaluated innovation algorithms

Algorithm	Pros	Cons
Baseline-innovation	Fast conversion to equilibrium	High variability Equilibrium of lower quality
Baseline-switchoff	Low variability	Unpractical for equilibrium analysis Interpretability issues
Anneal	Low variability Simple and robust formulation	Slow convergence
Greedo	Highest scores Minimal variability	Slow convergence Higher computational cost

6 Conclusion

6.1 Outlook

This thesis provides an analysis of the convergence towards equilibrium of a large scale agent-based transport model for Switzerland developed at the Swiss Federal Railways. With the perspective of transport planning practice, it analyzed statistics not only at global but also at the local scale, since the latter is closer to the demands of practical scenario evaluation.

The analysis brought insight into two dimensions of specification uncertainty, or error, that of stochastic variability and that of early-stop. For the first, it reproduced and expanded some of the analyses previously done in the literature by measuring variability, which translates into error due to using the output of a single iteration (within-run variability) or using the output of a single run (between-run variability). For the early-stop error, it looked into the issue of stopping the iterative process before convergence has reached by comparing it to a later stopping option.

The research here presented went on to investigate the actual presence and uniqueness of equilibrium, by applying rigorous non-parametric statistical tests for stationarity and ergodicity. Stationarity tests on global statistics identified the lack of equilibrium even after 1000 iterations. The analysis of local aggregates, such as the output of links and PT stops, showed that subsets of the model sometimes become stationary, although still subject to long-term effects which might break this state. The lack of overall stationarity prevented a complete analysis of ergodicity, but the test was still applied to local aggregates which passed the first test. Almost all aggregates which were stationary among all runs of the ensemble were also ergodic, giving a strong hint that the studied iterative process is, in fact, able to produce unique equilibria. An analysis of stochastic variability has also shown that, close to convergence, the within-run variability is representative of the between-run variability.

By comparing four different innovation algorithms and obtaining results *structurally* similar, this work adds to the evidence of the robustness of the iterative agent-based transport modelling approach, similarly to what Nagel et al. (2000) demonstrated when comparing different traffic assignment algorithms. It leaves a cautionary note however to the approach of arbitrarily switching innovation off after a certain stability threshold is verified, by showing the resulting instabilities in the final results. A positive relationship between innovation and convergence rate was also identified, although also pointing that high innovation rates cause a loss of quality in the final results due to high variability.

6.2 Recommendations for future research

For future research, it would be interesting to expand the analysis with results positively converged (i.e. overall stationarity reached), by running more iterations or a smaller model, and better test the usefulness of the two convergence criterions proposed in this thesis. An analysis of which dimensions of the model can be simplified to speed-up convergence along with the potential losses, could be very useful, e.g. in the lines of Llorca and Moeckel (2019).

The role of estimators, as proposed by Fourie et al. (2013) and applied in Greedo (Flötteröd, 2019), could be further investigated in the context of computationally heavy agent-based transport simulations. In their work, an estimation technique is used within the MATSim loop, but in principle, the same approach could also be applied to obtain estimations of the final results. This approach could speed-up the evaluation of simpler scenarios when computing the entire simulation until convergence becomes too costly.

6.3 Recommendations for practice

For practical applications, global analysis is only the first step, since scenario evaluations often are concerned with local effects. The results of the analyses performed in this thesis depend strongly on the model it is applied to and generalizations require that practitioners apply the tests to their own models, although the methods and results provided here offer a guide. It is thus recommendable that transport planners understand what are the uncertainties that underlie agent-based models, particularly at the scale relevant to a given analysis. This in turn pushes transport modellers to further improve the model.

In practical terms, improving convergence to achieve global stationarity is advised, since at this state it becomes possible to accurately estimate confidence intervals for variables of interest. The alternative algorithms tested in this thesis offer a starting point. The Anneal algorithm is an easy to implement option and could offer improvements to convergence without much overhead, being an almost immediate alternative to the switch-off procedure, which as shown throughout the thesis, has important limitations. Greedo, as a more advanced option, may offer more control and expanded possibilities for improvements in convergence and the overall equilibrium. Other possibilities to speed up convergence include simplifying the model by reducing the search space and randomness. Some ways of doing this include reducing the model's geographical scope, the population sample size and the granularity of the road network.

7 References

Chakirov, A. (2016) Urban mobility pricing with heterogeneous users, Ph.D. Thesis, ETH Zurich, Zurich.

Chakirov, A. and P. J. Fourie (2014) Enriched Sioux Falls scenario with dynamic and disaggregate demand, *Technical Report*, **978**, IVT, ETH Zurich, Zurich.

Flötteröd, G. (2016) MATSim as a Monte-Carlo Engine, in A. Horni, K. Nagel and K. W. Axhausen (eds.) *The Multi-Agent Transport Simulation MATSim*, 361–370, Ubiquity Press, London.

Flötteröd, G. (2019) A method for particle-based traffic assignment, KTH Royal Institute of Technology, Stockholm. Unpublished manuscript.

Flötteröd, G. and B. Kickhöfer (2016) Choice models in MATSim, in A. Horni, K. Nagel and K. W. Axhausen (eds.) *The Multi-Agent Transport Simulation MATSim*, 337–346, Ubiquity Press, London.

Fourie, P. J., J. Illenberger and K. Nagel (2013) Increased convergence rates in multiagent transport simulations with pseudosimulation, *Transportation Research Record*, **2343**, 68–76.

Gibb, J. and J. L. Bowman (2007) Convergence of an activity-based travel model system to equilibrium: experimental designs and findings, paper presented at the *11th National Transportation Planning Applications Conference*, Daytona Beach, Florida.

Grazzini, J. (2012) Analysis of the emergent properties: Stationarity and ergodicity, *Journal of Artificial Societies and Social Simulation*, **15** (2) 7.

Hayashi, F. (2000) *Econometrics*, Princeton University Press, Princeton.

Hemdan, S., A. M. Wahaballa and F. Kurauchi (2019) Travel time variability and macroscopic fundamental diagram relationships in multimodal networks, *International Journal of Intelligent Transportation Systems Research*, **17** (2) 114–124.

Horni, A., D. Charypar and K. W. Axhausen (2011) Variability in transport microsimulations investigated with the multi-agent transport simulation MATSim, *Technical Report*, **692**, IVT, ETH Zurich, Zurich.

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

Kendall, M. G. (1938) A new measure of rank correlation, *Biometrika*, **30** (1/2) 81–93.

Llorca, C. and R. Moeckel (2019) Effects of scaling down the population for agent-based traffic simulations, *Procedia Computer Science*, **151**, 782–787.

Mann, H. B. (1945) Nonparametric tests against trend, *Econometrica: Journal of the Econometric Society*, **13** (3) 245–259.

Meister, K. (2011) Contribution to agent-based demand optimization in a multi-agent transport simulation, Ph.D. Thesis, ETH Zurich, Zurich.

Nagel, K. and G. Flötteröd (2012) Agent-based traffic assignment: going from trips to behavioral travelers, in R. M. Pendyala and C. R. Bhat (eds.) *Travel Behaviour Research for an Evolving World: Select papers from the12th International Conference on Travel Behavior Research*, 261–294, Lulu publishers, Jaipur.

Nagel, K., M. Rickert, P. M. Simon and M. Pieck (2000) The dynamics of iterated transportation simulations, *arXiv e-prints*.

Nash, J. (1951) Non-cooperative games, *The Annals of Mathematics*, **54** (2) 286–295.

Park, K. I. (2018) *Fundamentals of Probability and Stochastic Processes with Applications to Communications*, Springer, Cham.

Paulsen, M., T. K. Rasmussen and O. A. Nielsen (2018) Output variability caused by random seeds in a multi-agent transport simulation model, *Procedia computer science*, **130**, 850–857.

Petrik, O., M. Adnan, K. Basak and M. Ben-Akiva (2018) Uncertainty analysis of an activity-based microsimulation model for Singapore, *Future Generation Computer Systems*, doi: 10.1016/j.future.2018.04.078.

Rasouli, S. and H. Timmermans (2012) Uncertainty in travel demand forecasting models: literature review and research agenda, *Transportation Letters*, **4** (1) 55–73.

Richiardi, M. G., R. Leombruni, N. J. Saam and M. Sonnessa (2006) A common protocol for agent-based social simulation, *Journal of Artificial Societies and Social Simulation*, **9** (1) 15.

Rickert, M. (1998) Traffic Simulation on Distributed Memory Computers, Ph.D. Thesis, University of Cologne, Cologne.

Scherr, W., P. Bützberger and N. Frischknecht (2018) Micro meets macro: a transport model architecture aiming at forecasting a passenger railway's future, paper presented at the *18th Swiss Transport Research Conference*, Ascona, May 2018.

Scherr, W., C. Joshi, P. Manser, N. Frischknecht and D. Métrailler (2019a) An activity-based travel demand model of Switzerland based on choices and constraints, paper presented at the *19th Swiss Transport Research Conference*, Ascona, May 2019.

Scherr, W., P. Manser and P. Bützberger (2019b) SIMBA MOBi: microscopic mobility simulation for corporate planning, paper presented at the *47th European Transport Conference*, Dublin, October 2019.

Sheffi, Y. (1985) *Urban Transportation Networks: Equilibrium Analysis with Mathematical Programming Methods*, Prentice-Hall, Englewood Cliffs.

Smith, L., R. Beckman, D. Anson, K. Nagel and M. Williams (1995) TRANSIMS: Transportation analysis and simulation system, paper presented at the *Fifth National Conference on Transportation Planning Methods Applications*, Seattle, April 1995.

ten Broeke, G. (2017) Sensitivity analysis methodologies for analysing emergence using agent-based models, Ph.D. Thesis, Wageningen University, Wageningen.

Wald, A. and J. Wolfowitz (1940) On a test whether two samples are from the same population, *The Annals of Mathematical Statistics*, **11** (2) 147–162.

Zhuge, C., C. Shao, S. Wang and Y. Hu (2019) Sensitivity analysis of integrated activity-based model: using MATSim as an example, *Transportation Letters*, **11** (2) 93–103.

A Appendix

Some of the plots were removed from the main text for brevity and are added here for reference.

Figure 22: Stationarities per aggregate volume for links

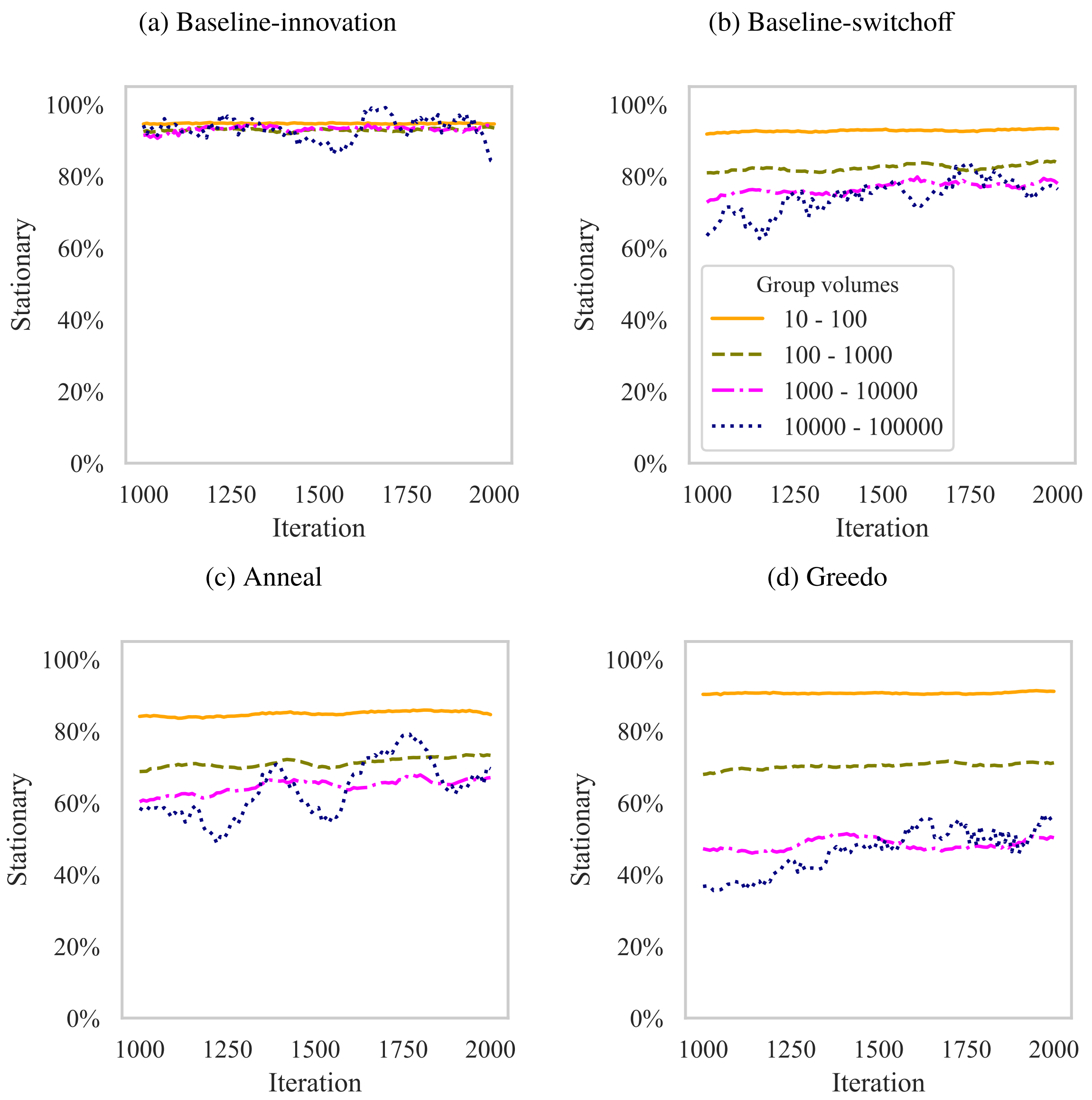


Figure 23: Ride mode shares