Aalto University
School of Science
Master's Programme in Mathematics and Operations Research

An activity-based model of travel demand using an open-source simulation framework

Master's Thesis
09.10.2017

Touko Väänänen

Aalto University
School of Science

Abstract of master's thesis

Author	Touko Väänänen
Title of thesis	An Activity-Based Simulation Model of Travel Demand using an open-source simulation framework.
Master's programme	Mathematics and Operations Research
Thesis supervisor	Professor Ahti Salo
Major or Minor/Code	Systems and Operations Research / SCI3055
Department	Department of Mathematics and Operations Research
Thesis advisor(s)	M.Sc. (Tech.) Taina Haapamäki

Date	**Number of pages**	**Language**
9.10.2017	82	English

Abstract

As self-driving cars, mobility as a service and other new transport services start to shape the future of travel, decision makers need to have accurate and well-justified information about the possibilities and consequences of their decisions. With the emerging travel modes, we face novel situations and therefore the forecasts of future travel should rest on theoretically and practically sound basis.

The notion that travel is a derived demand, is universally accepted. We rarely travel just for the sake of travelling, but to reach spatially separated activities which we ultimately aim to do. Despite understanding that activities are the drivers of our demand for travel, the models used for forecasting future travel demand ignore this fundamental attribute and concentrate on trip based approaches. Researchers have studied activity-based travel demand models for decades, but the use of these behaviourally more justifiable models has not gained much traction.

In this thesis, we construct an activity-based simulation model of travel demand using the MATSim simulation framework. MATSim's agent-based approach to travel demand simulation allows for a behaviourally justified model whose spatial resolution is greatly enhanced compared to the traditional models. The model is validated, and used to evaluate changes in travel behaviour induced by a large public transport infrastructure project, the new train line called Ring Rail line. The model performs fairly well in capturing the distance travelled distributions and aggregate measures of travel behaviour. In the case study, the model first predicts too low number of train passengers and then up to 600% increases in passengers on certain train stations after the changes in the public transport network.

The successes and possibilities of the model are discussed. Suggestions are given for the development of the model. It is discussed how to, on one hand, make the results more realistic and on the other hand, increase the model's capability to capture ever more subtle facets of people's travel behaviour.

Keywords agent-based, simulation, microsimulation, travel demand, discrete choice, open-source, activity-based

Aalto-yliopisto
Perustieteiden
korkeakoulu

Diplomityön tiivistelmä

Tekijä	Touko Väänänen
Työn nimi	Aktiviteettipohjainen liikennekysynnän malli avoimen lähdekoodin simulointiympäristössä
Koulutusohjelma	Matematiikka ja Operaatiotutkimus
Valvoja	Professori Ahti Salo
Pää tai sivuaine/koodi	Systeemi- ja Operaatiotutkimus / SCI3055
Työn ohjaaja(t)	Diplomi-insinööri Taina Haapamäki

Päivämäärä	**Sivumäärä**	**Kieli**
9.10.2017	82	Englanti

Tiivistelmä

Kun automaattiset ajoneuvot, liikkuminen palveluna sekä muut uudet liikkumisen palvelut alkavat muokata liikkumisen tulevaisuutta, päätöksentekoon tarvitaan tarkkaa ja hyvin perusteltua tietoa päätöksien mahdollisuuksista ja seurauksista. Kun uudet palvelut luovat täysin uudenlaisia toimintaympäristöjä, ennusteiden tulee olla mahdollisimman hyvällä pohjalla niin teoreettisesti kuin käytännöllisestikin.

Liikkuminen on johdettua kysyntää. Ihmiset kulkevat harvoin vain liikkumisen vuoksi vaan liikkumisella on päämäärä, jokin aktiviteetti, joka halutaan saavuttaa. Huolimatta siitä, että tämä liikkumisen erikoispiirre on tunnettu jo pitkään, käytetyt liikenteen ennustemallit eivät ota liikkumisen johdettua luonnetta huomioon vaan keskittyvät matkojen ennustamiseen pohjautuviin mallinnustapoihin. Aktiviteettipohjaisia liikkumiskysynnän malleja on tutkittu jo vuosikymmeniä, mutta näiden mallien mahdollisuudet ovat pitkälti hyödyntämättä palvelujen suunnittelussa ja poliittisessa päätöksenteossa.

Tässä diplomityössä rakennetaan aktiviteettipohjainen liikennekysynnän simulointimalli MATSim simulointiympäristössä. MATSimin agenttipohjainen lähestymistapa liikennekysynnän mallintamiseen mahdollistaa vakaalla teoreettisella pohjalla olevan mallin, jonka maantieteellinen tarkkuus on huomattavasti perinteisiä malleja parempi. Malli validoidaan ja sillä ennustetaan ison julkisen liikenteen infrastruktuuriprojektin, Kehäradan, liikennekäyttäytymisen muutoksia. Malli onnistuu kohtuullisesti matkan pituuden jakaumien, sekä laskettujen keskiarvomittareiden ennustamisessa. Case-tutkimuksessa malli ensin ennustaa liian alhaisia junaliikenteen käyttäjämääriä, minkä jälkeen jopa 600% matkustajamäärien kasvua tietyille juna-asemille Kehäradan avaamisen vaikutuksina.

Mallin onnistumisia ja mahdollisuuksia arvioidaan. Lisäksi tehdään ehdotuksia mallin kehittämiseksi. Kehityssuuntia pohditaan, jotta mallin tulokset olisivat totuudenmukaisempia ja jotta malli kykenisi mallintamaan yhä erilaisempia liikennekäyttäytymisen muotoja.

Avainsanat agenttipohjainen, simulointi, mikrosimulointi, liikennekysyntä, diskreetti valinta, avoin lähdekoodi, aktiviteettipohjainen

Contents

List of Abbreviations

FSM	Four step model
GTFS	General Transit Feed Specification
HLJ	Helsinki Region Transport System Plan
HSL	Helsinki Region Transport
JOSM	Java editor for Open Street Map
MATSim	Multi-agent transport simulation
MNL	Multinomial logit model
OSM	Open Street Map
TTB	Travel time bugdet
VTTS	Value of travel time savings
YKR	Monitoring system of the community structure

Chapter 1

Introduction

Travelling is a fundamental part of urban life and one of the basic sources of our wellbeing. As Hanson [2004] notes, transportation enables us to perform the spatially separated activities and tasks that make up our daily lives. Without the ease of movement people enjoy now, obtaining food, earning a living, getting medical care or visiting a friend, the basic things our lives and lifestyles depend on would become increasingly difficult.

Not only does transportation affect individual lives, but in larger scale transportation and its advancements also shape our cities and societies. Konishi [2000] argues that throughout history, cities have formed near rivers and coasts to serve as transportation hubs and interregional markets, because those places provided better access to other regions and thus smaller transportation costs. Moreover, an influential paper by Marchetti [1994] notes that the spatial size of cities has long been determined by the travel time required to get to the center of a city. As an example, there are no ancient cities with walls, with a diameter greater than 5 km, but when means of travel faster than walking are introduced, a city can grow in area. Marchetti [1994] gives Berlin as an example of this development. As new modes of travel were introduced to the city, first horse powered means, then electric trams, subways in 1925 and then cars in 1950, Berlin has grown in radius when the new means of travel have made faster relocation possible. The inspiration for

Marchetti [1994] was the preliminary work for a widely discussed [Mokhtarian and Chen, 2004] concept of travel time budget (TTB), originating from Zahavi and Talvitie [1980]. The idea of TTB, that people have a fairly stable time budget of a little over one hour for travel (Zahavi and Talvitie [1980], Schafer [2000], Vilhelmson [1999]), clearly suggests a travel speed based limitation for the geographic size of cities. If people are not ready to travel for more than a little over one hour per day, it is easily seen how a monocentric city cannot grow much larger than half an hour of travel from its commercial center.

Cities do not grow in area by themselves and the increased travel speeds merely enable their expansion. The real driver of the spatial growth of cities is the movement of people. The European Environment Agency [2017] has identified urbanisation as a megatrend that continues to shape the world. The United Nations has projected that 67% of the global population lives in urban areas by 2050 compared to 50% in 2010 and a mere 10%-15% in the early 19th century. Even though most urbanisation is estimated to happen in developing regions, this trend still continues in developed countries. Also in Finland the number of people living in city-like municipalities has been steadily rising since 1980, according to the urbanisation indicator of Findicator [2017], a joint project of Statistics Finland and the Finnish Prime Minister's Office.

The urbanisation process and growth of cities give rise to challenges especially in land use and transportation system planning. When more and more people live in cities, there is an increased need for housing, services, recreational facilities and locations for employment in these areas. Similarly, as people need to get access to the places they wish to visit, the demand for travel increases. Because especially car travel is currently based on fossil fuels, the increased demand for travel contributes to the megatrend of increasing pollution [European Environment Agency, 2017].

While cities continue to grow, at the same time transportation is facing an upheaval. Mobile phones and GPS services change the way to deliver useful traveller information, facilitate payment and charge for the use of transportation infrastructure [Ortúzar and Willumsen, 2011]. Electric vehicles, driverless cars and the possibilities of Internet and its effect on the sharing economy are beginning to shape the future of our movement. A report by a transportation research company Skift [Trivett and Staff, 2013] noted that trust in strangers, technology, cost consciousness and environmental concerns drive the boom of the sharing economy. In addition, they write that in the United States, car and driver's license ownership is declining and the younger generation is eager to participate in car sharing programs.

In the face of the myriad of changes and intertwined challenges in transportation, decision makers need high quality information about the consequences of their decisions regarding transportation. As a means to deliver accurate information for the decision maker, modelling plays an important role. For centuries, physical models in clay, wood or stone have been used to design equipment and infrastructure. Maps have been used as models of geographical space to navigate and plan routes. Even mathematical modelling has a long history, as priests in ancient times who could model the eclipse of the sun were revered in many civilizations [Hensher and Button, 2000]. Concerning transportation, mathematical models are used to understand and forecast various phenomena related to it. For example Ortúzar and Willumsen [2011] discuss models for trip generation, trip distribution, mode choice, traffic assignment, transport and freight demand, land use, car ownership, value of travel time, valuing external effects of transport, pricing and revenue. These mathematical models can be used to gain a better understanding of the systems, choices and consequences.

This thesis concentrates on travel demand models, especially activity based modelling, which has emerged from a better understanding of travel behaviour and the need to model ever more subtle changes in the transportation system. Traditionally travel demand modelling has been done with so called four step models (FSM) which model the travel demand in four separate and independent steps [Boyce, 2002]. This approach was first developed in the 1950s to forecast the effects of large infrastructure projects and has been celebrated for its successes at this task. However according to McNally [2000a], FSM has not been able to deliver good results in assessing the effects of relevant policy decisions. As the interest in transportation policy instruments such as road tolls and land-use planning grows, the need for travel demand models which capture the effects of such policies increases. Concerning the changing field of transportation modes, the effects and market shares of emerging travelling and mobility services and means such as shared or driverless cars, depend on the environment they operate in and the preferences of travellers. Therefore, to get a clear picture of the change in hand, a comprehensive model that takes into account a wide range of phenomena affecting travel demand is required. According to Bhat and Koppelman [2003] the inadequacy of the four step model for this task has led to the development of activity based models of travel demand. Rasouli and Timmermans [2014] write that academic work with these models began in the 1970s, but they have not yet been widely adopted by governments and other bodies involved in planning and developing the travel infrastructure.

In this thesis, an activity based model is developed for the Greater Helsinki region. The model is constructed using MATSim, an agent based simulation framework developed in several European universities [Horni et al., 2016]. The objective of the thesis is to explore the potential of activity based models to forecast travel behaviour adjustment induced by transportation infrastructure and system changes. The developed activity-based model is validated by comparing different travel statistics from the simulation with those of the Helsinki area travel survey [Transport, 2013]. As a case study, changes in the public transport system introduced by a large public transport infrastructure project in Helsinki, the Ring Rail line (Kehärata), are used to validate the forecasting potential of the model. With the model, we simulate a synthetic population's travel behaviour before the Ring Rail Line is opened and then after the changes have taken effect.

The thesis is structured as follows: In the first two chapters, the different frameworks of travel demand simulation are introduced with emphasis on the theoretical foundations of activity based modelling. In Chapter 3, the MATSim simulation framework is explained. Fourth Chapter describes models of discrete choice used in the model. Chapter 5 explains the data requirements of the activity based model and the way data is described for MATSim. Chapter 6 explains the development of the simulation model and Chapter 7 describes the technical details of the simulation runs. In Chapter 8, the results of the simulation runs are introduced. Chapter 9 analyses a case study with the model and the last chapter discusses the possibilities, shortcomings and future directions for the model.

Chapter 2

Simulation of travel demand

According to Banks et al. [2005], the act of simulation is mimicking a real life process or system. This process or system is studied by developing a simulation model, which usually is described as assumptions and rules concerning the behaviour and interaction of the system's parts. After the model is validated, it can be used to explore a wide variety of questions concerning the system's behaviour. The model can be used to predict the impact of, for example, changes in a system or studying the system before it is even built.

In some cases, it is possible to build a simple enough model which can be solved analytically. However many real world systems are so complex that it is not possible to derive an analytical solution for the models needed to study them. The models for these systems can then be simulated to imitate the system's behaviour over time. Data from the simulation can be collected as from the real world system, and these results can then be used to make inferences about the behaviour of the system.

Simulation can be a very useful tool in the field of transportation, as indicate by the number of papers concerning transportation in the Winter Simulation Conferences, a yearly conference on simulation organized by an association for operations research professionals, INFORMS. Shannon [1998] lists advantages of simulation, and in the field of transportation especially two of these advantages stand out:

1. New designs, layouts etc. can be tested without committing resources to their implementation.
2. New staffing policies, operating procedures, decision rules, organizational structures, information flows etc. can be explored without disrupting the ongoing policies.

Because many transportation systems are very complex and concern much of everyone's daily lives, one great benefit of simulating travel is that changes to the system can be analysed without the need to do costly modifications to the system in the real world.

Traditionally, travel demand modelling has been done with four step models, but already from the beginning of the 1970's, the activity based approach has been the prevalent way for travel demand forecasting in the research community.

2.1 The four step model

The traditional model for simulating travel demand is the so called four step model (FSM) of travel demand [Boyce, 2002]. Rasouli and Timmermans [2014] write that the theoretical basis for the four step model is in social physics, which postulates that, comparable to thermodynamics, the average behaviour of masses of individuals can be predicted with aggregate models even if the behaviour of one particular individual cannot be well represented. FSM is a trip based approach, in which the basic unit of analysis is a trip from one place to another. As the name suggests, FSM models travel demand via four steps, with different and independent models used in each step. McNally [2000a] introduces these four steps as: (i) Trip generation, (ii) Trip distribution, (iii) Mode choice and (iv) Route choice (Figure 2.1). These steps are modelled in succession, the output of the last step acting as input to the next one. Ortúzar and Willumsen [2011] note that the resulting travel time matrices obtained from the route choice step are likely to be different from the original ones, which means that there is a need for a feedback loop from the route choice step to trip generation. Yet according McNally [2000a], even when the feedback has been introduced, it has not been done in a consistent and convergent manner.

In trip generation, the number of trips made in the network is estimated. The time of day affects the amount of trips made and usually either morning or evening peak is modelled. The trips are usually divided into at least three categories: (i) the usual ones being home-based work trips, (ii) home-based other trips and (iii) non-home-based trips. The trip end points are modelled as productions (where the trip begins from) and attractions (where the trip is headed to). The trip productions and attractions can be modelled in zonal, household or individual level, household level being the most common in productions and zonal level in attractions. The productions and attractions are modelled independently and usually the productions are then scaled to match attractions due to attraction models being more reliable.

In trip distribution, the trip ends generated in the trip generation step are combined into trips. The trip distribution model can be viewed as a destination choice model. The result tells how many trips are made from one zone to another. This is based on the productions of the starting zone, attractions of the arrival zone and the travel impedance between the zones. Travel impedance is usually measured as travel cost consisting of time and money spent.

The third step, mode choice, considers the choice of transportation mode. It factors the trip distribution into mode specific trip tables. The mode choice models are usually estimated from choice samples and reflect the individuals' choice probabilities. Therefore these models are usually highly disaggregated and the results must be aggregated for different regions. Common, mode choice models include nested logit models which can reflect many trip-maker and performance variables such as travel time and travel cost.

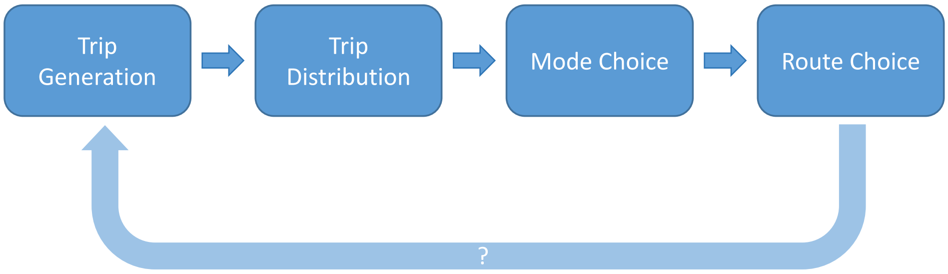

Figure 2.1: The four step model.

In the last step of the FSM, the performance of the route network is tested. The simulated travel demand is loaded on to the network, usually under the assumption of user equilibrium in which all paths utilized for a certain origin-destination pair have the same impedance. Typically, the user equilibrium is computed with the Frank-Wolfe algorithm.

In Finland, FSM has been used to forecast travel demand for the Helsinki region transport system plan [Transport, 2011] and for the whole Finland as a part of land use planning [Kalenoja et al., 2008], for instance. In addition, most cities in Finland have their proper FSM for traffic forecasting.

Although FSM is the most widely used model for forecasting travel demand, it has been criticised for decades. For example Kitamura [1996] identifies three main sources of problems in FSM: (i) lack of behavioural basis, (ii) lack of the time dimension and (iii) trip-based model structure. Rasouli and Timmermans [2014] also present some of the typical criticism of the FSM, arguing that most of it relates to (i). This is also seen in the summary of the weaknesses and limitations of FSM by McNally [2000b]. To verbatim:

1. “Ignorance of travel as a demand derived from activity participation decisions.
2. Focus on individual trips, ignoring the spatial and temporal interrelationship between all trips and activities comprising the individual

activity pattern.

3. Misrepresentation of overall behaviour as an outcome of a true choice process, rather than as defined by a range of complex constraints that delimit (or even define) choice.

4. Inadequate specification of the interrelationships between travel and activity participation and scheduling, including activity linkages and interpersonal constraints.

5. Misspecification of individual choice sets, resulting from the inability to establish distinct choice alternatives available to the decision-maker in a constrained environment.

6. The construction of models based strictly on the concept of utility maximization, neglecting substantial evidence relative to alternate decision strategies involving household dynamics, information levels, choice complexity, discontinuous specifications, and habit formation."

McNally [2000b] writes that these limitations contribute insensitivities to the four step model such that their performance in assessing the effects many transport related policy decisions is lacking. In the 1990s, increasing importance of travel demand management programmes brought up the need for more disaggregate models of travel demand that are more sensitive to such policy changes [Rasouli and Timmermans, 2014]. These needs and the criticisms of the FSM have given rise to the research of activity based models of travel demand.

2.2 Activity based models

To address the shortcomings of FSM, activity based models of travel demand were first studied in depth in the 1970s [McNally, 2000b]. According to Rasouli and Timmermans [2014], the approach began gaining attention at the latest in the 1990s when several pilot projects in the USA were funded.

The basis of the activity based approach stems from the observation that travel is a derived demand. One rarely travels just for the sake of travelling but rather, one needs to participate in activities such as work, shopping, hobbies or education, creating the need for travel. To participate in these activities, one needs to make trips which link one's day's activities together [Ortúzar and Willumsen, 2011]. This link between activity participation and travel was established already by Mitchell and Rapkin [1954].

The activity based approach is a less strict framework of travel demand modelling compared to FSM. Nevertheless, several themes characterize the activity based models all which try to address the problems of the conventional FSM [McNally, 2000b]. Specifically he notes that:

1. "Travel is derived from the demand for activity participation.
2. Sequences or patterns of behaviour, and not individual trips, are relevant unit of analysis.
3. Household and other social structures influence travel and activity behaviour.
4. Spatial, temporal, transportation and interpersonal interdependencies constrain activity-travel behaviour.
5. Activity based approaches reflect the scheduling of activities in time and space."

The basic unit of analysis is the travel-activity pattern, which represents the revealed pattern of behaviour expressed as the travel and activities over some period of time, usually one day. These activity patterns are referred to as household activity patterns and arise via the execution of household daily activity programs. Individual activity programs arise from the allocation of household activities for the members of the household via a decision process. Activity programs are taken as a plan of travel and activity participation which after scheduling, results in an activity pattern for an individual.

Rasouli and Timmermans [2014] provide a comprehensive review of the current state of activity based models. They divide the models into three categories: (i) constraints-based models, (ii) utility-maximizing models and (iii) computational process models. Out of these, the constraints-based models evolved first, the research beginning in the early 1970s. These models' functionality is based on evaluating whether a certain activity pattern can be executed within a space-time context. Even though these models contain some basic behavioural principles for adaptation to new policies, they are weak behaviourally compared to the other two types of models.

The utility maximizing models focus on individual preferences instead of time and space constraints. They are based on the premise that people maximize their utility when deciding on their activity schedules. These models are typically nested logit models predicting different facets of the activity schedule.

The third type of models, computational process models, try to relax the strict and unrealistic assumption of utility-maximizing behaviour and suggest rule-based activity schedule construction in order to mimic the decision heuristics of individuals.

Despite the promises of improving the deficiencies of FSM, the application of activity based models differs widely between countries and continents. According to Rasouli and Timmermans [2014], there is evidence of increasing adoption of activity based models in the US, but in Europe the application of activity based models has not gained traction. Asia, apart from South Korea, has not shown any interest in activity based models. In the European context, the Albatross model [Timmermans and Arentze, 2011] has been developed in the Netherlands.

2.3 Theoretical foundations of the activity based approach

One of the most widely accepted notions of travel is that it is a derived demand. The lives of individuals can be understood as a series of tasks that they perform such as work, shopping, playing or watching sports. These different tasks are called activities and the demand for travel arises from the need to perform these various spatially separated tassks (Bowman and Akiva [1996], Hanson [2004], McNally [2000b]). Ortúzar and Willumsen [2011] define an activity more precisely as:

> "[A] continuous interaction with the physical environment, a service or a person, within the same socio-spatial environment, which is relevant to the sample/observation unit. It includes any pure idle times before or during the activity (e.g. waiting at a doctor's office)."

The derived nature of travel demand is why it is necessary to consider travel decisions as a component in a broader framework of activity scheduling decisions and model the demand for activities. According to Bhat and Koppelman [2003], much of the current activity analysis is based on three seminal papers by Hägerstrand [1970], Chapin [1971] and Cullen and Godson [1975]. These three papers form a framework where activity patterns and their formation can be seen to surface from the interplay of different constraints. These constraints include constraints imposed by (i) individual motivations and societal constraints, (ii) spatial distribution of opportunities for activity participation and (iii) temporal considerations of an individual about activity participation, the two latter forming so called time-space prisms. It is also argued that the time-space prisms are mostly characterised by different rigidities in time and space, the rigidities of time being more dominant in the activity schedule formation. Another consideration for the activity schedule formation was given by Bowman and Akiva [1996], who note that especially

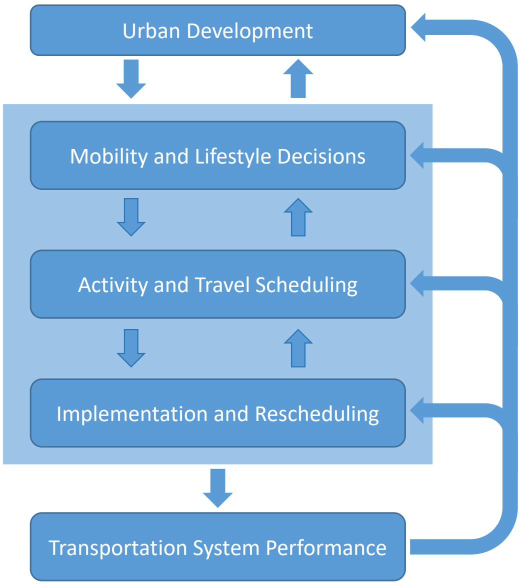

Figure 2.2: The activity and travel decision framework from Bowman and Akiva [1996], page 6.

with the increase of telecommunication, not all activities require travel and the activity model should also reflect this decision of travel versus non-travel activities.

Figure 2.2 shows the travel decision framework to which we mainly hold to in this thesis. This framework was introduced by Bowman and Akiva [1996]. Urban development decisions of governing bodies, firms and real estate developers influence the land use and transportation system, which affects the possibilities of households and individuals. The decisions of these individuals concerning their travel demand are divided into three categories. Further, the decisions have consequences in three different timeframes. The three decision categories are (i) mobility and lifestyle decisions, (ii) activity and travel scheduling decisions and (iii) implementation and rescheduling.

The decisions in the first category, that is mobility and lifestyle decisions, last generally longer and are made infrequently in a timeframe of years. These decisions include place of employment, residential location, car ownership and long term activity commitment decisions.

The next category of decisions are activity and travel scheduling decisions whose timeframe span from days to weeks. These decisions are made at more frequent intervals and include the choice and division of frequent and infrequent daily household activities such as grocery shopping, taking kids to school or visiting a bank for a loan. This category also includes decisions about personal activities such as going to work or gym. Furthermore, it includes the sequencing and timing of these activities. Mode choice decisions are also part of this category. For example, if one member of a household needs a car, this might affect the scheduling of activities for the other members of the household.

The last category of decisions made by the individual is implementation and rescheduling. This category has a time span of a few hours at most, and the decisions made in this category concern the execution of the activity schedule formed in the previous category. For example, these decisions include rerouting, lane change, accelerating, decelerating and travel speed decisions. Additionally, as the activity schedule unfolds, there may be some unforeseen circumstances that require changing or rescheduling the activities of the day. These include dropping an activity because of lack of time or changing the order of activities because a shop that was supposed to be open would not open until later.

The consequences of these decisions are ultimately embodied as transportation system performance when the individuals discharge their travel demand to the transportation system available. This performance manifests itself, for example, as travel times, travel volumes, congestion and environmental impact. The transportation system's performance then affects all of the aforementioned decision categories of individuals and organizations.

Because modelling travel demand is in many ways modelling complex decisions of individuals, knowledge of how people make these decisions is also crucial for the model. According to Bowman and Akiva [1996], every choice consist of three elements, (1) a set of alternatives, (2) a decision maker and (3) a decision protocol, meaning a set of rules with which the decision is made. The set of all feasible alternatives is usually called the universal set and the set of alternatives the decision maker actually consider is called the choice set. The alternatives in the choice set are defined as mutually exclusive and collectively exhaustive, meaning that the decision maker has to choose one and only one of the alternatives in the choice set.

As seen before, the decision making process concerning travel demand is very complex. The travel demand of an individual results from the choice of an activity schedule, which has many different dimensions in the decision process. Within any dimension, the number of alternatives can be very large, sometimes even infinite. Therefore in choosing an activity schedule, the individual faces a very large and complex set of alternatives.

The decision maker himself has limited resources in processing information. Information gathering takes time and energy, which are both limited. Combined with a large set of alternatives, the decision maker will be acting under incomplete information.

A variety of decision protocols may be chosen to make decisions. All of them can be described as a two stage process involving (1) choice set generation and (2) choice. In choice set generation, a set of alternatives are generated from the universal set. Then, an alternative is chosen from this set based on one criterion or several criteria. This process can be either deliberative or reactive. In a deliberate choice process, the choice set is generated from the universal set before any evaluation of the alternatives is done. The two stages of a decision process are conducted sequentially. In a reactive process, the evaluation of alternatives can lead to new alternatives being discovered, and the two stages of the process alternate until a final decision is made.

The choice set generation can be thought of as a search for alternatives and from a modelling perspective different methods vary in their style and rigor. The style of search can be random or structured. An exhaustive search finds all possible alternatives before making the choice and a non-exhaustive stops before all alternatives are found.

In the choice stage of the decision protocol, the alternatives are judged by one or more criteria and a choice is made based on a decision rule that takes the criteria into account. These rules employ unranked or ranked criteria, and compare these criteria between the alternatives. The decision rules are for example dominance, inferiority and satisfaction rules. Choice rules applying unranked criteria are not guaranteed to choose a single alternative, but rules applying ranked criteria can be formulated to always make an unambiguous choice. Decision rules also include composite rules in which the criteria are merged to a single metric describing the alternative. Then, usually the alternative with largest value of the metric is chosen.

Chapter 3

Multi-Agent Transport Simulation framework

Multi-Agent Transport Simulation (MATSim) is a Java based micro-simulation framework for activity based travel demand simulation [Horni et al., 2016]. It is open source, extendable and downloadable from the internet, for example Github. Rasouli and Timmermans [2014] do not classify MATSim to any of their three categories of activity based models, because it is based on a co-evolutionary principle and thus characterized by incremental improvement. Despite this, MATSim can be seen to have elements from both the utility maximizing and the computational process models.

In MATSim, every agent in the simulation iteratively optimizes its daily activity schedule while competing with other agents for space-time slots in the transportation network. The optimization is done with regard to a scoring function that can be interpreted as the agent's utility. The daily schedule is constructed heuristically based on the agent's experiences from previous iteration. Typical choice dimensions of the agents include route choice, mode choice, secondary activity location choice and departure time choice.

The MATSim iterative loop (figure 3.1) begins by loading the user generated initial demand to the transportation network. Each agent chooses a plan from its memory and executes it in the mobility simulation (mobsim). The plan is scored based on its performance in the simulation. Then some agents are allowed to change their plans following predefined innovation strategies and the loop begins anew. The loop is run until the average of the population scores stabilizes.

The transportation network is depicted as nodes and connecting links. Each link reflects a road segment and has a free flow speed, flow capacity telling how many cars may leave the link in certain time interval, and storage capacity telling how many cars fit into the link. The mobility simulation of MATSim uses a queue-based approach. When a car enters a network link, it is added to the end of a waiting queue. The car remains in the queue until the time it takes to reach the end of the link with free flow speed has passed, the car is in the front of the waiting queue and the next link allows entering. The approach is computationally efficient at the cost of some accuracy, because it does not capture some traffic phenomena such as overtaking or car following effects.

After the mobsim has run all agents' activity plans, the plans are scored based on a scoring function described in chapter 3.1.

In the replanning stage, part of the agent population is allowed to modify its activity plan. The old plan is copied and one aspect of the plan is changed. Currently MATSim supports changes to routes, leave times, leg modes and locations of secondary activities such as shopping. If the agent ends up with too many plans, the plans are deleted one at a time until there is a maximum allowed number of plans left. Plans can be deleted at random or based on the plan's score.

3.1 The scoring function

A crucial part of MATSim is the scoring function, which is used to evaluate the goodness of an agent's activity plan. The scoring function used in

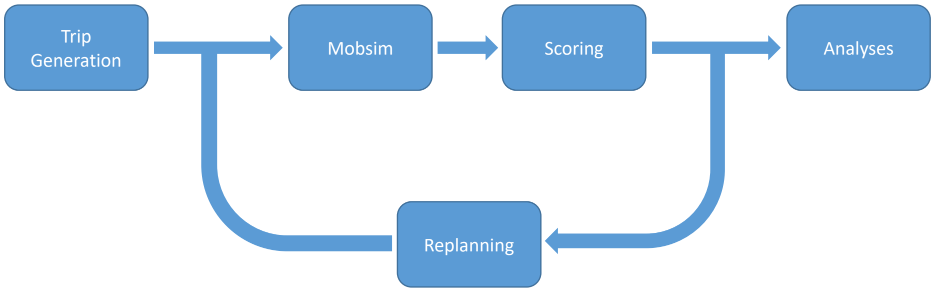

Figure 3.1: The MATSim iterative loop according to Horni et al. [2016], page 5.

MATSim is by Charypar and Nagel [2005] who describe a genetic algorithm for creating activity plans for agents. This function is loosely based on the Vickerey model of departure time choice (e.g. Arnott et al. [1993]). The scoring function is the sum of utilities of the activities performed, plus the sum of travel (dis)utilities

$$S(plan) = \sum_{i=1}^{n} S_{act}(i) + \sum_{j=1}^{n} S_{trav,mode(j)}(j), \tag{3.1}$$

where n is the number of activities, j is the trip that follows activity i, $S_{act}(i)$ is the score for doing activity i and $S_{trav,mode(j)}(j)$ is the (negative) score for trip j between activities i and $i+1$, done with mode $mode(j)$. For scoring, to produce a same number of trips and activities, the last activity is merged with the first activity.

The score of an activity depends on the time spent doing the activity, time spent waiting for the activity to start, penalty for arriving late to an activity, penalty for leaving too early from the activity and penalty for 'too short' activity. The expression for the activity's score is

$$S_{act}(i) = S_{dur,i} + S_{wait,i} + S_{late.ar,i} + S_{early.dp,i} + S_{short.dur,i}. \tag{3.2}$$

The individual components are defined as follows:

- The score for performing an activity for a certain amount of time

$$S_{dur,i} = \beta_{dur} * t_{typ,i} * ln(\frac{t_{dur,i}}{t_{0,i}}), \tag{3.3}$$

where $t_{dur,i}$ is the time spent doing activity i, β_{dur} is the marginal utility of time spent doing an activity (same for all activities), $t_{0,i}$ is the duration when utility starts being positive and $t_{typ,i}$ can be interpreted as a typical duration for activity i.

- The score for having to wait for an activity to start (e.g. because of shop is still closed)

$$S_{wait,i} = \beta_{wait} * t_{wait,i}, \tag{3.4}$$

where β_{wait} is the direct marginal utility of time spent waiting and $t_{wait,i}$ is the waiting time.

- The score (penalty) for arriving late to an activity is

$$S_{late.ar,i} = \begin{cases} \beta_{late.ar} * (t_{start,i} - t_{latest.ar,i}), & \text{if} \quad t_{end,i} > t_{earliest.dp,i} \\ 0, & \text{else,} \end{cases} \tag{3.5}$$

where $\beta_{late.ar}$ is the direct utility for arriving late, $t_{start,i}$ is the starting time of the activity i, $t_{latest.ar}$ is the latest penalty-free arrival time to the activity i (e.g. starting time of a play), $t_{end,i}$ is the ending time of the activity i and $t_{earliest.dp,i}$ is the earliest departure time from activity i.

- The penalty for leaving early from an activity is

$$S_{early.dp,i} = \begin{cases} \beta_{early.dp} * (t_{end,i} - t_{earliest.dp,i}), & \text{if} \quad t_{end,i} > t_{earliest.dp,i} \\ 0, & \text{else,} \end{cases} \tag{3.6}$$

where $\beta_{early.dp}$ is the marginal (dis)utility for leaving too early from an activity.

- The last expression telling the penalty for 'too short' an activity is

$$S_{short.dur,i} = \begin{cases} \beta_{short.dur} * (t_{short.dur,i} - t_{dur,i}), & \text{if} \quad t_{dur,i} < t_{short.dur,i} \\ 0, & \text{else,} \end{cases} \tag{3.7}$$

where $\beta_{short.dur}$ is the marginal utility for 'too short' activity and $t_{short.dur,i}$ is the shortest possible duration for activity i.

- The scoring function for travel (dis)utilities is defined as follows:

$$\begin{aligned} S_{trav,mode(j)}(j) = {} & C_{mode(j)} + \beta_{tt,mode(j)} * t_{trav,i} + \beta_m * \Delta m_j \\ & + (\beta_{d,mode(j)} + \beta_m * \gamma_{d,mode(j)}) * d_{trav,i} + \beta_{transfer} * x_{transfer,j}, \end{aligned} \tag{3.8}$$

where

- $C_{mode(j)}$ is a mode specific constant,
- $\beta_{tt,mode(j)}$ is the direct marginal utility of time spent travelling with mode $mode(j)$,
- $t_{trav,i}$ is the travel time between activities' i and $i+1$ locations,
- β_m is the marginal utility of money,
- Δm_j are the monetary costs of tolls and fares for the trip j,
- $\beta_{d,mode(j)}$ is the marginal utility of distance with mode $mode(j)$,

- $\gamma_{d,mode(j)}$ is the monetary distance rate of mode $mode(j)$,
- $d_{trav,i}$ is the distance travelled between activities' i and $i+1$ locations,
- $\beta_{transfer}$ is the penalty associated with public transport transfer,
- $x_{transfer,j}$ is a binary variable indicating whether there was a transfer between previous and current leg.

Horni et al. [2016] give some remarks about the scoring function. The penalties from Equation 3.8 describe the direct penalties from the travel. On top of these penalties, there is also the opportunity cost of not performing an activity while travelling. This opportunity cost comes on top of the penalty term $S_{wait,i}$ in the utility for performing activity i. Therefore reducing the travel time to some activity not only increases utility by reducing the penalties from travelling but also by increasing the amount of time spent performing some activity.

Another important remark about the scoring function is that in its current form, the scoring function cannot be used for dropping or adding of activities. This is because all activities give the same utility when performed for their typical duration. This leads to a situation where activities with smaller typical durations accumulate utility faster than those with larger typical durations. Therefore activities with long typical durations (e.g. home, work) get dropped first, which is not plausible. This limits the possible questions that can be studied with the MATSim framework, because the simulation cannot be used to forecast activity pattern changes based on changes, for example, in the road network or public transport network. Horni et al. [2016] offers a more extensive treatment of the scoring function.

3.2 Configuring the parameters

To realistically model the travel behaviour of the target population, the parameter values of the scoring function need to be estimated. Horni et al.

[2016] suggest a procedure for calibrating the parameters of the scoring function. The procedure is based on a mode choice model and the observation that most of the time, travelling is not less convenient than 'doing nothing' and therefore the direct marginal utility of travelling is close to zero. The calibration procedure is as follows:

1. Set β_m, marginal utility of money, to the prefactor of the monetary term in the mode choice model.
2. Set β_{dur}, marginal utility for performing an activity, to the prefactor of the car travel time in the mode choice model and change its sign from - to +.
3. Set $\beta_{tt,car}$, marginal utility of travelling by car, to 0.
4. Set all other marginal utilities of travelling by mode relative to the car value.
5. Set distance cost rates to some plausible negative value.
6. Set the alternative specific constants C_{mode} to calibrate the modal split.

Chapter 4

Discrete choice modelling

Discrete choice models are used in two stages of the construction of the simulation model. First, because in its current form the MATSim framework is not able to model workplace choice, a model is needed to define a place of occupation for the agents that are employed. Second, the parameters of the MATSim's scoring function are calibrated with the help of a mode choice model. Workplace and mode choice are both discrete in nature, because they involve a choice from a non-continuous set of alternatives.

According to Ortúzar and Willumsen [2011] the most common framework for discrete choice modelling is the random utility theory. The basic framework assumes that:

1. Individuals belong to a given homogeneous population Q, have rational preferences, and possess perfect information, therefore always choosing the alternative maximizing their utility.

2. There is a set $\mathbf{A} = \{A_1, A_2, ..., A_n\}$ of alternatives and a set $\mathbf{X}$ of vectors of observed attributes of the individuals and alternatives. An individual q has a set of attributes $\mathbf{x} \in \mathbf{X}$ and faces a choice set $\mathbf{A}(q) \in \mathbf{A}$.

3. For every individual q, each alternative A_j has a utility U_{jq} associated with it. The modeller does not observe all of the characteristics affecting the choice of the individual and therefore assumes that the utility

U_{jq} can be represented by two parts

1) a measurable part V_{jq}, which is a function of the measurable attributes $\mathbf{x}$,

2) a random part ϵ_{jq}, which represents the unobserved preferences or attributes of the choice.

Thus the modeller postulates that $U_{jq} = V_{jq} + \epsilon_{jq}$. This allows the modeller to explain two apparent 'irrationalities' that may present themselves in the data: (i) that two individuals with exactly the same preferences choose different alternatives, and (ii) that some individuals do not seem to choose the alternative that, from the modeller's viewpoint, maximizes their utility. For the specification of V_{jq}, a common choice is the linear form

$$V_{jq} = \sum_{k} \beta_{kj} x_{jkq} \tag{4.1}$$

where the parameters β are assumed to be constant for individuals but may vary across alternatives. Certain homogeneity is required from the sample population for this decomposition to hold. In particular, it is necessary that all individuals face the same alternatives and same constraints.

4. The individual always chooses the alternative that maximizes their utility. Given the representation of U_{iq}, this means that alternative A_i is chosen by individual q if and only if

$$U_{iq} > U_{jq}, \ \forall A_j \in \mathbf{A}(q) \tag{4.2}$$

so that

$$V_{iq} - V_{jq} > \epsilon_{jq} - \epsilon_{iq}. \tag{4.3}$$

Because the modeller cannot observe error terms, it is not possible to say with certainty whether (4.3) holds. Nevertheless, because the error terms are random variables from the point of view of the modeller, the probability of choosing alternative A_i is

$$P_{iq} = Prob\{\epsilon_{jq} < V_{iq} - V_{jq} + \epsilon_{iq}, \forall A_j \in \mathbf{A}(q)\}. \tag{4.4}$$

By using this knowledge, we can generate different model forms by fixing a certain distribution for the error terms ϵ.

4.1 Multinomial logit models

According to Ortúzar and Willumsen [2011] the simplest and most used discrete choice model is the multinomial logit model. This model specification is the result of assuming the error terms ϵ as independent and identically Gumbel distributed. The choice probabilities under this specification are

$$P_{iq} = \frac{\exp(\theta\, V_{iq})}{\sum_{A_j \in \mathbf{A}(q)} \exp(\theta\, V_{jq})}, \tag{4.5}$$

where the parameter θ is usually normalized to one, because it cannot be estimated separately from the parameters β.

Somewhat more complex form of the logit model is the conditional logit model. This form was proposed by McFadden with the idea of modelling the choice behaviour using the attributes of the alternatives. In a conditional logit model, the utility of choosing alternative i is

$$U_i = \alpha_i + \gamma z_i + \beta_i x_i + \epsilon, \tag{4.6}$$

where α_i is an alternative specific constant, z_i is a vector of alternative specific variables with generic coefficient γ, x_i is a vector of alternative specific variables with a vector of alternative specific coefficients β_i, and ϵ is an unobserved error term.

Once again, assuming that the error terms ϵ are independent and identically Gumbel distributed, it can be shown that the probability of choosing alternative i from the choice set of the individual $\mathbf{A}(q)$ is again of the form in (4.5).

Both model formulations satisfy the independence of irrelevant alternatives property. This means that the relative probabilities of alternatives i and j, P_{iq}/P_{jq}, should not change if new alternatives are introduced to the model. This requires some care in the choice of alternatives, as shown by the famous red bus/blue bus example by McFadden.

The coefficients for the logit models can be estimated with the maximum likelihood method from real choice data. In the data, for every choice, also the alternatives that were not chosen must be visible, which could prove to be a challenge for some data sets.

Chapter 5

Data requirements

An activity based model of travel demand requires a lot of data to realistically model travel in a certain geographic area. The simulation model consists of a synthetic population, a road network and a public transport network.

To make the model easily configurable for different settings and locations, the data requirements should be kept as small as possible. On the other hand, for the model to be realistic enough, the agents and their distribution across the simulated area should closely match the demographics of that area. In addition, because the model is activity based, some information about the travel and activity behaviour is required.

In Finland, the monitoring system of the community structure (yhdyskuntarakenteen seurantajärjestelmä, YKR), compiles commercial statistics of the community structure of Finland into a 250m x 250m grid. The statistics contains the following variables:

1. Population structure by sex and age group
2. Workforce by industry
3. Workplaces by industry
4. Floor area and number of buildings by usage class

5. Floor area and number of apartments
6. Length of commute by home and workplace by industry
7. Number of people in apartments, size of households and car ownership
8. Number of vacation buildings by year of construction
9. Commercial offices

Unfortunately, these statistics were not available for use in this thesis. Nevertheless, because the statistics are compiled from whole Finland and are relatively easily available, it sets a benchmark of the data to be used for the construction of the model.

The statistics that were used in place of the YKR were various statistics from Statistics Finland (Tilastokeskus) and from HLJ - Helsinki Region Transport System Plan (Helsingin seudun liikennejärjestelmäsuunnitelma). The data of Statistics Finland covers all of Finland and is therefore easily usable also for other geographical areas in Finland. The HLJ data on the other hand covers only the municipalities near Helsinki. The used statistics were:

1. Household-dwelling units by number of persons and type of building, year 2012 (Asuntokunnat koon ja asunnon talotyypin mukaan vuonna 2012)
2. Population by age, sex and family status, year 2012 (Väestö sukupuolen, iän ja perheaseman mukaan alueittain vuonna 2012)
3. Population by area, main type of activity, sex and age and year, year 2012 (Väestö alueen, pääasiallisen toiminnan, sukupuolen, iän ja vuoden mukaan, vuosi 2012)
4. HLJ forecast of the population and number of workplaces of Helsinki region in 2012 in a 250m x 250m grid

In addition, to model the travel behaviour of the agents, some kind of travel diary data is needed. This data usually contains information of mode choices which are required for the estimation of the mode choice model to calibrate the MATSim scoring function. The travel study of the Helsinki region (Helsingin seudun liikkumistutkimus) provides the data needed for this.

To model the performance of the transportation system, data from the infrastructure and services for travel is needed. The road network for cars was extracted from Open Street Map with the JOSM MATSim plugin. The whole road network is shown is figure 6.1 and a detail of it, the island of Lauttasaari, in figure 6.2.

Data is also needed of the public transport schedule. General Transit Feed Specification (GTFS) offers a good solution for depicting the public transport network. GTFS defines a common format for public transportation schedules and associated geographic information [Google, 2017]. It is developed by Google and major Finnish cities have their public transport schedule available in GTFS format. There is also a ready-made extension to MATSim, GTFS2MATSim, which generates the XML-files required by MATSim to simulate public transport from given GTFS data. As Helsinki Region Transport updates GTFS file of the public transport network whenever the schedule or routes change, this is also an easy way to maintain the model.

5.1 The MATSim data model

For MATSim to run the travel demand simulation, it needs a synthetic population, a road network, a public transport schedule, public transport vehicles, and facilities. These data are given for MATSim as XML-files of which each has its own structure. Another important file is the configuration file that binds all of these files together and allows the configuration of various other options in MATSim. These data files direct all aspects of the simulation.

The population file contains all information about the agents in the simulation. The file depicts the agents as XML objects that each contain the person id and the daily plan of the agent. The plan of the agent is given as a series of activities and legs between those activities, along with the activity type, one must specify a facility in which the activity is performed, coordinates for the facility and an end time for the activity. For the legs, only the travel mode is specified. If no route is specified for the leg, MATSim does the initial routing for the legs before the first mobsim begins. In addition to this mandatory data, one can specify for example the age, car availability, driver's license ownership and employment status for the agents.

The road network file consists of nodes and links of which each is its own XML object. The nodes must contain an id and coordinates of the node. The links, on the other hand, must contain the id of the link, the ids of the nodes the link starts from and ends to, the length, freespeed, capacity, number of lanes, whether the link is one- or two-way link and which modes are allowed to traverse the link. The length of the link is given in meters, freespeed in meters per second, and capacity in number of cars that can pass the link in an hour.

The public transport schedule file is the most complicated of the data files. The schedule defines all of the transit stops, which have the id, name and coordinates of the stop and the information if a vehicle that has stopped to take passengers blocks the traffic on the particular link. A public transport line is defined in several steps. First the line id is given. Next, a public transport line can contain several transit routes for which a transit mode is given. In addition to the mode, the transit route contains references to the stops it uses and the departures from the first stop on the route. The arrival times for each stop are given as times telling how long it takes for the transit vehicle to arrive to the stop from when it first departs for the route. After this, a transit itinerary is defined for the route. This is done as a sequence of links the vehicle traverses as it goes through the itinerary. The last part of defining a route is the departures, which tell at what time a transit vehicle departs for the route. All departures also contain an id for the departure and the id of the transit vehicle departing for the route.

The public transport vehicle file describes all vehicles used in public transport. The file consists of vehicle types, which describe all the vehicle types and the vehicles themselves. The vehicle type object contains the id and description of the vehicle type, how many seats there are in the vehicle and the length of the vehicle. The vehicle objects contain the id of the vehicle and the id for the type of the vehicle.

The facilities file contains information about the facilities where the agents conduct activities. In the facilities file, the coordinates and the type of activity conducted there must be defined for each facility. All homes and workplaces of the agents must also have their proper facilities. It is also possible to define certain opening hours for the individual facilities.

Lastly, the configuration file, called config file, is a long XML document containing a wide range of technical and simulation related parameters. For example the parameters of the scoring function, the innovation modules, modes in the simulation and the length of the simulation day are set in this file. The config file also defines the output directory and data files used. Config file is therefore indispensable for the simulation run. It is possible to generate a default config file containing all of the modifiable parameters with a Java command from the MATSim distribution. Examples of these files can be found in the MATSim's Github repository.

Chapter 6

The simulation model

The simulation model for travel demand consists of a synthetic population, a road network, a public transport network and facilities, where the activities are conducted. The study area of the model is the HLJ population forecast region which includes the Helsinki region and municipalities around it, totalling 29 municipalities. Private car, public transport and walking are modelled. Cars and public transport move on their respective networks and walking is modelled as a teleported mode. Public transport and cars have no traffic interaction between them. Parking is not modelled and there is no heavy traffic.

First, the synthetic population is constructed to match as closely as possible, within the limits of the available data, the demographics of the real population. Then, a place of occupation is simulated for all of the agents that are employed. For the agents that go to school, the closest education facility is chosen as their place of education. For other activities, the closest facility where the activity can be conducted is chosen. All in all, 1 575 544 agents are constructed for the simulation. The travel demand simulation is run with 10% of the total population.

The model simulates one working day, starting from midnight and ending at six o'clock the next morning, simulation time. During the travel demand simulation, agents execute their activity schedule, score it based on its performance and then change some aspect of it. After the modification of plans, called innovation, the activity plans are executed again. The simulation is run with 100 iterations with innovation modules disabled for the last 10 iterations.

6.1 Synthetic population

The creation of a synthetic population for the simulation involves matching the demographic statistics of municipalities from Statistics Finland to the number of people in each 250m x 250m square in the study area provided by Helsinki Region Transport System Plan.

The statistics from municipalities were directly used to generate distributions, from which different features were drawn for the agents. Because most of the statistics used are only available on a municipality level, the features of the population are aggregate on that level.

First, synthetic households are formed to each municipality according to the household sizes of each municipality. These synthetic households are then assigned to a certain square so that the number of people living in each square matches the amount in the HLJ grid data. Then, the number of adults and children living in each household is generated. The age and sex of the agents are generated according to the municipality level statistics of family size and type. Next, the main activity is drawn for an agent based on the municipality and age group of the agent. The main activities are school, work and other. For all agents under the age 18, school is assigned as the primary activity. The age groups are the same as in the division by Statistics Finland which is 0-17, 18-64 and 65-. After generating a primary activity, an activity plan is drawn for each agent over seven years old. For the generation of an activity plan, the synthetic population is divided into five different age groups: 6-17, 18-24, 25-40, 41-65 and 66-. In the HSL region travel study, the respondent could choose a trip end type from 21 different types. For the purposes of this thesis, the trip ends are aggregated into four different activity types: school, work, shopping and leisure. The aggregation is shown in Table 6.1. An activity plan is drawn for each agent from the activity patterns of the actual respondents to the travel study. The plan is drawn based on the agent's age conditional on that the activity plan includes the agent's primary activity.

Model trip ends	**HSL trip ends**
Home	Home Other domicile Summer cottage Hotel
Work	Place of employment Second place of employment Work errand
School	Own school
Shopping	Daily shopping Shopping mall Other shopping place
Leisure	Errand Restaurant Sports Culture Visit Other leisure
Other	Reclaim/Delivery Daycare Other Not available

Table 6.1: The aggregation of the trip ends from HSL travel survey for the purposes of this thesis.

6.2 Primary activity location choice

A primary activity location is chosen for each agent that has a work or school activity in its activity schedule. For the school activity, the closest education facility is chosen. This is a reasonable assumption for the younger students, because in Finland most children attend to the school closest to their home. Especially for university students though this assumption is too restrictive, because they have more freedom in their choice of domicile. This also leads to very different age groups to study in same facilities. However, since as age is not a variable examined for the results, it is unlikely for this to cause problems.

A work location is chosen for each employed agent by going through the 250m x 250m grid one square at a time and choosing a work square for each agent in that square. For the work location choice, a MNL model is used. For the MNL model, a utility function is needed to describe the different workplaces the agent may choose. To keep things simple but intuitive, the utility function chosen is decreasing in the travel time to a certain work square. Also the longer the travel time, the sharper the reduction in the utility. The chosen utility function is

$$V_{iq} = -t_{iq}^2, \tag{6.1}$$

where i is the index of the work square, and t_{iq} is the travel time to work from home of agent q. With this specification, we see from (4.5) that the probability of an agent to choose square i to work in is

$$P(worksquare = i) = \frac{\exp(-t_{iq}^2)}{\sum_{j \in \mathbf{A}(q)} \exp(-t_{jq}^2)} \tag{6.2}$$

where $\mathbf{A}(q)$ is the set of all squares in which there are workplaces left and to which the travel time from the agent's home square is under two hours. Limiting the travel time to work to two hours comes from the observation that people are not willing to travel to work that is too far away. In addition, a requirement for unemployment benefit in Finland is that one must accept work offered by employment office if the travel time to the workplace is up to one and half hours. A buffer of 30 minutes is added to get the limit used in the model.

To find out the travel times between the home squares and work squares, Open Trip Planner is used to route the center of the home square to the center of the work square.

After the work square is decided for all agents, exact coordinates for the workplace are chosen randomly from inside the square.

6.3 The car and public transport networks

The road network is extracted from Open Street Map (OSM) with the JOSM MATSim extension. Given a geographic area, the extension automatically downloads the necessary information from the OSM server and writes the road network in MATSim format. The final road network consists of 89 814 nodes and 189 164 links. The whole road network is depicted in figure 6.1 and a detail of it, the island of Lauttasaari, in figure 6.2.

The public transport network and timetables were extracted from GTFS data provided by Helsinki Region Transport. A MATSim extension called GTFS2MATSim was used to convert the GTFS data to MATSim format. The public transport network is extracted from the GTFS data of 16.4.2015. To simulate the situation after the opening of the Ring Rail line, GTFS data of 16.4.2016 is used. The public transport network is modelled on top of the car network, wherefore public transport and individual cars have no interaction between each other. This is a limitation to the model, because the effect of congestion to the public transport cannot be observed.

Walking mode is modelled as a teleported mode so that the length of a walking trip is 1.264 times the beeline distance between the starting and

Figure 6.1: The road network on which the micro-simulation is done.

arrival location, and the walking speed is 4.2 km/h, which is the assumed walking speed of Helsinki Region Transport's journey planner. The multiplier for the beeline distance was estimated from the travel study by comparing the beeline distances of the trips made by respondents to the distances of the same trips routed by Open Trip Planner. For some walking trips in the travel study the starting and ending locations had the same coordinates and these trips were excluded from the estimation.

6.4 Choice dimensions

Between the simulation iterations in MATSim, the agents can change parts of their daily schedules. According to Horni et al. [2016], typical choice dimensions include route choice, mode choice, departure time choice and secondary activity location choice. These are also the choice dimensions of the agents chosen for this thesis. After each iteration, and for each choice dimension, 10% of the population is chosen at random to change the corresponding di-

Figure 6.2: A detail of the road network, the island of Lauttasaari in Helsinki.

mension of their daily schedule.

The algorithms for the three first choice dimensions are fairly straightforward. Route choice is made by a route optimization algorithm taking into account the traffic flows from an earlier iteration. Mode choice is made by randomly choosing from the three possible modes - car, public transport and walking - and departure time choice is made by randomly changing each departure time by ± 30 minutes. Secondary activity location choice follows the more complex algorithm in Horni et al. [2012].

6.5 Facilities

Facilities are needed by the secondary activity location choice algorithm. Facilities are places where different activities are conducted. Each facility is associated with an activity that can be done there. The facilities were extracted from OSM, by downloading the coordinates of buildings by their tags. OSM uses building tags to describe the building's function and this system was used to assign an activity for each building. Because many activity places such as shops are open only during certain hours, MATSim also makes it possible to define the opening hours for the different facilities. Table 6.2 shows the division of tags into activities and the opening hours of the different activity types. The sport tag is an exception, because it is an OSM key instead. In the table, also work and home activities are shown, as they also need their proper facilities. The agents' home and work places were already defined earlier and these facilities were constructed based on the coordinates simulated before. The inclusion of facilities to the simulation also introduces the possibility of monitoring the use of different facilities during the day.

Activity	Opening hours	OSM tags
Education	7.00 - 20.00	school
Shopping	7.00 - 22.00	commercial shop
Leisure	6.00 - 2.00	sport public civic chapel
Other	Always	cabin kindergarten
Work	6.00 - 22.00	-
Home	Always	-

Table 6.2: Activities, their opening hours and associated Open Street Map tags.

6.6 Mode choice model for score function calibration

Horni et al. [2016] suggest that the parameters of the scoring function be calibrated with a mode choice model. After estimating the mode choice model, the scoring function's coefficients for monetary cost, transfers and travel time with different modes can be calibrated. Because mode choice is a discrete choice, it is typically modelled with logit models. The estimated mode choice model is a conditional logit model so that generic and alternative specific coefficients can be estimated. The models for discrete choice are treated in chapter 4.

In the mode choice model, the choice of a mode for a trip is explained by travel time, number of transfers and cost of travel. A generic coefficient is estimated for monetary cost and number of transfers. Mode specific coefficients are estimated for travel time with different modes. The model contains mode specific constants.

The data that is used for the estimation of the coefficients consists of travel diary data from the travel survey conducted by HSL in 2012. The data contains the chosen mode for the trip, but the time and distance reported in the data are only estimates by the respondent. The other possible modes that were not chosen are not reported in the data. We assume that possible alternatives for every trip are walking, public transportation and car. As the data contains the coordinates for the start and end points of the trip, Open Trip Planner is used to find the fastest route from the starting coordinates of the trip to the ending coordinates of the trip with the three modes used in the simulation model - car, public transport and walking. The distance, time and number of public transport transfers from the routing was used in the estimation of the mode choice model.

For estimating a coefficient for the monetary cost, the cost of a trip by car was calculated as

$$c_{car} = d_{car} * 0.1325 \text{ €}/km, \tag{6.3}$$

where d_{car} is the distance of the trip by car in kilometers and 0.1325 is an estimate for the variable costs (e.g. gas, wear of the car) of driving one kilometer by car given by Transport [2011].

The cost of a trip by public transport was based on the municipalities the trip began from and ended to. If both municipalities were within the HSL service area, the ownership of an HSL travel card was taken into account. If the traveller had a travel card and had bought unlimited trips for a certain period of time, the fare for the trip was the cost of 14 days of unlimited travel divided by 14∗2.0375, where 2.0375 is the average number of trips conducted by travel card holders in the HSL travel survey. If the traveller had a travel card but had not bought unlimited trips, the fare for the trip was the cost of a single ticket bought by travel card. If the traveller did not own a travel card, the fare for the trip was the price of a single ticket paid in cash. If one or both of the municipalities were outside the HSL service area, the fare of the trip was based on the distance of the trip and fares by Matkahuolto for the year 2013.

6.7 Simulation runs

A sample of 10% of the total population was used for the final simulation runs. The simulation was run with the public transport schedule of 16.4.2015 for a ‘before Ring Rail line’ scenario and the public transport schedule of 16.4.2016 was used for an ‘after Ring Rail line’ scenario. Both simulations ran for 100 iterations with the innovation modules disabled for the last 10 iterations.

For scenario testing, the runtime and hardware requirements of the model are important. If the simulation requires extensive hardware to run quickly, the widespread use of the model becomes burdensome. In this thesis the simulations were run on a server with a 2.40GHz 8 core Intel Atom processor and 12GB of RAM. The simulation was run with 8 threads.

The before simulation took 154 hours to complete and the after simulation took 93 hours. Most of the time was spent running the innovation modules while the mobility simulation itself took considerably less time. The runtime of the innovation modules varied between 4000 and 6400 seconds in the before simulation and between 1900 and 4000 seconds in the after simulation. The mobility simulations took between 200 and 400 seconds to run in both simulations. Thus reduction in the running time can be attributed to the shorter innovation module run times. The exact reason for the difference in innovation module runtimes is not known.

Chapter 7

Results

We evaluate how the simulation model performs in recreating the observed travel behaviour of people. Towards this end, different statistics were calculated from the output of the simulation and compared with those from the HSL travel survey of 2012 [Transport, 2013]. The validation was done separately for the work location choice model and the travel demand model. The plausibility of the coefficients of the mode choice model are evaluated by comparing them to the results of value of travel time savings (VTTS) studies.

7.1 Work location choice model

For the work location choice model, beeline distances between home and work locations were examined. The simulated and real beeline distances between homes and workplaces seem to match quite well in the municipalities that participated in the survey. The simulated average beeline distance between home and workplace is 15.489 km and the average beeline distance calculated from the HSL travel study is 15.08 km. The average beeline distances between work and home locations by municipality are shown in Figure 7.1. If the

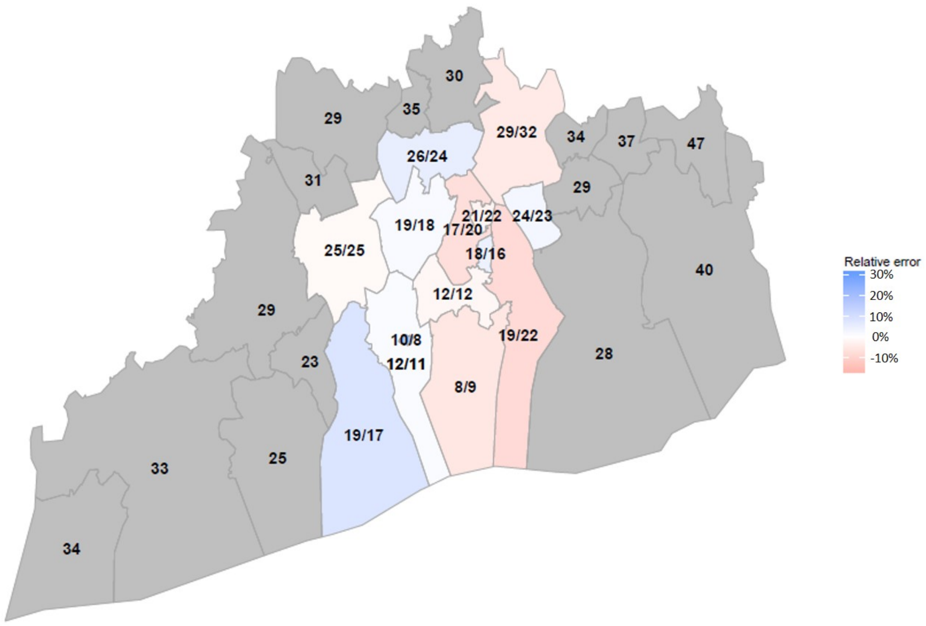

Figure 7.1: The average beeline distance of the work trip in each of the travel survey municipalities (simulated/survey). The color of the municipality corresponds to the relative error and grey means there was no empirical distance available.

municipality did not participate in the survey, only the simulated distance is shown. In the figure, blue color indicates on average too long simulated distances compared to the travel survey and red color indicates that, on average, the distances are too short.

Figure 8 shows that the home-work distances in the municipalities that participated in the travel survey match fairly well the simulated distances. However, in the fringes of the Helsinki working area some distances seem unreasonably high. The very long distances might be attributed to the fact that, in reality, some people in the border municipalities work outside of the Helsinki work area, and the work place assignment model does not allow agents to commute outside this region. Also in the simulation, some of the working squares filled up very quickly. This led to some choice set problems and the lack of choice in the model may also skew the results towards too long work trips.

7.2 The mode choice model and parameters of the scoring function

The mode choice model is used to calibrate the coefficients for the scoring function of MATSim.

The estimated coefficients of the mode choice model, their standard errors and p-values are in Table 7.1. All coefficients are statistically significant in 1% confidence level. The estimated coefficients imply a VTTS of 21.86 €/h by car, 11.13 €/h by public transport and 14.52 €/h by walking. A meta-study of VTTS [Shires and de Jong, 2009] examined VTTS studies from 30 countries, estimated regression models to explain the values and used them to calculate the VTTS by trip purpose for the respective countries. From this study the VTTS by car in Finland varied from 8.18 €/h to 28.54 €/h, the average between the purposes being 15.67 €/h. The VTTS by bus varied from 5.77 €/h to 22.91€/h, the average between the trip purposes being 11.13 €/h. VTTS studies of walking were not found in the Finnish context, but a British study by Wardman et al. [2000] found the VTTS of walking in the UK to be 17.67 pence/minute, resulting in 16.06 €/h with the average exchange rate of 1999, the year for which the study's values are based on. This value can be considered quite comparable with Finland based on the VTTS estimated for Finland and the UK in Shires and de Jong [2009].

Coefficient	Estimate	Std. Error	p-value
cost	-0.223	0.015	$< 2.2 \cdot 10^{-16}$
transfer	-0.288	0.054	$1.059 \cdot 10^{-7}$
car time	-4.875	0.190	$< 2.2 \cdot 10^{-16}$
pt time	-2.483	0.105	$< 2.2 \cdot 10^{-16}$
walk time	-3.238	0.071	$< 2.2 \cdot 10^{-16}$

Table 7.1: The estimated coefficients of the mode choice model. The p-values are calculated by the R package 'mlogit' which used in estimating the MNL model.

Coefficient	Value
$C_{mode,car}$	-3.5
$C_{mode,transit}$	2.5
$C_{mode,walk}$	-2.5
β_{dur}	2.483
$\beta_{tt,mode(car)}$	-2.392
$\beta_{tt,mode(transit)}$	0.0
$\beta_{tt,mode(walk)}$	-0.755
β_m	0.223
$\beta_{transfer}$	-0.288

Table 7.2: The coefficients used in the scoring function (3.2) for the final simulation runs.

Comparing the estimated coefficient and the VTTS found in studies, the estimated coefficients seem plausible. The VTTS of car and public transport are within the range of the values found in Shires and de Jong [2009] and the VTTS of walking is 10% smaller than the one found in Wardman et al. [2000].

From the coefficients of the model, the coefficients for the scoring function are calculated based on the calibration process in section 3.1. Because public transport possesses the lowest value among the coefficients for the travel time, unlike in the process, the travelling disutility of public transport is set to zero. This is done in order to eliminate positive travelling disutilities. To calibrate the constants of the different travel modes, the simulation was run with the public transport schedule for 16.4.2015. To speed up the simulations, a sample of one percent of the total population was used and the simulation was run for 100 iterations. After each run, the modal split of the trips was observed. For the next simulation run, the coefficients were changed in a direction to make the modal split closer to the one in the travel survey of HSL. This was repeated until the number of trips made with each of the modes differed less than three percentage points from the modal split in the travel survey. The final coefficients for the scoring function are shown in Table 7.2.

In addition to these parameters, the ones for the typical durations of activities must be configured. The typical activity durations are set as home 13 hours, work 8 hours, education 6 hours, leisure 3 hours, shopping 1 hour and other 3 hours.

7.3 Travel demand simulation

Some aggregate statistics from the simulation are shown accompanied with the same statistics calculated from the travel survey are shown in Table 7.3. All statistics are averages and, as in the travel survey, only trips under 100km are taken into account when calculating the time and distance averages. The percentage of trips made by walking includes the bike trips.

Statistic	Model	Survey	Difference
Trips by car (%)	52.0	42.5	9.5
Trips by walking (%)	28.7	33.7	−5.0
Trips by public transport (%)	19.3	23.8	−4.5
Time spent travelling (min)	86.0	73.0	13.0
Distance travelled (km)	35.6	28.0	7.6
Length of trip(km)	9.8	8.4	1.4
Length of trip by car (km)	16.9	11.7	5.2
Length of trip by walking (km)	1.5	1.4	0.1
Length of trip by transit (km)	8.3	11.5	−3.2

Table 7.3: Different aggregate statistics from the output of the simulation model and travel survey. The percentage of trips by walking includes trips by bike.

Table 7.3 shows that most of the statistics derived by the model are fairly close to the ones observed in the travel survey. Nevertheless, the relative differences between the simulation and the survey are fairly large. Besides, averages are not very descriptive of the underlying distribution and, because the area for the simulation is big and spatially diverse, there can be large differences among different spatial regions. Therefore, also the distributions behind these statistics are studied.

The HSL travel survey gives some statistics split into two regions, the capital region and rest of the municipalities. The capital region consists of four municipalities, Helsinki, Vantaa, Espoo and Kauniainen. Table 7.4 shows the statistics of Table 7.3 split into these two spatial regions.

Statistic	Model, capital	Survey, capital	Model, rest	Survey, rest
Trips by car (%)	42.3	37.7	66.8	59.3
Trips by walking (%)	31.7	34.1	25.1	32.5
Trips by public transport (%)	26.0	28.1	8.1	8.1
Time spent travelling (min)	66.0	73.0	106.0	74.0
Distance travelled (km)	23.7	25.0	48.8	41.0

Table 7.4: Different aggregate statistics from the simulation model and travel survey in the capital region and rest of the municipalities. The percentage of trips by walking includes trips by bike.

Looking at these statistics from the split into two regions, it appears that the simulation model works better for the capital region than for the other municipalities. Exception for “Trips by public transport”, all the statistics calculated from the results of the simulation are closer to the survey in the capital region than in the other municipalities.

Mode choice distributions

Figure 7.2 shows the percentage of trips starting from each of the travel survey municipalities that are made by car. From the figure, it is apparent that the model estimates too many trips by car in almost every municipality. The largest discrepancies between the model and the travel survey are in Helsinki, Vantaa and Pornainen. The differences in Helsinki and Vantaa could be explained by the fact that the model does not account for the time needed to find a parking space at the end of a trip. This should increase the number of car trips made in municipalities where there is a lack of parking space in reality, like Helsinki.

Figure 7.3 shows the percentage of trips made by walking in each of the travel survey municipalities. Here, the results of the model compared to the travel survey vary. Municipalities with too many walking trips include Kauniainen, Mäntsälä and Vihti. On the other hand municipalities with too few walking trips include Pornainen, Sipoo and Kirkkonummi. There are also many municipalities where the simulated number of walking trips matches very well to the travel survey.

The percentage of public transport trips made in each travel survey municipality is shown in figure 7.4. The model suggests too few public transport trips in almost every municipality. The difference compared to the travel survey varies from about 9 percentage points too few trips to about 1.5 percentage points too many. This discrepancy could be explained in some municipalities by the too high number of car trips, but another possible explanation is that the spatial distribution of home and work locations does not match reality well. Because the work location choice model only accounts for the travel time by car, municipalities with good public transport connections can have agents going to work in different places than in reality. This would reduce the number of public transport trips made in municipalities like Mäntsälä, Riihimäki, Järvenpää and Hyvinkää which have good train connections to other municipalities.

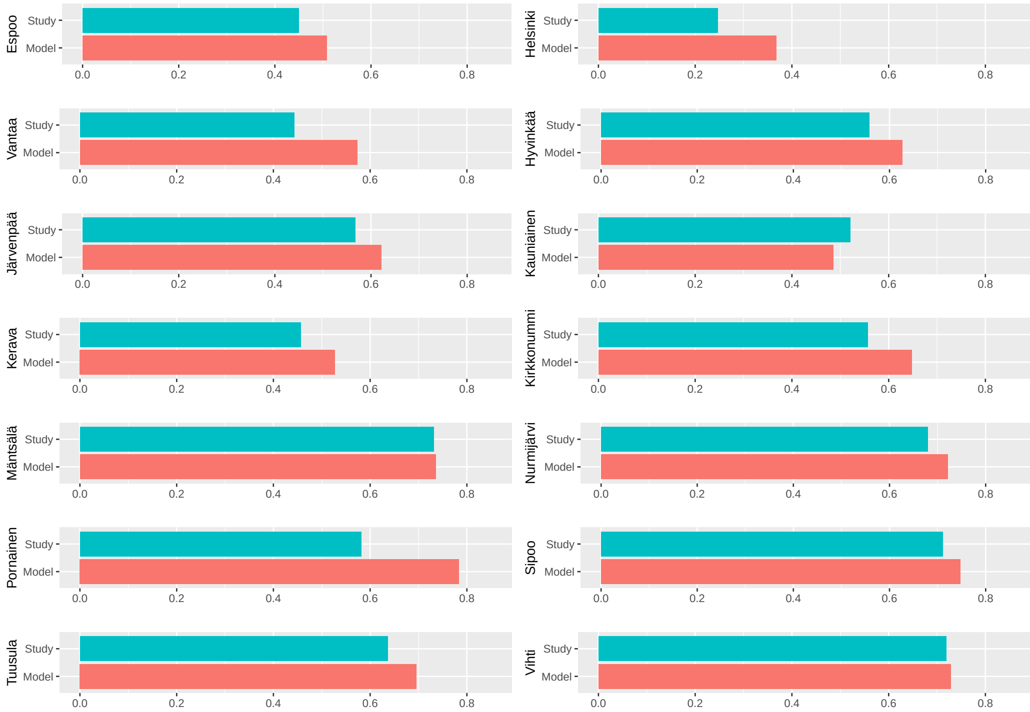

Figure 7.2: Percentage of trips made by car in each of the travel survey municipalities.

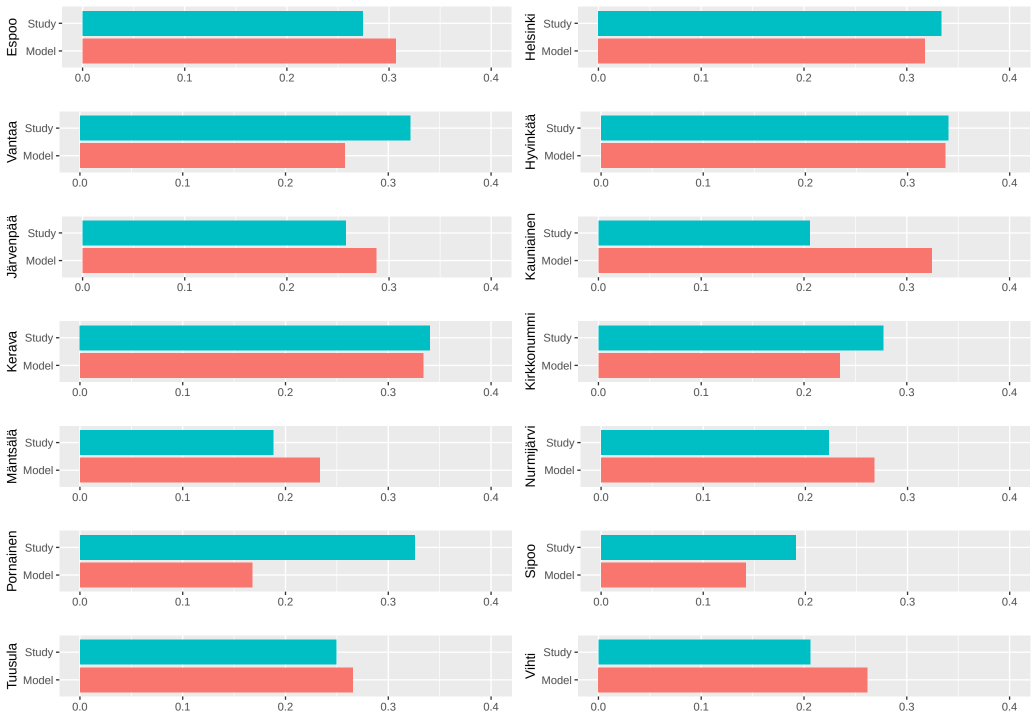

Figure 7.3: Percentage of walking trips in each of the travel survey municipalities.

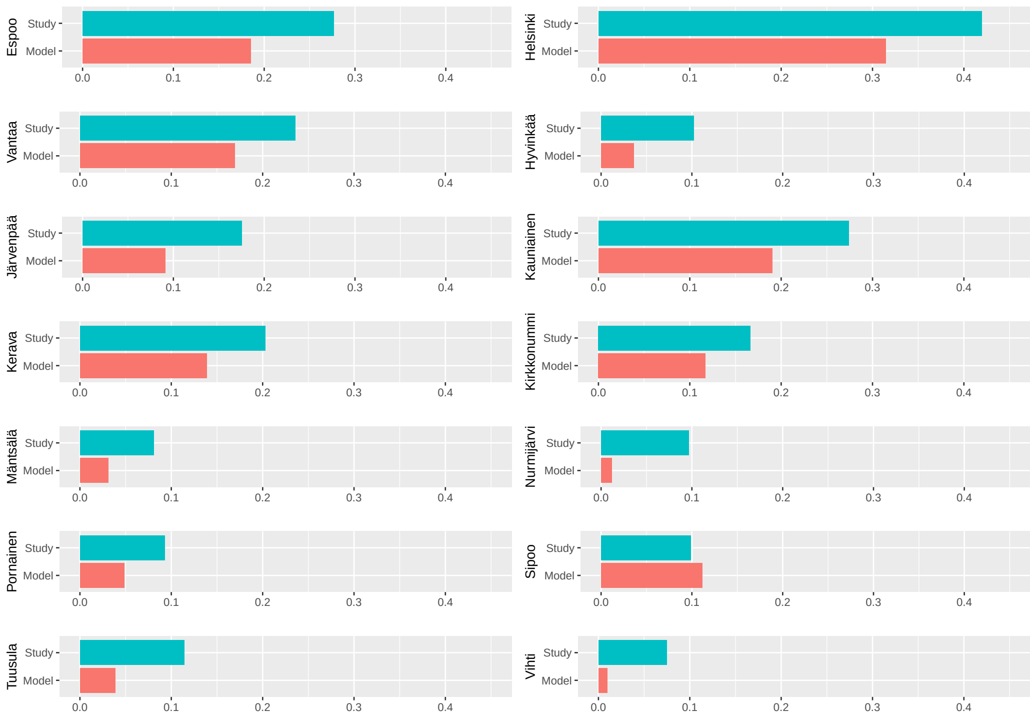

Figure 7.4: Percentage of trips made by public transport in each of the travel survey municipalities.

Length of trips

Figure 7.5 shows the length of trips distributions made by different modes. In general, the simulated trips are too short. All modes show an overrepresentation of trips that are shorter than 5km. This can be seen especially in the public transport trips. In the simulated car trips, there are too many trips that are over 45km long. Nevertheless, when looking at the combined distribution of trips lengths, it matches the study fairly well. This suggests that the spatial choices of destination of the agents match those of real people, but the mode choices do not represent the true ones.

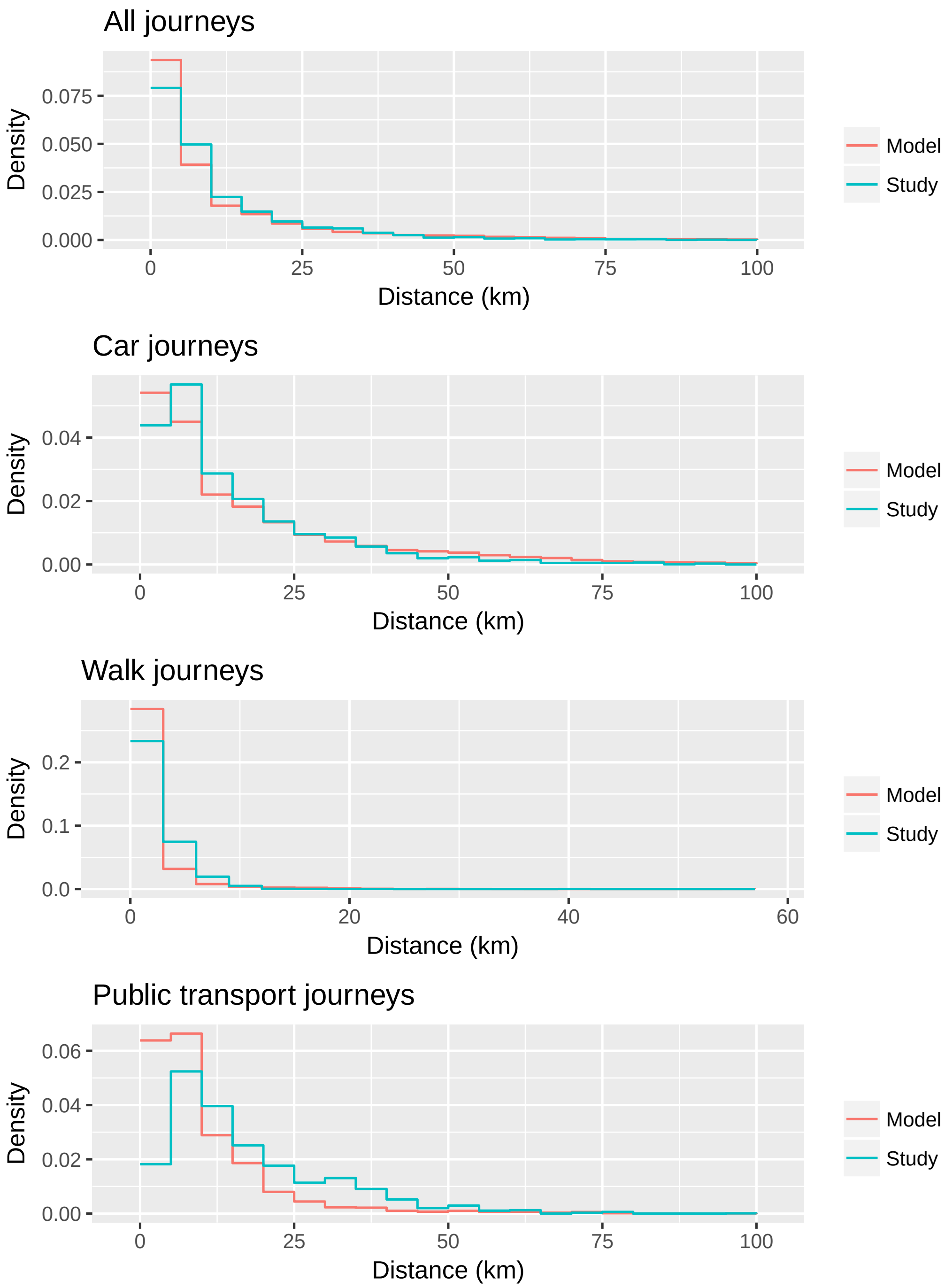

Figure 7.5: Distributions of length of trips by different travel modes.

Chapter 8

Case: The Ring Rail line

For a case study, the effects of Ring Rail line on the behaviour of train travellers are studied. The Ring Rail line is a large infrastructure project costing €774 million and completed in autumn 2015. With the Ring Rail line, five new train stations opened, one of which was a station for the Helsinki-Vantaa airport. The Ring Rail line connected two previous railways going in different directions, the new train connections starting and terminating in the Helsinki main railway station, hence the name Ring Rail line. The opening of the Ring Rail line meant also an overhaul of the Vantaa bus lines. The objective of the new design was to make the bus connections in Vantaa work better in conjunction with the newly opened train stations. The Ring Rail line was opened together with the new bus timetables in August 1st 2015.

To forecast the travel behaviour changes induced by the Ring Rail line, the simulation was run again with the exact same setting as before, but the public transport network was changed for the one extracted from the GTFS data of 16.4.2016. This day was a Thursday. The Ring rail line along with the rest of the public transport network changes had been implemented by this date.

Table 8.1 shows the number of passengers (boardings and alightings) in each of the train stops along the Ring Rail line before its opening. The number of passengers are according to the simulation and according to a study estimating the number of passengers in the train stops before and after the Ring Rail line. Figure 8.1 shows the number of passengers according to the study and figure 8.2 according to the simulation.

In view of Figure 8.1, Helsinki and Pasila are clear hubs for the train passengers to get on and off the train. The next busiest stations are Huopalahti, Malmi and Tikkurila. The results from the simulation show also that Helsinki and Pasila are big hubs for the passengers and Huopalahti, Tikkurila and Malmi are the next three busiest stations. Unfortunately despite this encouraging qualitative result, the exact number of passengers according to the simulation differ quite a lot from the ones in the study. From Table 8.1 we see that the relative errors of the number of passengers vary from -62% to 22% and the total relative error in the number of passengers in these stations is -47.2%. Figure 8.3 shows the absolute and relative errors of the simulation and the study.

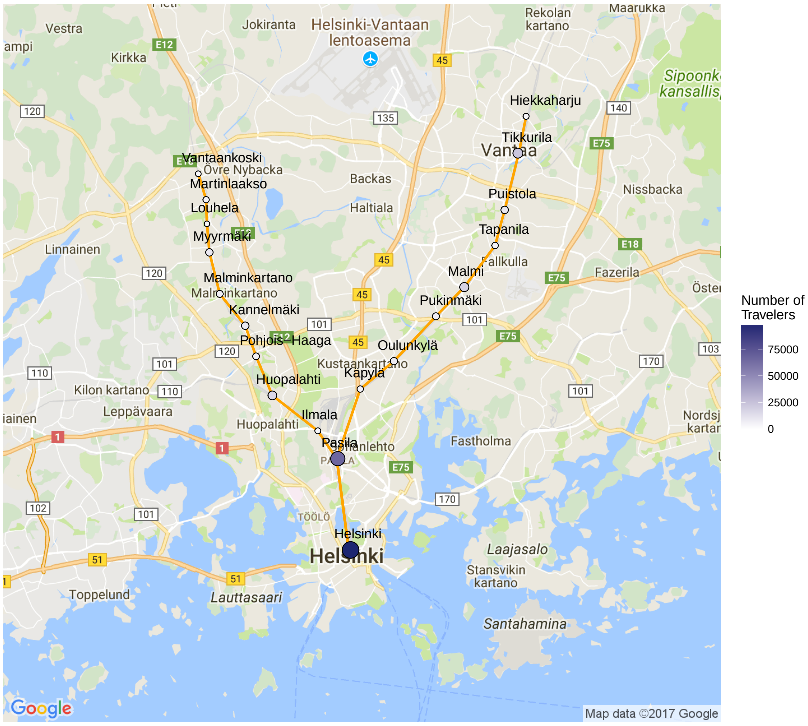

Figure 8.1: Number of boardings and alightings at the train stops along Ring Rail line before the completion of the Ring Rail line according to study.

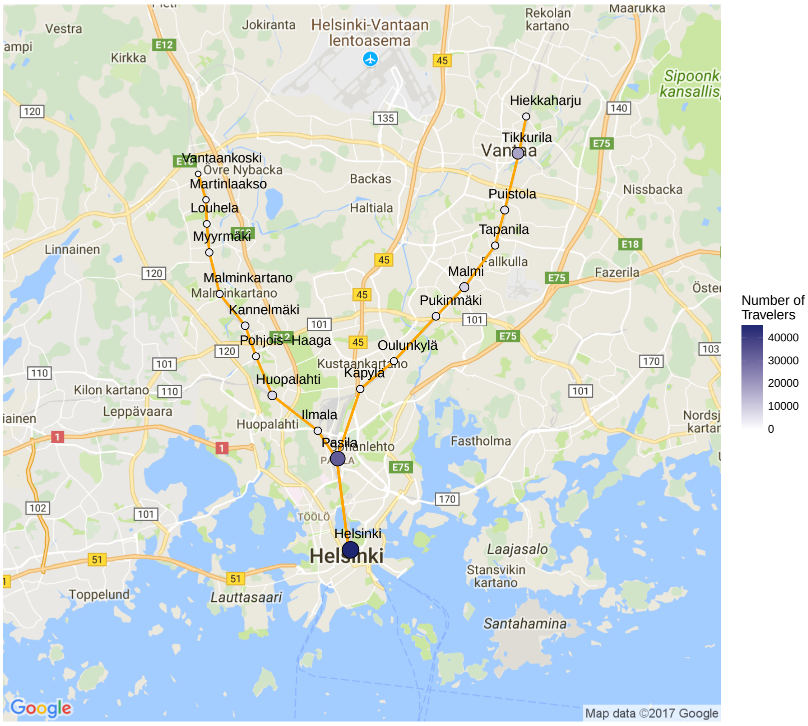

Figure 8.2: Number of boardings and alightings in the train stops along Ring Rail before the completion of the Ring Rail line according to the simulation.

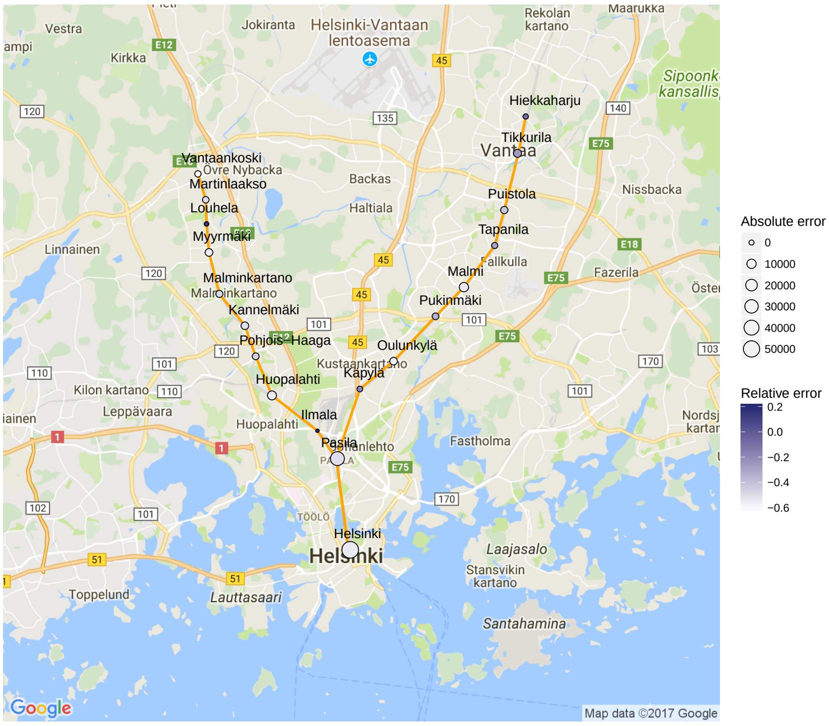

Figure 8.3: The error in the number of boardings and alightings of the simulation compared to the study.

Station	Travellers, model	Travellers, study	Abs. Error	Rel. Error
Helsinki	44910	98395	−53485	−54%
Pasila	32440	65777	−33337	−51%
Ilmala	4160	3395	765	22%
Huopalahti	7090	15755	−8665	−55%
Pohjois-Haaga	3240	6034	−2794	−46%
Kannelmäki	4440	9072	−4632	−51%
Malminkartano	3480	6763	−3283	−49%
Myyrmäki	3990	8882	−4832	−55%
Louhela	2820	2438	382	15%
Martinlaakso	2220	4381	−2161	−49%
Vantaankoski	1090	2880	−1790	−62%
Hiekkaharju	3370	3689	−319	−9%
Tikkurila	17450	23275	−5825	−25%
Puistola	5320	9104	−3784	−42%
Tapanila	3620	5064	−1444	−29%
Malmi	8280	18187	−9907	−54%
Pukinmäki	4600	7086	−2486	−35%
Oulunkylä	3580	8174	−4594	−56%
Käpylä	3880	4754	−874	−18%

Table 8.1: The number of travellers per station before the opening of Ring rail according to the simulation and study.

Table 8.2 shows the number of passengers boarding and alighting at each of the Ring Rail line stations according to the simulation and the study after the opening of the Ring Rail line. Figures 8.4 and 8.5 show the changes in the number of passengers in each of the Ring Rail line stations. The simulation suggests that the largest relative changes to passengers take place in Hiekkaharju, Vantaankoski, Oulunkylä and Tikkurila. The study shows that the largest relative changes in the number of passengers occurred in Hiekkaharju, Martinlaakso, Louhela and Myyrmäki. In general the model does not capture the qualitative nor quantitative changes in the number of passengers very well. The model also predicts more passengers would use the five new stations that opened with the Ring Rail line than what was the case.

Station	Absolute, model	Absolute, study	Relative, model	Relative, study
Helsinki	30980	7455	70.0%	7.5%
Pasila	48400	−1593	149%	−2.4%
Ilmala	7030	7	169%	0.2%
Huopalahti	15480	1682	218%	10.7%
Pohjois-Haaga	4820	−131	149%	−2.2%
Kannelmäki	5370	19	121%	0.2%
Malminkartano	3730	693	107%	10.2%
Myyrmäki	7980	3025	200%	34.1%
Louhela	3780	859	134%	35.2%
Martinlaakso	5160	1878	232%	42.9%
Vantaankoski	5370	−99	493%	−3.4%
Hiekkaharju	21980	2752	652%	74.6%
Tikkurila	48810	5215	280%	22.4%
Puistola	9520	1319	179%	14.5%
Tapanila	83700	−90	131%	−1.8%
Malmi	4750	2433	184%	13.4%
Pukinmäki	15240	110	106%	1.6%
Oulunkylä	4900	997	371%	12.2%
Käpylä	4930	219	127%	4.6%
Vehkala	1200	644	NA	NA
Kivistö	5290	3533	NA	NA
Aviapolis	6220	2285	NA	NA
Lentoasema	11950	5275	NA	NA
Leinelä	8000	2785	NA	NA

Table 8.2: The change in the number of travellers after the opening of Ring rail comparing the simulations together and according a study.

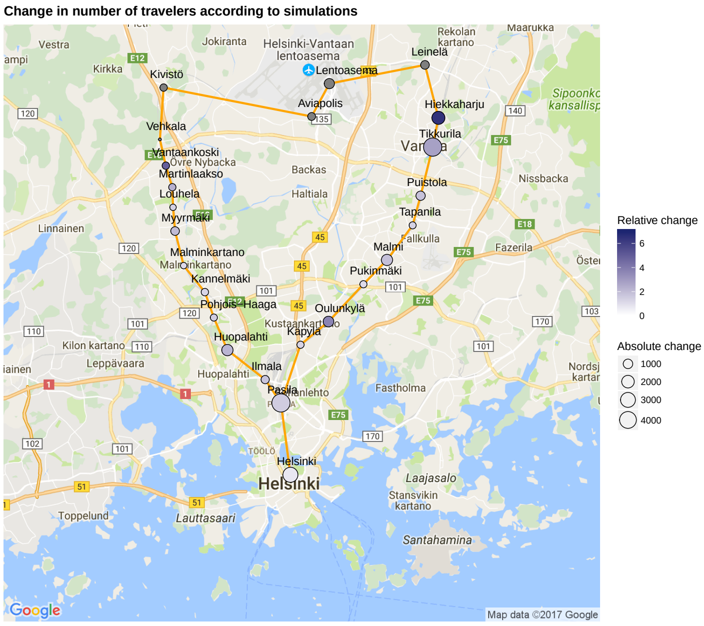

Figure 8.4: Change in the boardings and alightings along the train stops of the Ring rail according to the simulations.

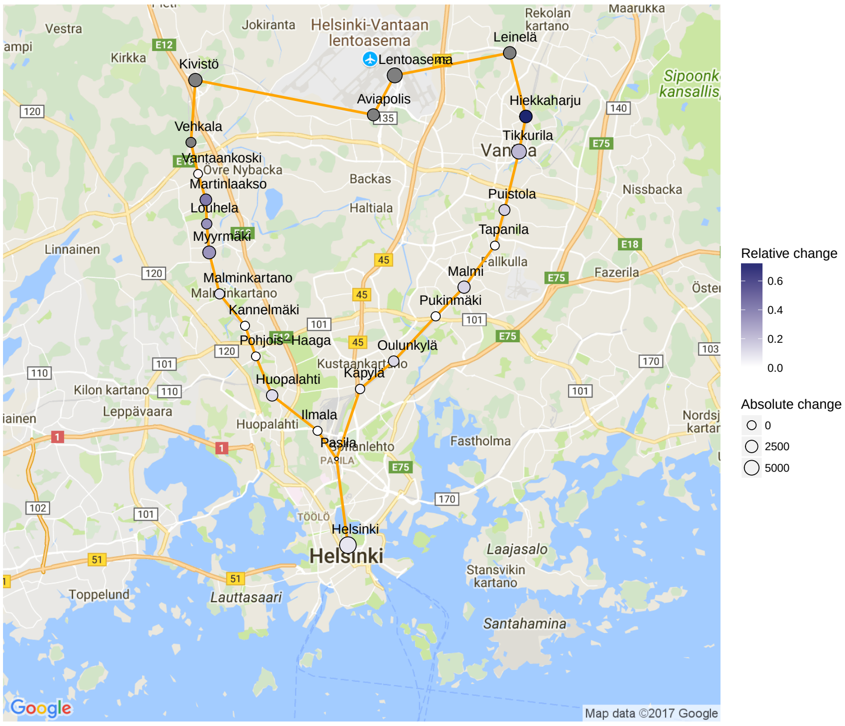

Figure 8.5: Change in the boardings and alightings along the train stops of the Ring rail according to study.

Chapter 9

Discussion and Conclusions

In this thesis, we have constructed an activity based travel demand simulation model. The model is validated and used to forecast travel behaviour changes after a large public transport infrastructure project. To validate the results, different aggregate statistics are calculated (Table 7.3). The statistics match the ones from the Helsinki Region Transport's travel survey [Transport, 2013] fairly well. Regarding the model's modal split, car is chosen too often compared to the other modes. This can be seen from Figure 7.2, in which almost every municipality too many trips are made by car. The model predicts too few car trips only for the municipality of Kauniainen.

Comparing the study and the model, large differences in the popularity of car can be seen in municipalities of Helsinki, Vantaa and Pornainen. At least for Helsinki, and probably also for the neighbouring municipality of Vantaa, the abundance of car trips can be partly attributed to the fact that parking is not modelled in the simulation. Because downtown Helsinki is quite congested in parking space, the fact that there is no need to spend time finding a parking slot in the model, should increase the popularity of car in the model.

Other factors playing a big role in the overrepresentation of car trips, and underrepresentation of public transport trips, are the assumption that all agents that are over 18 years old have access to a car and the workplace location model. In the workplace location model, an agent chooses its place of employment based on the amount of time the work trip takes *by car*. This leads to the result that those workplaces which are accessible well by public transport, rather than by car, are not easily chosen. These places are especially ones near train or metro stations. As a large portion of all trips are trips to work, the prevalence of car as the mode of choice for work trips should influence the aggregate measures of modal split. Based on this, we would expect to see large underrepresentation of public transport trips in municipalities with good train connections like Espoo, Järvenpää, Hyvinkää and Mäntsälä. Based on Figure 7.4, this seems to be the case.

Looking at Table 7.3, another large difference when comparing the model to the travel survey is the length of trip by car. This may be caused by the workplace location model. Because the model emphasizes trips by car, some of the longer work trips that would be made by public transport, are then made by car, increasing the average length of trips by car and simultaneously shortening the average length of trips by public transport compared to the travel survey. This would at least partly explain the lack of longer trips by public transport seen in Figure 7.5.

Observing the length of trip distributions from the model and the travel survey, it seems that the aggregate trip length distribution matches the distribution from the travel survey quite well. The trip length distributions by modes on the other hand seem to match less well. Comparing the trip length distributions of car and public transport, we see that there is a lack of longer transit trips and some overabundance of car trips compared to the study. This suggests that discrepancies in mode choice distribution is the cause of on average too long car trips and too short public transport trips.

9.1 The case study

The model was used in examining a case study on the travel behaviour changes caused by the opening of a new train connection - the Ring Rail line - accompanied by bus line changes.

The examination of travellers (boardings and alightings) in train stations along the Ring Rail line does not look good. According to Table 8.1, the relative error in the number of travellers of each stop varies between -62% and 22%. All but two stations have too few travellers during the simulation day. This too may be caused by the workplace location model. As pointed out before, workplaces along train tracks suffer most from the car time based choice of workplace. Agents in the simulation whose domicile is near train stations probably have their workplace further away from train stations than they really do. This would lead to the agents using less trains in travelling.

Simulating the travel behaviour after the opening of the Ring Rail line, the model forecasts huge, up to 652%, increases in the usage of train stations. This is not plausible. The cause of the huge increase in the popularity of train travel is not clear. Some of it could be attributed to the too low popularity of train travel in the first place, resulting from the workplace location model, but as the home and work locations stayed the same across the simulations, it is not clear how much this would affect the results. In the five new stops, the excessive number of travellers can be attributed to the possibility that the changes in the travel behaviour of people have not yet been fully realised [Kiiskilä et al., 2017]. Unfortunately, it is hard to tell how much of the changes may not have realised yet, and it is therefore not possible to estimate to what extent this really affects the comparison. A third source of error in the forecast of the number of train travellers is land use. Between the before and after scenarios, the land use around the simulation area, particularly around the train stops of the Ring Rail line, has changed. At least the number of people living in Myyrmäki, Kivistö, Leinelä and Tikkurila has increased [Kiiskilä et al., 2017]. This has not been taken into account in the model, but could be done by creating their proper synthetic populations for the different scenarios. It must be noted, though, that the relative change in the number of travellers on the stops around the areas whose population has increased is already too high. Therefore, this source of error cannot be used to explain the too large changes in the number of passengers between the scenarios.

In view of the results, in its current form the model would seem to be most suitable for analysing the car traffic of the Helsinki area. The outputs of the MATSim simulation are suitable for many analyses of this kind: for example, the hourly traffic amounts can be extracted from them. This is an improvement compared to the traditional models of travel demand which only model the peak periods of travelling by scaling the forecast daily trips. Considering the travel behaviour of people through a whole day can capture changes in the timing of the peak traffic periods induced by increased population or changes in the transportation system. The observation that the model seems to work better in the municipalities belonging to the capital region, suggests that the model should be used in evaluating the travel demand of the capital region at large.

9.2 Development directions for the model

As an activity based simulation, the model is a step in to the right direction in the behaviourally realistic modelling of travel demand. The results of the model are an encouraging start for development, but there are still aspects of the model that need to be implemented and improved upon. These include:

1. Workplace location model
2. Increasing the number of possible modes
3. Adding new choice dimensions
4. Increasing agent heterogeneity

5. Adding an activity model

Many shortcomings of the model seem to be related to the workplace location model. The development of the simulation model should therefore be focussed on this model. The main deficiency of the workplace choice model is only using car travel time as a choice variable. In choosing a workplace, an agent should also take into account the possibility of travelling to work by public transport and perhaps also by walking. Further, there are other factors influencing the choice of a workplace, namely education and salary. Data requirements of the model and the availability of data must be taken into account when increasing the number of choice variables. A somewhat larger overhaul regarding the workplace choice model is to model residential location choice instead of occupation location choice. This means essentially reversing the home and work location choice process. As there are more residents around any given area compared to workplaces, shrinking of the choice set would not prove to be as considerable a problem as it is now in the model. Also, yet again with an increase in data requirements, the choice of residential location could be based on a more comprehensive measure of accessibility which accounts for possibilities for shopping, leisure and other activities in addition to travel time to work. Additionally, this change would require some changes in the way the synthetic population is constructed.

Another future direction for the model is that of expanding it. Interesting and useful expansions could include adding new modes, adding new choice dimensions and increasing agent heterogeneity. In addition, to forecast changes in travel demand further away in the future, an activity model would be beneficial.

As pointed out in the introduction, new travel modes are emerging to the transportation markets. Forecasting the impact and possible market shares of these new modes is intriguing for many different parties in the field of transportation. Also some significantly used current modes, particularly bicycles and taxis, are excluded from the current model and could be added for increased accuracy. In addition, the effects of transportation system changes for the popularity of bicycles and taxis is valuable information by itself.

Out of new choice dimensions, especially activity choices would bring new and interesting analysis possibilities to the model. Currently, there is no activity schedule model implemented and agents' activity schedules are directly drawn from those observed in a travel survey. Adding the possibility for agents to change their activities or the order of their activities in-between iterations would make it possible to forecast changes in the agents' activity schedules based on changes in the transportation system.

Increasing agent heterogeneity would make it possible to study the effects of transportation system changes to different groups of people. One way to introduce increased heterogeneity is to implement different scoring functions for different agents. As people have different preferences over modes and travelling in general, clustering of travellers could make it possible to differ the parameters of the scoring function for agents that are in a given cluster. Combined with simulating new and emerging travel modes, this could be used to forecast the possibilities of the modes among different clusters of travellers, makingit easier to identify the target market of the modes, for example.

The scoring function in the MATSim enables us to estimate monetary benefits derived from abstract improvements of the transportation system, such as reachability. Because the scoring function includes a monetary term, the score of a plan can be transformed to monetary value. By making changes to the transportation system that increase the reachability of some parts of the city, these changes are reflected in the scores of the plans and can then be transformed to monetary values. With the same logic, one can estimate the gains of other improvements to the transportation system in monetary terms.

Lastly, an activity model would construct activity schedules for the agents based on characteristics of the agents. Because travel demand is derived from the need to participate in activities and there are no guarantees that current activity patterns will last in the future. Therefore, a model predicting the activity patterns would bring added credibility for forecasts. Combining this model to the MATSim simulation would be preferred, as not only the characteristics of people influence their daily activity schedules, but also the changes in the transportation system and governmental and employer policies have an effect on them. Bowman and Akiva [1996] give an example of how an employer policy could affect the activity schedule of a worker.

9.3 Future prospects and deployment

Another important topic in activity-based travel demand models is the deployment of such models. Because the public sector is the main user of travel demand models, efforts should be made for them to adopt the best-practice models in its decision making. Comparability is important when proposing new tools for decision making to the public sector. Because the decision processes are a continuum and span over many areas in the public sector, it is crucial that the information that is used can be compared to that of others and of the past. Another aspect in adopting new tools for decision making is that they need to provide better information than those that have been used before. The superiority of activity-based models over the four step model becomes more evident the farther away to future we try to forecast the travel demand. Because the environment that the model needs to forecast becomes more different from the present one, it becomes more important to base forecasts on well-founded theory. In the case of traffic and travelling, this means that the patterns and needs that create a demand for travel, namely activities, and their spatial and temporal restraints that affect this demand need to be studied and understood. In this way, activity-based models' representation of travelling as a derived need is likely to outperform the four step model's trip based modelling approach as we try to forecast the distant future.

In the future, activity-based models could be used for project evaluation and to forecast trip-chain changes, possibilities of emerging transportation modes, benefits of increased accessibility in monetary terms and activity and movement pattern changes that emerge from increased accessibility.

Such uses require efforts in the study and development of the activity-based models. First, to make the results comparable with the one's derived from currently used four step models, the forecasts should be made with similar data and the results of the more disaggregate activity-based models should be aggregated to match the current models. In addition, especially the decision making process of households for activity participation and scheduling should be studied.

For a long time, the research community has viewed the activity-based approach as a superior method to model travel demand, compared to the four step model. This stance has not been reflected in the practice of travel demand simulation where the traditional models still dominate. To increase the level and detail of travel demand simulation to match the needs of the users, the practice of activity-based travel demand simulation calls for further development of the models with the aim of evaluating projects and policies even better.

Bibliography

R. Arnott, D. Palma, and R. Lindsey. A structural model of peak-period congestion: A traffic bottleneck with elastic demand. *The American Economic Review*, 83(1):161, 1993.

J. Banks, J. S. Carson II, B. L. Nelson, and D. M. Nicol. *Discrete-Event System Simulation*. Pearson, Michigan, fourth edition, 2005.

C. R. Bhat and F. S. Koppelman. *Handbook of transportation science*. Springer US, 72 Spring Street, New York, NY 10012, second edition, 2003.

J. L. Bowman and B. M. Akiva. Activity based travel forecasting. In *Activity-Based Travel Forecasting Conference (New Orleans, Louisiana, June 2, 1996)*. US Department of Transportation and the Environmental Protection Agency, 1996.

D. Boyce. Is the sequential travel forecasting paradigm counterproductive? *Journal of Urban Planning and Development*, 128(4):169–183, 2002.

S. F. Chapin. Free time activities and quality of urban life. *Journal of the American Institute of Planners*, 37(6):411–417, 1971.

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

I. Cullen and V. Godson. Urban networks: the structure of activity patterns. *Progress in Planning*, (4):1–96, 1975.

European Environment Agency. Assessment of global megatrends

— an update. https://www.eea.europa.eu/themes/sustainability-transitions/global-megatrends/global-megatends, 2017. Accessed 4.9.2017.

Findicator. Urbanisation. http://www.findikaattori.fi/en/56, 3 2017. Accessed 23.3.2017.

Google. GTFS static overview. https://developers.google.com/transit/gtfs/, 3 2017. Accessed 23.3.2017.

S. Hanson. The context of urban travel. In S. Hanson and G. Genevieve, editors, *The geography of urban transportation*, pages 3–24. The Guilford Press, 72 Spring Street, New York, NY 10012, 2004.

David A. Hensher and K. J. Button, editors. *Handbook of Transport Modelling*. Elsevier Science Ltd., Oxford, first edition, 2000.

T. Hägerstrand. What about people in regional science? *Papers in Regional Science*, 24(1):7–24, 1970.

A. Horni, K. Nagel, and K. W. Axhausen. High-resolution destination choice in agent-based demand models. In *Annual Meeting Preprint*. Transportation Research Board, Washington, 2012.

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

H. Kalenoja, K. Vihanti, V. Voltti, A. Korhonen, and N. Karasmaa. Liikennetarpeen arviointi maankäytön suunnittelussa (Evaluation of traffic needs in land use planning). Technical report, Ympäristöministeriö, 2008.

K. Kiiskilä, J. Tuominen, N. Frösén, R. Valli, and A. Herneoja. Kehäradan liikenteelliset vaikutukset (Impacts of the ring rail line on traffic). Technical report, Helsingin Seudun Liikenne, 2017.

R. Kitamura. Applications of models of activity behavior for activity based demand forecasting. In *Activity-Based Travel Forecasting Conference (New Orleans, Louisiana, June 2, 1996)*. US Department of Transportation and the Environmental Protection Agency, 1996.

H. Konishi. Formation of hub cities: transportation cost advantage and population agglomeration. *Journal of Urban Economics*, 48(1):1–28, 2000.

C. Marchetti. Anthropological invariants in travel behavior. *Technological Forecasting and Social Change*, 47(1):75–88, 1994.

D. McFadden. Conditional logit analysis of qualitative choice behaviour. In *Frontiers in Econometrics*.

M. G. McNally. The four step model. In David A. Hensher and K. J. Button, editors, *Handbook of Transport Modelling*, pages 35–52. Elsevier Science Ltd., Oxford, 2000a.

M. G. McNally. The activity-based appoach. In David A. Hensher and K. J. Button, editors, *Handbook of Transport Modelling*, pages 53–69. Elsevier Science Ltd., Oxford, 2000b.

R. B. Mitchell and C. Rapkin. *Urban Traffic: A Function of Land Use*. Columbia University Press, New York, 1954.

P. L. Mokhtarian and C. Chen. TTB or not TTB, that is the question: a review and analysis of the empirical literature on travel time (and money) budgets. *Transportation Research Part A: Policy and Practice*, 38(9):643–675, 2004.

J. Ortúzar and L. G. Willumsen. *Modelling Transport*. John Wiley & Sons, Ltd., fourth edition, 2011.

S. Rasouli and H. Timmermans. Activity-based models of travel demand: promises, progress and prospects. *International Journal of Urban Sciences*, 18(1):31–60, 2014.

A. Schafer. Regularities in travel demand: an international perspective. *Journal of Transportation and Statistics*, 3(3):1–31, 2000.

R. E. Shannon. Introduction to the art and science of simulation. In *Proceedings of the 1998 Winter Simulation Conference, Washington DC*. IEEE Computer Society Press, 1998.

J. D. Shires and G. J. de Jong. An international meta-analysis of values of travel time savings. *Evaluation and Program Planning*, 32(4):315–325, 2009.

H. J. P. Timmermans and T. A. Arentze. Transport models and urban planning practice: Experiences with albatross. *Transport Reviews*, 31: 199–207, 2011.

Helsinki Region Transport. Helsingin seudun työssäkäyntialueen liikenne-ennustemallit 2010 (traffic forecast models for the helsinki region commuting area 2010), 2011.

Helsinki Region Transport. Liikkumistottumukset Helsingin seudulla 2012 (HLJ 2015 Travel behavior in the Helsinki region in 2012). Technical report, Helsingin Seudun Liikenne, 2013.

V. Trivett and Skift Staff. What the sharing economy means to the future of travel. Technical report, Skift, 2013.

B. Vilhelmson. Daily mobility and the use of time for different activities. the case of Sweden. *GeoJournal*, 48(3):177–185, 1999.

M. Wardman, M. Page, M. Tight, and Y.L. Sin. Cycling & urban commuting: Results of behavioural mode and route choice models., 2000.

Y. Zahavi and A. Talvitie. Regularities in travel time and money expenditures. *Transportation Research Record*, 750:13–19, 1980.