THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Agent-based Models for Evaluating Sustainable Transportation Systems

ÇAĞLAR TOZLUOĞLU

Department of Space, Earth and Environment
Division of Physical Resource Theory
CHALMERS UNIVERSITY OF TECHNOLOGY

Göteborg, Sweden 2024

Agent-based Models for Evaluating Sustainable Transportation Systems

ÇAĞLAR TOZLUOĞLU

ISBN 978-91-8103-135-5
ISSN 0346-718X
Doktorsavhandlingar vid Chalmers tekniska högskola, Ny serie 5593

Department of Space, Earth and Environment
Division of Physical Resource Theory
Chalmers University of Technology
SE-412 96 Göteborg
Sweden
Telephone: +46 (0)31-772 1000

Cover artwork by Ezgi

Chalmers Reproservice
Göteborg, Sweden 2024

ABSTRACT

Many countries and organisations are adopting strategies to promote a sustainable transition of the transportation system and mitigate the adverse effects of climate change. To guide these efforts, decision-makers need tools that are capable of evaluating complex transportation policies and assessing sustainable transportation scenarios.

This thesis presents the development of transportation models that simulate individuals' travel behaviours and evaluate sustainable transportation technologies. Agent-based models (ABMs) are well suited to address the complexities of transportation systems through individual-level modelling, while activity-based travel demand generation integrates the behavioural dynamics driving demand into the model. Specifically, the thesis aims to answer three research questions: (1) How can a synthetic population of Sweden be created for use in ABMs? (2) How can realistic travel demand be generated for the developed population? (3) How can state-of-the-art methods and big data sources be incorporated into the modelling process?

Papers I and II introduce the Synthetic Sweden Mobility (SySMo) model, a large-scale agent-based transportation model. Paper I details the creation of a synthetic population for the entire Swedish population, addressing a significant data gap (Q1). Paper II focuses on SySMo's travel demand generation module, proposing a novel methodology that preserves heterogeneity in travel demand using machine learning algorithms (Q2-Q3). Paper III explores the integration of big data sources into activity-based models, introducing a generative model that synthesises multi-day activity-travel plans for over 263,000 individuals residing in Sweden using mobile data (Q3). Paper IV demonstrates an application of SySMo by analysing sustainable travel modes, specifically assessing the potential of e-bikes to replace passenger car trips and reduce greenhouse gas emissions.

The developed models may assist decision-makers by capturing individuals' travel behaviours and evaluating emerging technologies, enabling informed policy formulation. To support collective efforts toward sustainable transportation, this research adheres to open science principles by making the models publicly accessible.

Keywords: Agent-based modelling, Activity-based modelling, Synthetic population, Activity-travel plans, Big data.

Appended publications

This thesis consists of an extended summary and the following appended papers:

Paper I Ç. Tozluoğlu, S. Dhamal, S. Yeh, F. Sprei, Y. Liao, M. Marathe, C. L. Barrett and D. Dubhashi (2023b). A synthetic population of Sweden: datasets of agents, households, and activity-travel patterns. *Data in Brief* **48**, p. 109209.

Paper II Ç. Tozluoğlu, S. Dhamal, S. Yeh, F. Sprei, Y. Liao, M. Marathe, C. Barrett and D. Dubhashi (2022b). Heterogeneous activity generation for a large-scale synthetic population. *Working manuscript.*

Paper III Ç. Tozluoğlu, Y. Liao and F. Sprei (2024b). Mobile phone application data for activity plan generation. *Under review.*

Paper IV Ç. Tozluoğlu, Y. Liao and F. Sprei (2024c). Potential of e-bikes to replace passenger car trips and reduce greenhouse gas emissions. *Journal of Cycling and Micromobility Research* **2**, p. 100043.

Author contributions

Paper I: ÇT, SD, SY, FS, YL, MM, CB and DD conceptualised the study. ÇT and SD processed the data and developed the software. ÇT and SD wrote the original draft. ÇT, SY, FS, YL reviewed, edited and approved the final version of the manuscript.

Paper II: ÇT, SD, SY, FS, YL, MM, CB and DD conceptualised the study. ÇT and SD processed the data and developed the software. ÇT and SD wrote the original draft. ÇT, SY, FS, YL reviewed, edited and approved the final version of the manuscript.

Paper III: ÇT, YL, FS conceptualised the study. ÇT and YL processed the data. ÇT developed the software. ÇT wrote the original draft. ÇT, YL, FS reviewed, edited and approved the final version of the manuscript.

Paper IV: ÇT, YL, FS conceptualised the study. ÇT and YL processed the data. ÇT and YL developed the software. ÇT wrote the original draft. ÇT, YL, FS reviewed, edited and approved the final version of the manuscript.

Other publications

Other publications by the author not included in the thesis:

I. Ç. Tozluoğlu, S. Dhamal, Y. Liao, S. Yeh, F. Sprei, M. Marathe, C. Barrett and D. Dubhashi (2022a). Synthetic Sweden Mobility (SySMo) model documentation. URL: https://research.chalmers.se/publication/531094/file/531094_Fulltext.pdf.

II. Y. Liao, Ç. Tozluoğlu, F. Sprei, S. Yeh and S. Dhamal (2023). Impacts of charging behavior on BEV charging infrastructure needs and energy use. *Transportation Research Part D: Transport and Environment* **116**, p. 103645.

Data repositories:

I. Ç. Tozluoğlu, S. Dhamal, Y. Liao, S. Yeh, F. Sprei, M. Marathe, C. Barrett and D. Dubhashi (2023a). A synthetic population of Sweden: datasets of agents, households, and activity-travel patterns. *Mendeley Data*. DOI: https://doi.org/10.17632/9n29p7rmn5.2.

II. Y. Liao, Ç. Tozluoğlu, G. Kaniska, S. Dhamal, F. Sprei and S. Yeh (2024). Integrated agent-based modelling and simulation of transportation demand and mobility patterns in Sweden. *zenado*. DOI: https://doi.org/10.5281/zenodo.10648078.

III. Ç. Tozluoğlu, Y. Liao and F. Sprei (2024a). Synthetic multi-day activity-travel schedules for Swedish residents. *zenado*. DOI: https://doi.org/10.5281/zenodo.14012139.

Acknowledgements

Time has flown since I started this journey. Looking back now, I am aware that it has been filled with rewards and challenges, none of which I could have faced alone. My heartfelt gratitude goes out to everyone who shared this incredible journey with me.

I express my deepest gratitude to my supervisor, Frances Sprei, for her trust in my work and inspiring guidance. Her willingness to listen and offer support at every step made this thesis possible. I am also thankful to my examiner, Sonia Yeh, for initiating this journey. Her insightful questions have been instrumental in advancing my research. I am grateful to my co-supervisor, Yuan Liao, for her direct feedback and hands-on support. Additionally, I thank Swapnil Dhamal for accompanying me in countless online coding sessions during the COVID-19 pandemic. Your intellectual and scientific support created an exceptional research atmosphere, and the time you always reserved for discussion was invaluable.

I extend my thanks to Madhav Marathe and Christopher Barrett from the University of Virginia for their valuable feedback on my research and for warmly hosting me during my time in Virginia.

Throughout my work, I was surrounded by wonderful people in the Physical Resource Theory division. I am grateful to Martin Persson and Angelica Linnehav for their care and support, making life easier. Special thanks to my office mates, Ankit and Avi, for being my partners in crime; to Ahmet for our many conversations; and to Hanna for sharing the beauty of Dalarna. To Manuel, Caroline, Lola, Yodefia, Jinxi, Ella, Devon, Kaniska, and those not mentioned by name, you made my time incredibly enjoyable with excellent tea parties, ping-pong matches, after-work gatherings, and various events.

My utmost gratitude goes to my family: my parents, Aynur and Mustafa, for their unconditional support from the very beginning; my brother, Çağrı, and sister-in-law, Melda, for being sources of inspiration. Last but certainly not least, my heartfelt thanks to my wife, Ezgi, for encouraging me to embark on this journey. Your love, patience, and unwavering support have made life not only manageable during the toughest times but truly wonderful. Here's to the many adventures that still lie ahead of us.

Göteborg, Dec 2024
Çağlar Tozluoğlu

Contents

CHAPTER 1

Introduction

The world faces a significant challenge in reducing greenhouse gas (GHG) emissions to mitigate the negative effects of climate change. The Intergovernmental Panel on Climate Change (IPCC) reports with high confidence that achieving climate mitigation goals necessitates transformative changes in the transportation sector [1]. In Sweden, transportation is responsible for 32% of GHG emissions, and passenger cars within the sector have the most significant share, with 62% of the emissions [2].

In response to this challenge, many countries and organisations have implemented strategies to restructure transportation systems and accelerate the transition toward sustainability. In passenger transport, efforts are primarily focused on promoting the adoption of zero- or low-emission vehicles through subsidies, contributing to the global growth in electric vehicle sales [3]. Additionally, mobility innovations such as shared autonomous electric vehicles and on-demand ride-sharing services have the potential to transform human travel behaviours toward more sustainable future transportation systems [4, 5]. To understand the undergoing transformations in the system and evaluate emerging transportation technologies, transportation models are commonly employed, providing abstract representations of the system.

Traditionally, aggregate transport models, such as four-step models, have been widely adopted in transportation modelling [6, 7]. These models employ a trip-based approach and generate aggregated travel flows between areas within a certain time period. However, these models have been criticised for inadequately representing human travel behaviour, making them less sensitive to evaluating complex transportation policies. Furthermore, emerging transportation technologies require individual-level analysis, which traditional aggregate models cannot effectively provide [8]. For example, evaluating shared autonomous cars can be challenging in the traditional models, as these models do not simultaneously track individuals and vehicles. These limitations highlight the need for decision-support tools that are capable of reflecting individuals' travel behaviours and integrating new transportation technologies to inform policymaking.

The agent-based modelling (ABM) framework, equipped with an activity-based travel demand generation approach, provides a tool for simulating scenarios that support transport planning and policy formulation. This integrated approach enables a detailed representation of individuals with their travel behaviours and interactions, making it well-suited for addressing the complexities of transportation systems. Over the past few decades, the increased computational power, rise of big data sources and the need for more sophisticated models have expanded these models' applications [9], including the simulation of various scenarios with emerging transportation technologies [e.g., 10, 11].

An essential step in developing ABMs is generating a synthetic travel demand, which includes a synthetic population of individuals with socio-demographic characteristics and daily activity patterns. While several countries, including Switzerland [12], Denmark [13], and France [14], have developed synthetic populations, Sweden lacks a publicly available synthetic travel demand representing its entire population, limiting the ability to conduct detailed simulations. Furthermore, recent advancements in computer science and the availability of emerging big data sources present new opportunities to develop methodologies that realistically capture individuals' mobility behaviours on a larger scale.

This thesis aims to develop decision-support tools that can capture individuals' travel behaviours and evaluate sustainable transportation scenarios in Sweden. It investigates the developments of large-scale transportation models with advanced synthetic populations using state-of-the-art methods from recent advancements in transportation modelling and computer science, thereby enhancing the realism of the simulations. This thesis also explores the use of emerging big data sources in synthetic travel demand generation. The developed models are designed to assist decision-makers by reflecting individuals' travel behaviours and transportation technologies, enabling the formulation of informed policies. To contribute to the collective efforts supporting sustainable transportation and maximise the impact of these models, the research adheres to open science principles, making the models publicly accessible. This approach not only promotes transparency and collaboration but also ensures that the research can effectively inform and influence sustainable transportation planning and policymaking.

Scope and contributions

This thesis focuses on modelling human travel behaviour in Sweden using an agent-based modelling framework combined with an activity-based approach. The main research question is:

- How can we develop transportation models that are capable of simulating individuals' travel behaviours and evaluate sustainable transportation technologies to study transitions toward sustainability?

To address this overarching question, the thesis further investigates the following sub-questions:

- How can we generate a synthetic population that is a statistically accurate representation of the Swedish population with certain attributes for use in ABMs? (Q1)
- How can we generate realistic individual-level travel demand for the developed synthetic population? (Q2)
- How can we incorporate state-of-the-art methods in computer science and emerging big data sources into the modelling process to increase the realism of the simulations? (Q3)

In answering these questions, **Papers I–III** present the development of large-scale ABMs for Sweden, and **Paper IV** demonstrates the application of one of the models.

Paper I and **Paper II** introduce the Synthetic Sweden Mobility (SySMo) model. **Paper I** details the generation of the synthetic population with various socio-demographic variables and activity-travel schedules across the entire Swedish population (Q1). This study makes the model outputs publicly available, addressing a significant data gap in Sweden and facilitating the building of agent-based transport simulations by other researchers.

Paper II focuses on the travel demand generation module within the SySMo model. It proposes a novel methodology that combines probabilistic neural networks with random sampling to generate daily activity schedules for the synthetic population (Q2-Q3). This methodology generates detailed average weekday activity-travel patterns, including activity sequences, types, start and end times, and locations. This approach captures heterogeneous activity patterns within sub-populations, ensuring diverse mobility plans based on socio-economic attributes.

Paper III explores the integration of big data sources into activity-based travel demand generation (Q3). It introduces a generative model that synthesises activity-travel plans from extensive but incomplete mobile phone application data. This model is capable of generating multiple weekday activity schedules, including activity sequences, types, start and end times, and locations, while incorporating daily variability. The study also proposes a novel approach to improve the state-of-the-art home and work location identification approaches to realistically synthesise activity-travel plans.

Paper IV exemplifies the practical applications of the developed models in promoting sustainable transportation systems. The study applies the SySMo model to evaluate the potential of e-bikes to replace passenger car trips and reduce greenhouse gas emissions. This study leverages individuals' socio-demographic variables, daily activity patterns, and the availability of biking infrastructure to realistically identify trips that e-bikes can replace. The findings highlight the benefits of e-bike adoption in achieving sustainable transportation systems and reducing emissions.

Disposition of this thesis

The thesis consists of five chapters briefly introducing my doctoral studies, followed by four papers. Chapter 2 reviews transportation modelling tools with an emphasis on individual-level approaches. It introduces key concepts in the transportation field and explains agent-based modelling techniques and activity-based travel demand modelling. Chapter 3 provides a summary of the appended four papers and presents the main findings. Finally, chapter 4 overviews the developed models, reflects on my PhD research and outlines potential future research directions. It discusses the potential and limitations of the models and open science principles and presents various applications of the models. The end of the chapter reflects on my research and outlines my future research.

CHAPTER 2

Background

This chapter provides an overview of transportation models, focusing on individual-level transportation modelling approaches. It begins by introducing the fundamental concepts of transportation modelling, establishing a foundation for understanding two key modelling frameworks: four-step and agent-based transportation modelling. Section 2.2 discusses four-step transportation models that provide aggregated trip flows between areas. Section 2.3 explores agent-based models, which simulate individuals' travel behaviours. The section also presents agent-based models' main components, including the activity-based travel demand generation approach. Since the papers included in this thesis employ agent-based transportation modelling combined with activity-based travel demand generation, the chapter dedicates more space to this framework.

Before discussing the concepts in transportation modelling and various modelling frameworks, it is essential to define what transportation is. As described by Hensher et al. [15]:

> "Transportation is concerned with the movement of goods and people between different locations and systems used for this movement. Included in the former would be the journey to work, trade flows between nations, commodity flows within a single nation, passenger flows by various modes, and so forth, and those factors that affect these flows. In general, movement within a single industrial firm or building, or the migration of population, is not included in this area."

This definition highlights the broad scope of transportation, which encompasses the movement of both people and goods across various geographic scales, thereby playing a pivotal role in shaping numerous aspects of society and the economy. A well-functioning transportation system supports economic growth by facilitating trade, enhancing productivity, and creating jobs [16]. It also shapes urban development by influencing land use and the organisation of cities [17]. Additionally, transportation impacts social equity by providing access to essential services, employment, and opportunities, especially for disadvantaged groups [18]. Furthermore, transportation has significant environmental implications, as it is a major contributor to greenhouse gas emissions that drive climate change and to local pollutants that adversely affect air quality and public health, especially in dense urban areas.

This thesis primarily focuses on the movement of people and the modelling frameworks employed to simulate and analyse human travel behaviours within transportation systems.

2.1 Transportation modelling

Models are essential tools for scientists, planners, and decision-makers to understand complex phenomena. They provide simplified representations of real-world systems, focusing on the key elements relevant to a particular perspective. Through these representations, modelling plays a crucial role in most decision-making processes. Moreover, experimentation allows modellers to gain significant insights into the system's behaviour and inner workings. This thesis explores transportation models and various transportation modelling techniques.

Transportation models create an abstract representation of the transportation system, including key components such as infrastructure, vehicles, users, and control mechanisms and their interactions with each other, either as a whole or as a particular subsystem, using mathematical techniques grounded in specific theories and assumptions [6]. For informed decision-making, it is essential to understand how transportation systems function across space and time, both in the present and in various future scenarios. These models generate quantitative outputs and are critical tools for addressing a wide range of questions related to human mobility. Although these models provide simplified representations, they need substantial amounts of data for their application and can be highly complex.

To address a wide range of transportation issues, transportation modellers developed numerous models over the years. The first operational models, developed in the 1950s, were primarily concerned with planning investments to meet the increasing demand for road infrastructure in the United States [19, 20]. These models focused mainly on car travel during peak hours. However, as transportation systems have changed and decision-makers have increasingly prioritised environmental sustainability, models have significantly advanced to address concerns about the system's environmental impacts and the need for sustainable transport solutions. Models have expanded to incorporate multiple transport modes over more extended time periods and have started using disaggregated modelling techniques [21]. These advancements in transportation models have also diversified their applications, moving beyond investment planning to areas such as travel demand management [22], environmental pollution assessment [23], and energy needs evaluation [24, 25]. This shift highlights the models' increased capacity to capture transportation systems' complexity and varied impacts.

2.1.1 Types of transportation models

Transportation models can be categorised into three main types based on their applications: strategic decision-making, scenario analysis, and project/operational design models [26]. Each type serves distinct purposes, from long-term policy formulation to daily traffic operations management. This categorisation assists in selecting the appropriate model type depending on the specific objectives and requirements of a transportation analysis or planning project.

- **Strategic decision-making** models address high-level questions that inform policy development and long-term planning. These models analyse transportation systems over large geographic areas and extended time frames, such as assessing the impacts of public transportation expansion or land-use policy changes on transportation systems over the years.
- **Scenario analysis** explores a range of hypothetical situations to understand the impacts of different factors on transportation systems, such as technological innovations or policy interventions. These models typically simulate specific time points,

such as base or future years. Future scenarios are generated exogenously, considering key factors like population growth and employment levels, without internal feedback mechanisms.

- **Project/Operational design** modelling focuses on the management and performance of transportation systems, providing detailed insights into traffic dynamics. These models simulate vehicle movements and driver behaviours, enabling granular analysis of traffic flow and operational performance. They are typically applied to small geographic areas, such as consecutive intersections or specific road segments, to identify and mitigate traffic delays, queues, and bottlenecks, thereby enhancing the overall functionality of transportation systems.

This thesis focuses on developing a model capable of incorporating individuals' behaviours and addressing the complexities of transportation systems for sustainable transportation scenario analysis. Travel demand generation is a crucial part of model development. The following subsection introduces travel demand generation for large-scale transportation models, establishing a foundation for the four-step and agent-based transportation models, which are commonly used in scenario analysis.

2.1.2 Modelling travel demand

The modelling travel demand is a crucial part of transportation modelling. The complexity of transportation systems primarily arises from travel demand driven by human behaviours and its interaction with limited transportation supply [27]. Demand for transportation is influenced by demographics, economic activity, available transport options, service quality, pricing, and land use, which fluctuate over time and across regions [28]. Understanding these factors assists planners and policymakers in developing more efficient and sustainable systems. This thesis primarily focuses on modelling transportation demand, which often serves as the starting point for modelling studies.

Travel demand is considered a derived demand rather than being an end in itself [29]. People travel to fulfil their needs by engaging in activities at specific locations such as work, leisure, or grocery, with possible exceptions, such as sightseeing and hiking [30]. To understand travel demand, we also need to understand how these activities are distributed across space and time. These spatial and temporal distributions influence when and where people choose to travel, thereby shaping the travel demand.

The demand for transport services is highly qualitative and diverse, varying significantly based on factors like time of day, day of the week, journey purpose, speed, frequency, and cost [6]. For instance, commuters may prioritise speed and reliability during weekday mornings, while leisure travellers might value comfort and cost during weekends. A transport service that does not align with these specific needs may be ineffective or underutilised. This diversity in demand adds complexity to the analysis and forecasting, making it challenging for planners and policymakers to design services that meet the varied needs of the population.

Travel demand can be modelled in various ways. One primary categorisation is based on the level of aggregation: models can be either aggregated or disaggregated. Additionally, depending on the unit of analysis, there are different modelling approaches, such as trip-based and activity-based modelling. These modelling approaches are elaborated below.

Aggregate and Disaggregate Modelling

Travel demand models are developed to represent either the travel behaviour of individuals within a population or the population as a whole. Depending on this representation, models are categorised as either aggregate or disaggregate.

Aggregate models estimate travel behaviour between spatial regions for a population [31]. Travel demand occurs over space, and the activities' spatial distribution impacts this demand. Therefore, a common method to address spatial considerations involves dividing study areas into zones. These models often represent flows between origins and destinations without detailing individual choices, focusing instead on aggregate movement patterns. The aggregation level of the model is defined depending on various factors, such as the modelling objectives and the data structure.

Cascetta [27] formulates a travel demand model organised as flows between two points or regions as a function of the population's socioeconomic characteristics and transport infrastructures features, according to the given travel, considering specific travel attributes such as purpose or mode. This relationship can be represented as:

$$\mathbb{D}_{o,d}(K_1, K_2, ...K_n) = f(SE, T, \beta) \tag{2.1}$$

where, $\mathbb{D}_{o,d}$ denotes the travel demand flow from origin o to destination d with characteristics K_n. SE shows the socioeconomic variables of the decision-makers in the transportation system, and T denotes the level-of-service attributes of the transportation supply system. β specifies the model parameters related to the travel flow between o, and d. The utilised parameters vary depending on the adopted model (see more in Chapter 8 in the book [27]).

Disaggregate models, on the other hand, represent the travel behaviours of individual decision-makers or groups with similar characteristics, such as households [27]. These models capture the decision-making processes at an individual level, allowing for a more detailed analysis of factors influencing travel choices. Disaggregate travel demand models typically consist of a combination of submodels, such as mode choice models or destination choice models, which are used to forecast various aspects of trips.

The choice between aggregate and disaggregate modelling depends on several factors, including modelling objectives, data availability, time constraints, and domain knowledge. Disaggregate models offer detailed insights but require extensive data and computational resources. Conversely, aggregate models are less data-intensive but may overlook individual behavioural nuances. In practice, disaggregated model results may need to be aggregated to effectively inform planning and policy-making processes. While aggregating individual travel behaviours to obtain overall flows is relatively straightforward, disaggregating aggregate flows to understand individual behaviours requires more rigorous methods [31].

Trip-Based and Activity-Based Modelling

In the travel demand modelling literature, there are two widely used approaches: trip-based and activity-based modelling. The trip-based modelling approach is the first developed approach and calculates travel demand using trips as the unit of analysis. Early applications of this modelling approach assumed that trips occur independently of previous and subsequent trips, focusing on forecasting travel demand between zones at an aggregate level. These models do not consider the interdependencies between trips and the activities in individuals' schedules. Some studies have extended a trip-based approach by adopting tour-based models, which account for all trips within a tour [32–34]. In the tour-based modelling approach, the unit of analysis is tours that are defined as the trips from home to one or more locations and then back home [35]. There are also a limited number of disaggregated trip-based modelling studies that independently estimate each individual's travel [36]. Although the trip-based modelling approach has progressed over time, it often lacks a valid explanation of the underlying causes of travel behaviour.

To address these limitations, activity-based travel demand models have been developed, adopting a holistic approach that considers travel demand in connection with individuals' activity patterns. This modelling approach aims to jointly deduce individuals' activity schedules and the associated travel between activities over a specific time period, usually one day [37, 38]. By focusing on the underlying reasons for travel, the inadequacy of trip-based modelling in reflecting behavioural realism is overcome by the activity-based modelling approach.

The conceptual framework of activity-based modelling consists of several key ideas. First, travel demand originates from the need to participate in activities and is a derived demand rather than an end in itself [39]. From the economic perspective, people undertake travel only when the utility gained from participating in activities exceeds the disutility of the travel itself; most people do not travel for the sake of travel. The second is the time geography concept, which explains how individuals implement their activity agenda within the confines of time and space [40]. This concept highlights the constraints imposed by factors such as time availability, transportation options, spatial distances, and mandatory activities, all of which influence travel behaviour and activity participation.

To enhance the understanding of travel behaviour and explore various aspects of activity-based modelling, numerous studies have been conducted. For example, Pas [41], and Hanson [42] investigate the correlation between activity-travel patterns and socioeconomic attributes such as age, gender, and employment status. Kitamura [43] identifies the interdependence of activity locations within the activity sequence. Golob and McNally [44], and Pooley et al. [45] deal with interactions between household members and their activity patterns.

While this overview covers only the principal studies, more comprehensive reviews of activity-based modelling can be found in Buliung and Kanaroglou [46]. The various methodologies and applications of the activity-based approach are further discussed in Section 2.3.3, where we delve into specific models and their contributions to understanding and forecasting travel demand.

2.2 Four-step transportation models

Four-step transportation models (FSMs) are the most widely applied models in the transportation modelling field [6, 7]. These models use a trip-based approach and are a primary tool for evaluating large-scale infrastructure projects and policy interventions. FSM generates aggregated travel flows based on defined travel characteristics between regions over a certain time period, typically simulating peak hours or an average day. The overall framework of the FSM consists of four sequential and independent steps: trip generation, trip distribution, mode choice, and traffic assignment. These steps are defined as follows:

- **Trip generation:** This first step predicts the number of trips produced (originated) and attracted (destined) in each zone. Using zone-level statistics, the production and attraction numbers are calculated independently for each zone at an aggregate level. Trip production is estimated using variables such as population, number of households, income level, number of cars, and residential density. Trip attraction is calculated using variables such as office space, number of retail buildings, number of employees, and student capacity.
- **Trip distribution:** The second step aims to match trip origins and destinations by estimating where trips generated in one zone will end. The most common technique for calculating travel flow between regions is the gravity model, which distributes the number of trips between two locations inversely proportional to their distance or travel cost. This step results in the creation of origin-destination (OD) matrices representing the number of trips between each pair of zones.
- **Mode choice:** In this step, the total number of trips between zones is distributed among the available transportation modes,

such as car, public transit, walking, or cycling. Discrete choice models, like the multinomial or nested logit models, are often used to determine the modal split based on factors such as travel time, cost, and individual preferences. This step produces mode-specific OD matrices from the matrices generated in the previous step.

- **Trip assignment:** The final step assigns trips to specific routes within the transportation network, including roads or public transport networks, to simulate travel flows and network performance. Various assignment algorithms, such as user equilibrium or stochastic assignment, are used to model how travellers choose their routes based on factors like travel time and congestion. This step provides outcomes regarding the aggregated travel behaviour of the population and the operational characteristics of the network, such as link volumes and travel times.

Despite their widespread use, FSMs have been criticised for inadequately representing human travel behaviour [36, 47]. One of the main criticisms is that FSMs, adopting a trip-based approach, lack behavioural foundations associated with the creation of travel demand [48]. These models treat trips as independent events, disregarding the spatial and temporal interdependencies between trips in the same trip chain or tour [7]. Moreover, FSMs inadequately reflect the interrelationships among different characteristics of an individual's travel, such as time of day and destination selection.

These behavioural inadequacies make traditional FSMs less sensitive when evaluating complex transportation policies that target specific times of the day or aim to influence specific travel behaviours. For instance, most FSMs are not capable of predicting responses to travel demand management policies, such as strategies to increase car occupancy [49] or assessing innovative transport services like autonomous vehicles that may transport one household member before returning empty to collect another [8]. The aggregate nature of traditional models cannot accurately determine shared ride occupancy or ensure that individuals and vehicles are only in one location at any given time. FSMs are better suited for assessing infrastructure measures than behavioural interventions. In response to these limitations, there has been a shift toward more advanced models that incorporate behavioural theories and consider the interdependencies of travel decisions.

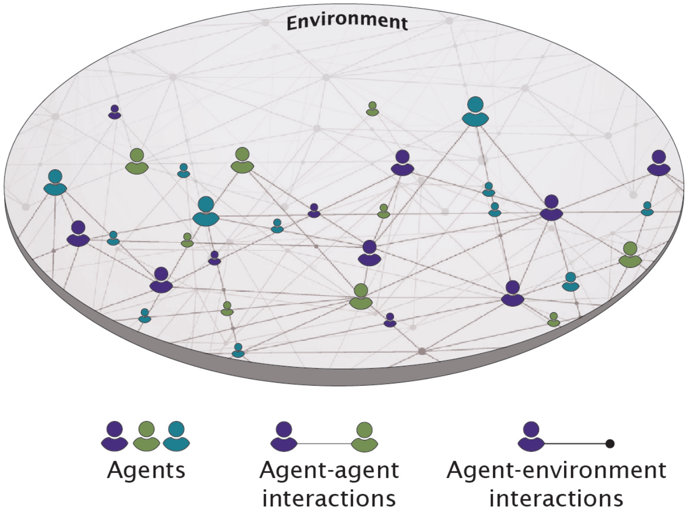


Figure 2.1: An overview of agent-based models with their typical elements.

2.3 Agent-based transportation models

Agent-based modelling (ABM) is employed in modelling studies across various fields, including transportation[46], [50], power markets [51], and financial systems [52]. It provides a framework for simulating complex systems by dividing them into individual actors interacting with each other and the environment based on predefined rules and characteristics [53, 54]. Figure 2.1 provides an overview of ABMs and their typical components.

ABMs have been increasingly used in transportation studies since the early 2000s. In these models, agents typically represent individual travellers with specific attributes, such as socio-demographic characteristics and travel behaviours that influence their interactions with other agents and the transportation environment. By modelling these interactions, ABMs aim to replicate how individual behaviours collectively influence and shape the transportation system. The disaggregated nature of ABMs makes them particularly valuable for capturing the complexity and heterogeneity of the transportation system. While the traditional models often struggle to track individuals and vehicles, ensuring they are in only one location at a time, ABMs simulate agents within the transportation network, ensuring realistic spatial and temporal consistency. Moreover, this modelling framework enables a detailed analysis of policy interventions' impacts on different population segments.

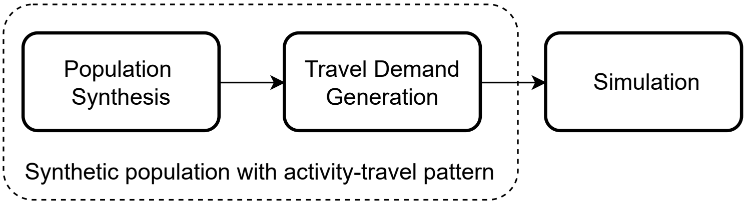


Figure 2.2: Agent-based transportation model workflow.

To model a transportation system using an ABM framework, detailed modelling of agents and their travel demand is essential. The typical ABM workflow involves three key steps: **population synthesis**, **travel demand generation**, and **simulation of agents** (see Figure 2.2). Each of these steps is crucial for building a realistic, comprehensive model.

The population synthesis step involves generating agents with attributes that reflect the target population's characteristics. The second step, travel demand generation, assigns travel demand to each agent based on their attributes. Activity-based models are commonly used for travel demand generation [55]. Combining agent-based modelling frameworks with activity-based travel demand modelling approaches allows for more behaviourally realistic simulations. The last step simulates the agents with the travel demand on the transport network. In this step, agents interact and compete with each other across time and space in an interaction-based traffic model. Before diving into the modelling steps, I will discuss the modelling structures of ABMs in the next section.

2.3.1 Model structures

The integration of an activity-based modelling approach with ABMs can be achieved through various modelling structures, each differing in the level of agent interaction and feedback mechanisms between demand generation and network assignment. Figure 2.3 explores possible modelling structures for the ABM framework.

Figure 2.3a illustrates an activity-based demand model combined with the static traffic assignment. In this structure, the activity-based model does not place in an ABM framework. The activity-based model generates detailed activity-travel plans for agents, which are then converted into tours or trips based on activity locations. These trips are assigned to the transportation network using traditional traffic assignment models, such as equilibrium models. There is no interaction among agents during the assignment process; instead, the focus is on accurately representing individual travel behaviour derived from activities.

Figure 2.3b depicts an ABM framework with an activity-based travel demand model with a feedback loop. In this structure, the activity-based model generates detailed activity-travel plans for each agent, including activity types, start and end times and locations. These plans are executed within an agent-based simulation, where agents interact with each other and the environment across time and space. Agents compete for limited resources, such as road capacity and vehicle availability, leading to emergent traffic patterns and congestion effects. The feedback loop allows the traffic simulation to provide data—such as travel times, delays, and network conditions—back to the demand model. This feedback mechanism enables the demand model to update activity-travel plans, resulting in more realistic outcomes.

Figure 2.3c shows an ABM framework with an activity-based travel demand model. However, in this structure, the activity and travel plans generated by the demand model are fixed and used as inputs to the agent-based simulation without a feedback loop. Agents execute their activity schedules within the simulation environment, interacting with other agents and the transportation network. While this structure captures the interactions among agents during the simulation, such as congestion effects and dynamic route choices, the absence of a feedback mechanism means that agents do not adjust their activity plans based on the simulation results.

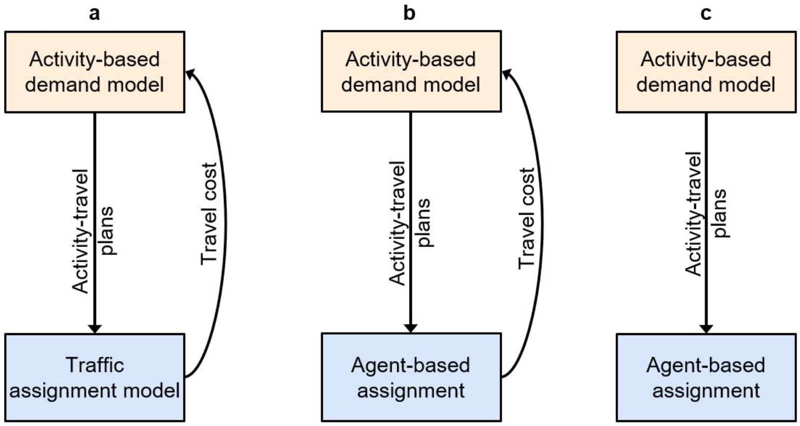


Figure 2.3: Model Structures of agent-based frameworks with activity-based demand models. Adapted from the source: Figure 1.3 [p10, 56].

2.3.2 Population Synthesis in Transportation Modelling

Population synthesis creates a statistically representative set of individuals within a specific geographical area Rich [13]. This synthetic population serves as the foundational input for agent-based transportation models. Agents are assigned socioeconomic details, such as age, gender, household information, and car ownership, based on the model's requirements. To capture social dynamics and shared resources such as household vehicles or parental responsibilities, agents are often grouped into households [57]. While some synthesis methods are dynamic, projecting future populations using fertility and mortality rates [58], this thesis focuses on generating populations using static methods for a base year.

There are two commonly used population synthesis methods, *reweighting* and *synthetic reconstruction* [59]. The reweighting method produces the population by adjusting the weights of micro-data obtained from national surveys so that, when aggregated, they match known population totals and distributions at smaller geographic levels. Traditionally, combinatorial optimisation and generalised regression techniques have been employed for this purpose. However, recent research has introduced new methodologies to meet the increasing demand for more detailed populations within smaller geographic zones [60]. These studies utilise machine learning-based algorithms, such as Markov models [61] and Variational Autoencoders (VAEs) [62], to simulate population characteristics with greater accuracy and granularity.

The synthetic reconstruction method generates a synthetic list of individuals and households whose aggregated characteristics match the known totals by the statistical data. This method does not need microdata, making them suitable when micro-data is scarce. The synthetic reconstruction method typically uses the iterative proportional fitting (IPF) technique and has widespread usage in the transportation modelling field (see, e.g., Smith, Beckman and Baggerly [63], Frick [64], Arentze, Timmermans and Hofman [65] and Guo and Bhat [66]). Paper I applies this technique to generate a synthetic population for Sweden, while Paper II and Paper III leverage it to maintain certain distributions in the activity-travel patterns.

Iterative proportional fitting

IPF technique consists of three main steps: initialisation, iterative scaling, and termination. Initialisation involves assigning initial counts to the joint distribution of the population's attributes. Iterative scaling then adjusts the initial counts so that they match marginal totals for each attribute. During each iteration, attributes are adjusted to fit their corresponding marginal totals while maintaining the changes from previous iterations. The process continues until the termination step, which occurs when the changes between successive iterations are minimal, indicating that the joint distribution has sufficiently aligned with the marginal totals.

To illustrate the application of IPF, a synthetic population with multiple attributes is generated. Let $\mathbf{x} = (x_1, x_2)$ represent the set of agent attributes, such as age group (x_1) and gender (x_2). Let $\mathbf{z} = (z_1, z_2)$ denote the zones where the population resides. Marginal totals for these attributes $N(x_1)$ and $N(x_2)$ are known, representing the number of agents with attributes x_1 and x_2. We also know the population sizes for each zone $N(z_1)$ and $N(z_2)$.

To estimate the joint distribution $n(z, x_1, x_2)$ that aligns with the known marginal totals, we can employ the IPF technique. The process begins with initialisation (Equation 2.2), where initial estimates of $n(z, x_1, x_2)$ are uniformly assigned by dividing the total population of each zone by the number of attribute combinations. Following initialisation, the technique proceeds to iterative scaling. In each iteration t, the cell counts are adjusted to fit each set of marginal totals sequentially. The adjustment rules are applied as shown in Equations 2.3 and 2.4.

$$n^{(0)}(z, x_1, x_2) = \frac{N(z)}{\text{Number of combinations of } x_1 \text{ and } x_2} \tag{2.2}$$

For each combination of attributes z, x_1, x_2:

$$n^{(t+1)}(z, x_1, x_2) \leftarrow \frac{N(x_1)}{\sum_{\substack{z' \\ x_{2'}}} n(z', x_1, x_{2'})} \times n(0)(z, x_1, x_2) \tag{2.3}$$

$$n^{(k+2)}(t, x_1, x_2) \leftarrow \frac{N(x_2)}{\sum_{\substack{z' \\ x_{1'}}} n(z', x_{1'}, x_2)} \times n^{(t+1)}(z, x_1, x_2) \tag{2.4}$$

$$n^{(k+3)}(t, x_1, x_2) \leftarrow \frac{N(z)}{\sum_{x_1, x_2} n^{(k+1)}(z, x_1, x_2)} \times n^{(t+2)}(z, x_1, x_2) \tag{2.5}$$

This iterative process continues until the changes between iterations fall below a predefined convergence threshold, ensuring that the joint distribution aligns with the marginal totals for each attribute. The IPF technique effectively estimates the joint distribution while maintaining the known correlation structures between attributes, making it a robust choice for population synthesis in transportation modelling.

2.3.3 Activity-based travel demand generation

Activity-based models generate travel demand based on the assumption that people travel to participate in activities. This approach is often utilised in ABMs to generate the travel demand [55], as well as used standalone, as shown in Figure 2.3a. Activity-based models integrate seamlessly with the AMB framework by providing a behavioural basis for agents' travel decisions [67]. Compared to traditional trip-based models, activity-based models offer a more detailed and realistic representation of travel behaviour.

Activity-based models predict agents' activities, including when and where they occur, their duration, purpose, the individuals involved, and the travel choices made to reach the activities. Figure 2.4 illustrates an individual's daily plan. Due to the complexity of decision-making and the level of detail required, these models can incorporate various sub-models of travel demand, including activity schedule generation, location assignment, and travel mode choice.

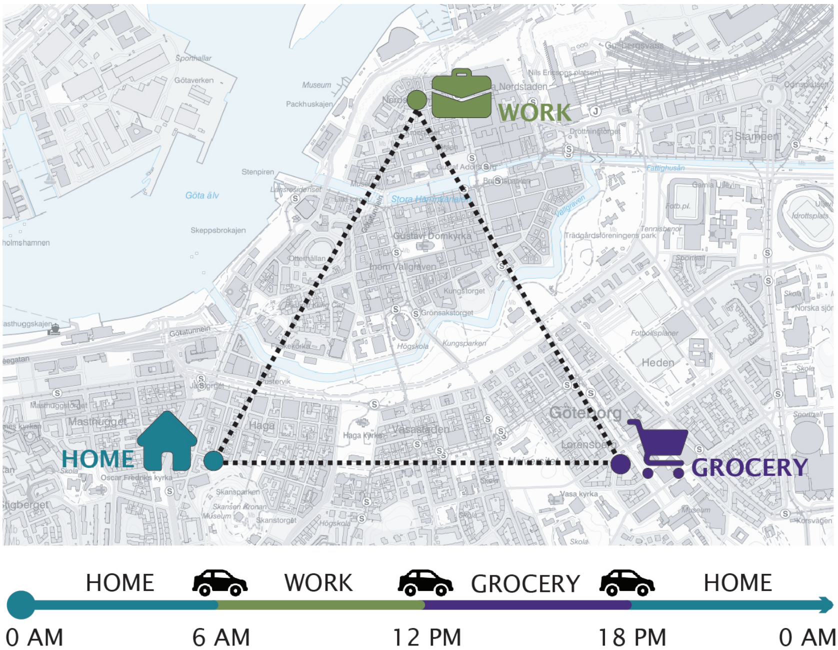


Figure 2.4: An activity schedule characterised by activities' type, start end time, location and travel mode between activities.

Methods to develop activity-based models

Methods used to develop activity-based models are generally grouped into three categories: constraint-based models, econometric models, and computational process models [68]. While some well-known transportation models rely on a single method, others integrate multiple methods to model different components. For example, TASHA [69] and ADAPTS [70] use a combination of methods to build activity-based models.

Constraint-based models evaluate the feasibility of a given activity schedule within certain constraints rather than predicting activity-travel patterns [71]. These models ensure that activity schedules can be executed without violating Hägerstrand's time geography concept

[40] by considering three main constraints: capability, coupling, and authority. Capability constraints refer to physical or biological limitations on an individual's movement, such as the need for rest or restrictions on travel speed. Coupling constraints involve the need to be in specific places at specific times with others, like attending meetings or appointments. Authority constraints pertain to institutional or legal restrictions on accessing certain spaces or times, such as operating hours or restricted areas. These constraints jointly shape the feasibility and execution of activity schedules in spatially distributed environments.

Examples of constraint-based models include PESASP [71], CARLA [72], BSP [73], MAGIC [74], and GISICAS [75]. These models vary in their level of spatial and network detail and the considered constraints. For instance, BSP and GISICAS incorporate more detailed transport network representations at the individual level compared to PESASP, while CARLA explores household-level accessibility. Constraint-based models are relatively straightforward to develop and implement, ensuring that generated schedules are consistent with spatial and temporal constraints. However, these models are often behaviorally limited, as they don't capture the decision-making processes of individuals or households. They often overlook the variability and uncertainty in travel behaviour and depend on fixed parameters like opening hours and maximum speed limits. [68, 76]. More recent models have addressed some of these limitations by incorporating varying travel speeds over time and space [77, 78], and decision-making under uncertainty [79, 80].

Econometric models, also known as utility-maximising models, are based on the theory that individuals aim to maximise the utility of their choices [68]. For example, when planning a weekend trip, a person might choose a nearby small park over a distant larger one due to the shorter travel time and easier access, which they perceive as more benefits. Econometric models utilise a series of discrete choice models, particularly nested logit models, to represent individuals' decision-making processes. This technique generates activity schedules that provide maximum utility to the individuals from their activity-travel choices [68, 81]. Ben-Akiva and Bowman [82] and Bowman and Ben-Akiva [83] developed prominent econometric models that formalise individuals' travel decision-making through five nested decision stages: choosing whether to travel or not, deciding the time of the primary tour of the day, selecting the primary destination and mode of transportation, determining the time of the secondary tour of the day, and choosing the secondary destination and mode of transportation. Other notable examples include CEMDEP [84] and PCATS [85].

Econometric models are widely used in travel behaviour modelling due to their behavioural richness and their grounding in the well-established theory of utility maximisation. They are particularly effective in predicting responses to policy changes. However, they have been criticised for assuming perfect rationality, suggesting that all decision-makers act as rational utility maximisers, consistently evaluating their options and selecting the one offering the greatest benefit [86]. This assumption does not always hold true. Individuals may make decisions that are less beneficial or whose benefits are uncertain, driven by bounded rationality and habitual behaviour.

Computational process models, also known as rule-based models, are a more recent method in activity-based modelling. These models employ heuristic rules such as if-then statements to simulate individuals' decision-making processes. For example, a rule might state that "if an individual is an employee, then the individual participates in work activity during the day". By applying these rules at various decision-making stages, the models generate activity-travel schedules. SCHEDULER [87] provides a conceptual framework for understanding how individuals organise their daily activities under constraints aligned with the time geography concept. TASHA [69] creates activity schedules using empirical distributions and heuristic rules. ADAPTS [88] generates activity schedules sequentially, considering long-term decisions like housing, job choices, and vehicle ownership. Although computational process models offer detailed representations of decision-making, they are often highly complex and require extensive data and expertise to define the rules accurately. Additionally, deriving non-linear dependencies can be challenging.

To address these limitations, practitioners have started to use machine learning techniques such as neural networks, support vector machines, and decision trees to extract rules from data in recent applications. These techniques, widely used in other fields for their predictive accuracy, robustness, and flexibility, have shown potential for identifying complex human travel patterns [89]. As a result, they have gained increasing attention over the past decade [90].

ALBATROSS [91] is one of the first implementations of a rule-based machine learning technique. Using decision trees, the model sequentially generates activity schedules. AgentPolis is another example, an open-source simulation framework using neural networks to model dynamic activity-travel behaviour [92]. Hafezi, Liu and Millward [90] identified twelve daily activity pattern clusters using a fuzzy C-means (FCM) clustering algorithm. The study then explored the dependencies between activity schedules and socio-demographic characteristics [93]. Allahviranloo and Recker [94] model individuals' daily activity schedules using support vector machines (SVM). Recently, Data-Driven Activity Scheduler (DDAS) [95] generates activity schedules employing supervised machine learning techniques at various steps.

The use of machine learning techniques makes activity generation relatively easier compared to traditional methods that depend on expert knowledge [95]. However, many machine learning models lack interpretability, often functioning as "black boxes," which makes it difficult to understand the underlying decision-making processes. Moreover, the accuracy of these models remains highly dependent on the availability of reliable data.

Constraint-based, econometric, and computational process methods each offer unique strengths in activity-based travel demand modelling. Depending on data availability, expertise, and project focus, practitioners select the most suitable methods and develop corresponding methodologies. With the increasing availability of data and advancements in computing power, particularly through the incorporation of machine learning techniques, there is significant potential to develop more comprehensive and accurate activity-based travel demand models. This integration could address existing limitations and advance the field by enabling more robust and reliable predictions of travel behaviour. For instance, Paper II explains activity generation for a large-scale agent-based transportation model using the neural networks ML technique. Paper III shows activity-travel pattern generation from big data.

2.3.4 Simulation of agents

Simulation is a crucial component of ABMs, enabling interactions among agents and between agents and their environment. While interactions can occur within the activity-based demand model, such as agents competing for shared resources like household cars, the interactions largely take place in the simulation where agents move through the transportation network.

During the simulation, agents' trips are assigned to a transport network, where they compete with each other across time and space. The simulation outcomes provide detailed information regarding each agent's trips, including route choice and travel duration. These outcomes facilitate the analysis of how individual behaviours aggregate to produce emergent phenomena such as congestion, bottlenecks, and variations in travel times. Consequently, simulation models offer insights into the intricate supply-demand relationships within the transportation system.

Agent-based traffic simulation platforms like MATSim [96] and TRANSIMS [63] are used to simulate agents in transportation networks. These platforms enable the integration of activity-based travel demand models with ABMs, facilitating the study of dynamic interactions between agents and the network.

Multi-Agent Transport Simulation (MATSim)

Multi-Agent Transport Simulation (MATSim) is one of the most commonly used agent-based transportation simulation tools. MATSim can simulate large-scale transportation systems involving extensive networks and populations, making it suitable for projects requiring high computational power. MATSim has been implemented in various recent transportation studies, including analysis of EVs [97], autonomous vehicles [98], car-sharing systems [99].

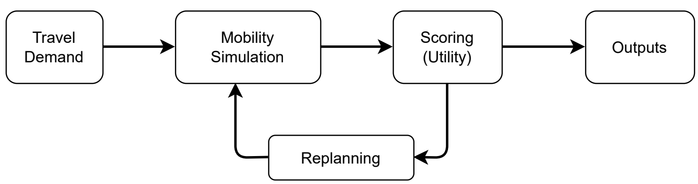


Figure 2.5: Mobility simulation steps by The Multi-Agent Transport Simulation (MATSim) tool. Adapted from the source: Figure 1.5 [p4, 96].

To simulate agents' travel behaviour within a transportation network, MATSim employs a co-evolutionary algorithm. The concept of co-evolution in MATSim refers to the mutual adaptation between agents and their environment, leading to an equilibrium in the transportation system [96]. The algorithm enables the simultaneous evolution of agents' travel plans and the overall state of the transportation system through iterative interactions [100].

The simulation process in MATSim consists of five main steps: initial demand, mobility simulation, scoring, re-planning, and analysis, with an iterative loop between re-planning, mobility simulation, and scoring (Figure 2.5). The simulation begins by feeding the initial travel demand derived from agents' daily activity schedules generated by activity-based travel demand models. Once the initial setup is complete, the iterative process starts. Each iteration begins with mobility simulation, where agents execute their activity schedules through the transportation network and make route choices based on their activity plans. As agents move through the network, they interact with each other and encounter various traffic conditions, influencing their activity and travel times. After the simulation, each agent receives a score reflecting their daily performance. This score is a combination of activity utility, derived from performing activities and travel disutility, derived from the costs associated with travel [101]. In the re-planning phase, a subset of agents is randomly selected to adjust their activity schedules. These agents adjust their plans, including route departure times and transportation modes, based on the simulation setup to improve future scores.

The iterations of simulation, scoring, and re-planning continue until the system reaches termination, where the changes between successive iterations fall below a predefined convergence threshold. At this point, the simulation achieves a user equilibrium, characterised by a stable state where agents' travel plans are optimised, and the overall system reflects a balanced supply-demand relationship [100].

MATSim efficiently simulates individual decision-making and interactions at a granular level, making it a powerful tool for analysing complex transportation scenarios. Its open-source nature and modular architecture allow for customisation by various transportation studies. However, the software is large and complex, with details that can be difficult to understand for users with limited programming knowledge. The calibration and validation of the model can require detailed input data, such as individual activity plans and high-resolution network information, which may not always be available. Furthermore, the stochastic nature of the simulation can result in small variability in outcomes between runs, necessitating multiple iterations to achieve stable results [96].

CHAPTER 3

Included papers

This chapter provides an overview of the papers included in the thesis. Section 3.1 introduces the Synthetic Sweden Mobility (SySMo) model through Paper I and Paper II. Paper I outlines the SySMo model's methodology and presents its outputs, including a synthetic population which is available on the internet to support open science initiatives. Paper II focuses on the methodological advancements of the SySMo model, particularly its heterogeneous activity generation module.

Section 3.2 explores the integration of emerging data sources into activity-based models. In this context, Paper III introduces a novel generative model that synthesises activity-travel plans using mobile phone application data and compares its performance with similar activity-based models. Then, section 3.3 showcases the practical applications of these models in promoting sustainable transportation systems. Paper IV applies the SySMo model to evaluate the potential of e-bikes to replace passenger car trips and reduce greenhouse gas emissions.

3.1 Synthetic Sweden Mobility (SySMo) model

The Synthetic Sweden Mobility (SySMo) model is a large-scale agent-based transportation framework that employs an activity-based travel demand modelling approach. By providing a Swedish synthetic population with detailed activity-travel schedules, the model serves as a test bed for simulating and analysing emerging transportation technologies.

3.1.1 A synthetic population of Sweden: datasets of agents, households, and activity-travel patterns (Paper I)

Highlights

- The first publicly available nationwide synthetic population for Sweden, addressing a significant data gap.
- The model generates a synthetic population for the Swedish population, incorporating advanced socio-demographic attributes.
- The model assigns travel demand for each individual with high spatial and temporal resolution, enabling detailed transportation analyses.
- Publicly available model outputs facilitating the use of agent-based transport simulations by other researchers.

Introduction

A synthetic population is a crucial component for developing Agent-Based Models (ABMs), providing a statistically accurate representation of a population in time and space. Synthetic populations have been widely utilised in ABMs within transportation research and are also increasingly applied in analysing population energy demand, land use, and modelling disease spread [14]. They allow modellers to analyse trends and test different scenarios while preserving individual privacy.

Several countries, such as Switzerland [12], Denmark [13], and France [14], have developed comprehensive synthetic populations. However, Sweden lacks a synthetic population representing its entire population, limiting the ability to conduct detailed simulations. To address this gap, we developed the Synthetic Sweden Mobility (SySMo) model, a large-scale transportation model that generates a synthetic replica of over 10 million Swedish individuals (agents), including their household characteristics and detailed activity-travel plans.

This paper presents the SySMo model's methodology and describes the generated dataset. To create a synthetic population with attributes associated with human transport behaviour, we employ state-of-the-art techniques, including machine learning (ML), iterative proportional fitting (IPF), and probabilistic sampling. The model first generates a synthetic replica of the Swedish population by incorporating socio-demographic attributes such as age, gender, civil status, residential zone, personal and household income, car ownership, household size, number of children ≤ 6 years old, employment status, and student status. It then generates activity-travel patterns using an activity-based travel demand generation approach. These activity-travel schedules include activity types, start and end times, durations, sequences, locations, and travel modes between activities.

By providing a detailed, disaggregated representation of the Swedish population with travel behaviour, the SySMo model enables decision-makers and researchers to evaluate policy-relevant questions. It employs an agent-based modelling framework equipped with an activity-based travel demand generation approach to realistically model individuals' average weekday activity schedules. For example, Liao et al. [97] use the synthetic population to assess the impact of charging behaviours on battery electric vehicle infrastructure in Sweden. By making the model outputs accessible, we aim to contribute to open science and enhance research transparency. Furthermore, since the SySMo model leverages data commonly available in many countries, such as census data and national travel surveys, its methodology can be applied to other regions with similar datasets.

Methodology

The *Synthetic Sweden Mobility (SySMo) Model* consists of three key components: population synthesis, activity generation, and location and mode assignment. Figure 3.1 illustrates the model workflow describing each component and its sub-components.

The first component is *population synthesis*, where all agents are generated in three sub-steps. The process starts with the generation of agents with basic attributes such as age, gender, civil status, residential zone (DeSO), household size, and the number of children<6) using the Iterative Proportional Fitting (IPF) technique (see details in Subsection 2.3.2). Subsequently, we create households based on agents' age, civil status, and household size attributes. We then add the advanced attributes to agents such as employment and student statuses of agents, car ownership, and personal income. To maintain the correlations between the attributes, we use a novel method that combines ML, IPF technique, and probabilistic sampling.

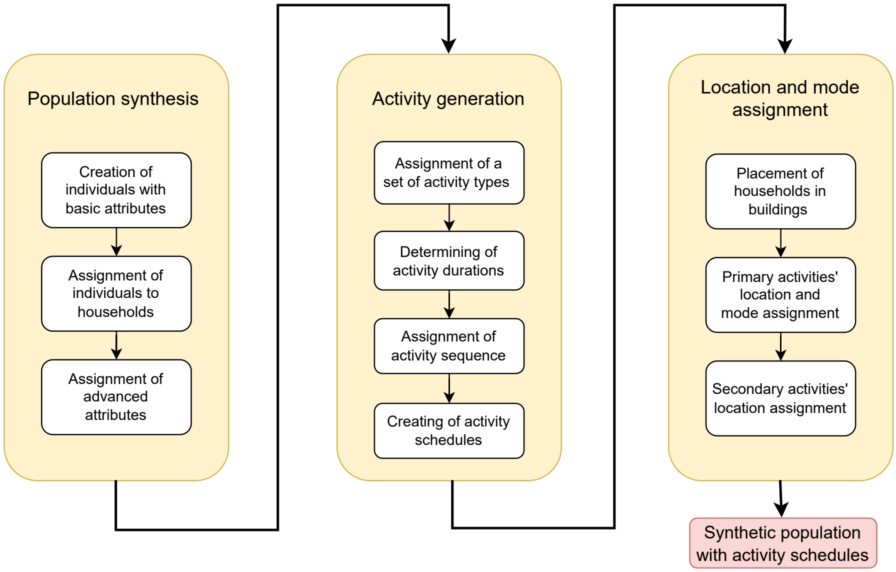


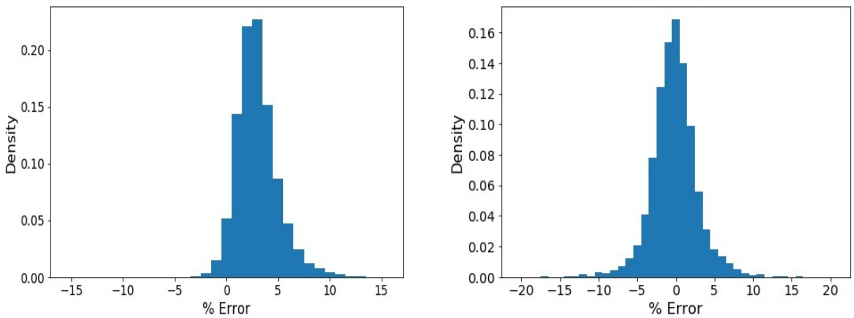
Figure 3.1: Methodology overview of *Synthetic Sweden Mobility (SySMo) Model*. Yellow rectangles: three main components of SySMo model with sub-components; pink rectangle: the final output, a spatially explicit agent-based mobility model.

Activity generation component creates travel demand using an activity-based approach for the population using ML techniques as explained in Subsection 2.3.3 of Chapter 2. The component assigns an activity schedule for each agent using the Swedish national travel survey. These activity schedules are characterised by activity sequences, types, durations, and start and end times. We first determine agents' daily activity participation by home, work, school and other activity types. Then, we jointly calculate the total daily duration of the activity types in which the agents participate. Next, we assign an activity sequence to each individual by matching them with a person from the travel survey based on similarities in their attributes and activity type durations. Finally, we create activity schedules for each agent using the calculated durations and assigned activity sequences.

The *location and mode assignment* component assigns locations to all activities in the schedules and determines the travel modes between them, such as walking, biking, driving, and public transport. We first spatially place each household in a residential building, classified as either a detached house or an apartment building. Next, we assign the locations of primary activities, such as home, work, and school and the travel modes using Origin-Destination (OD) matrices from Trafikverket's (the Swedish Transport Administration) Sampers model or a variant of the gravity model developed for SySMo based on the Swedish national travel survey. Finally, we assign locations for secondary activities, whose positions depend on the locations of the primary activities, using the gravity model informed by the national travel survey.

Results

This section briefly presents the assessment and validity of the SySMo model's results for each model component that generates the synthetic population data. The results of the activity generation component are presented under Section 3.1.2 (Paper II).



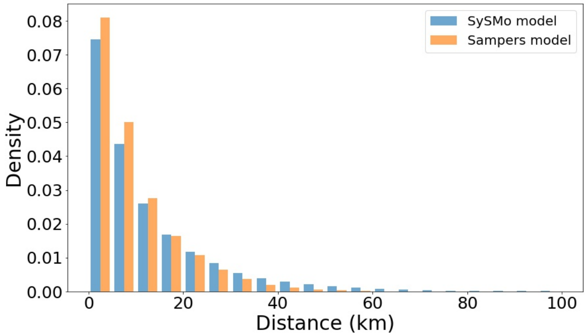
(a) Employees **(b)** Cars

Figure 3.2: The percent error in the number of employees in each DeSO zones(a) and the percent error in the number of cars in each DeSO zones(b).

First, we compare the synthesised population to official statistics from Statistics Sweden (SCB) [102] at the DeSO zone level, where the average population is 1,706 individuals per zone. We calculate the percentage differences for both basic attributes, such as gender and age and advanced attributes, such as employment and car ownership. We find that in over 92% of DeSO regions, the percentage difference in the number of people assigned an incorrect gender by the statistical data is a range between 0.5% and 0.5%. Similarly, more than 78% of

DeSO zones exhibit age group percentage differences between –1% and 1%.

(a) Car



(c) Public transport

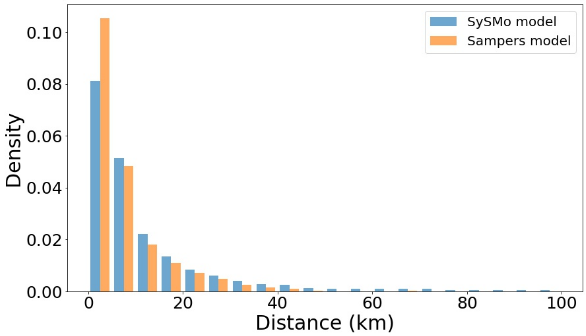


Figure 3.3: Daily travel distance distribution between home and work by travel modes in Vastra Gotland.

Figure 3.2a shows the percentage error in the number of employees per DeSO zone, with errors between –3% and 3% in over 55% of the zones. Figure 3.2b displays the percentage error in the number of cars per DeSO zone, with errors ranging from –3% to 3% in over 76% of the zones. As these are advanced attributes (i.e., derived from basic attributes), the errors are slightly higher than those for basic attributes. These evaluations demonstrate that the attribute distributions of the synthetic population closely align with the official statistical data.

We also assess the performance and validity of the synthetic individuals' travel patterns by comparing them with other models, such as Sampers [103] and a model from Trafikanalys [104]. Figure 3.3 illustrates the distribution of travel distances between home and work locations for car and public transport travel modes in both the SySMo model and Sampers's west regional model. To compare the synthetic individuals' travel pattern with the Sampers model, we calculate the spherical travel distances between activity locations for modes including car, car passenger, public transport, bike, and walk. Although there are slight differences in the peak values, the overall distributions closely align with the Sampers model. The evaluation results indicate that the activity-travel patterns in the synthetic data reasonably approximate other models' patterns.

3.1.2 Heterogeneous activity generation for a large-scale synthetic population (Paper II)

Highlights

- A novel methodology combining probabilistic neural networks and random sampling to generate realistic daily activity schedules for a synthetic population.
- The approach captures heterogeneous activity patterns within sub-populations, ensuring diverse mobility plans based on socio-economic attributes.
- The methodology is applied to a large-scale synthetic population in Sweden.
- The model performance is evaluated by comparing its output with the national travel survey data.

Introduction

Agent-based modelling (ABM) framework equipped with an activity-based travel demand generation approach is a widely adopted method for modelling travel behaviour. Buliung et al. [46] claim that the activity-based approach provides a rigorous perspective for generating travel demand within transportation models. However, proper implementation of the activity generation component still plays a crucial role in developing a model that accurately represents the population's mobility pattern.

Emerging big data sources and increased computational power have accelerated the development of activity-based models, improving their temporal and spatial resolution and behavioural realism [68]. These models increasingly incorporate machine learning (ML) techniques to model human mobility behaviour, as detailed in Subsection 2.3.3 of Chapter 2. Although ML-equipped models can capture the overall distribution of travel patterns (e.g., [91, 93–95]), they often capture limited heterogeneity in travel patterns within sub-populations. Additionally, data limitations and high computational demands typically restrict these models to small-scale populations. The lack of representation and the assumption of homogeneity may hinder insights into human travel behaviour and the effects of policy interventions targeting significant behavioural changes. Achieving a sustainable transport system requires models that accurately represent behavioural characteristics, thereby supporting informed policy-making.

This study introduces a novel methodology for generating daily activity schedules, including activity type, start and end times, duration, and sequence for a synthetic population in Sweden. By integrating probabilistic neural networks (NN) with random sampling, our approach captures the heterogeneity in activity generation, producing realistic daily mobility plans for individuals. This combination of machine learning and probabilistic sampling ensures accurate distributions while reflecting diverse behaviours. This paper outlines the methodological framework for travel demand generation in the Synthetic Sweden Mobility (SySMo) Model [105], a unique large-scale mobility model of the Swedish population across various travel modes. Our study advances machine learning research in transportation modelling and addresses the lack of large-scale synthetic populations with heterogeneous activity patterns.

Methodology

The proposed activity generation framework comprises four main steps: assigning activity types, determining the duration of each activity, sequencing the activities, and creating individual activity schedules. This framework is applied to the Swedish synthetic population created within the SySMo model. Figure 3.4 illustrates the proposed workflow, detailing each step involved in generating travel demand.

The methodology begins by *assigning daily activities* to each individual, including home, work, school, and other activities such as visiting shops and restaurants. We utilise NN classifier, trained on

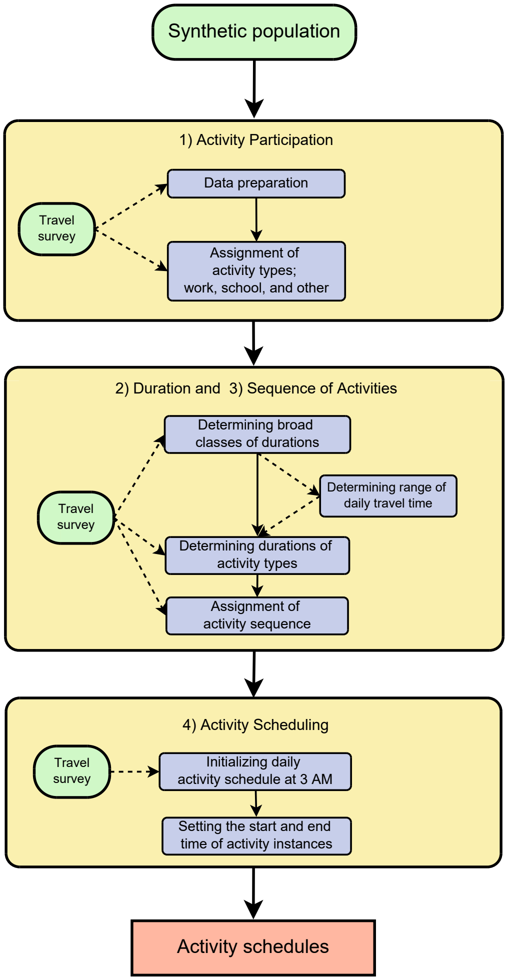


Figure 3.4: Methodology overview of the activity generation module of *Synthetic Sweden Mobility (SySMo) model.* Yellow rectangles: major steps of the activity generation; purple rectangles: steps of the calculations; green ellipses: input data; pink rectangle: model outputs of activity schedules for each individual.

data from the Swedish national travel survey, to predict an individual's participation in each activity type based on their socio-economic attributes. We assume that each individual visits home at least once per day. Next, we calculate the total *daily duration for each activity* type using a two-step process that integrates NN classifiers with sampling techniques. In the first step, we jointly deduce broad duration classes for the different activity types, capturing the correlation between their durations. Using these broad classes and the individuals' attributes, we estimate the range of total daily travel time. In the second step, we sample specific durations for each activity type for each agent, ensuring that the sum of activity durations and the estimated daily travel time does not exceed 24 hours.

The third step assigns an *activity sequence* to each individual by matching them with respondents from the national travel survey based on their socio-economic attributes and travel times. Statistical matching, commonly used in activity-based modelling, defines daily activity patterns for synthesised individuals by leveraging their similarity to survey participants [14]. In the final step, we calculate the duration of each activity within the assigned sequences and *create activity schedules* for each individual, showing the activity type, sequence, and start and end times.

Results

We evaluate the proposed model's performance by comparing its outputs with underlying travel survey data and present the model's outputs.

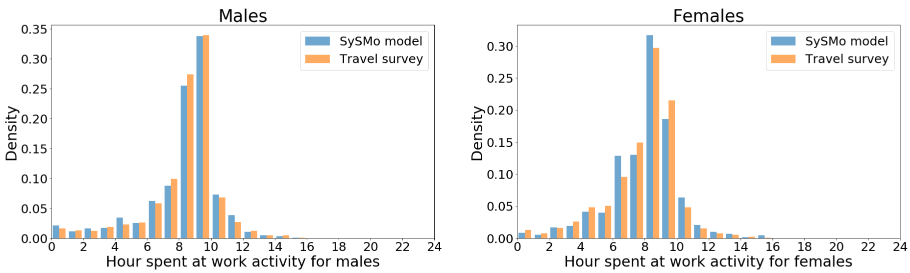


Figure 3.5: Comparison of work activity duration by gender. The left panel shows the number of hours spent on work activity for males (JS distance=0.05) and the right panel females (JS distance=0.08).

We assess the model's performance by comparing its generated activity participation, duration, and temporal profiles with data from the national travel survey. To quantify the similarity of distributions, we calculate Jensen–Shannon (JS) distances [106]. JS distances take values in the range [0,1], where 1 denotes the maximum distance between the distributions. Figure 3.5 presents the work activity duration distributions by gender. The JS distance for work activities is 0.05 for males and 0.08 for females, indicating a close alignment with the survey data. For school, home, and other activities, the JS distances range from 0.05 to 0.13, demonstrating that the model accurately captures activity durations across different genders and activity types. We repeat comparisons for each activity type and attribute. The results demonstrate the model maintains correlations between socio-demographic attributes and activity patterns.

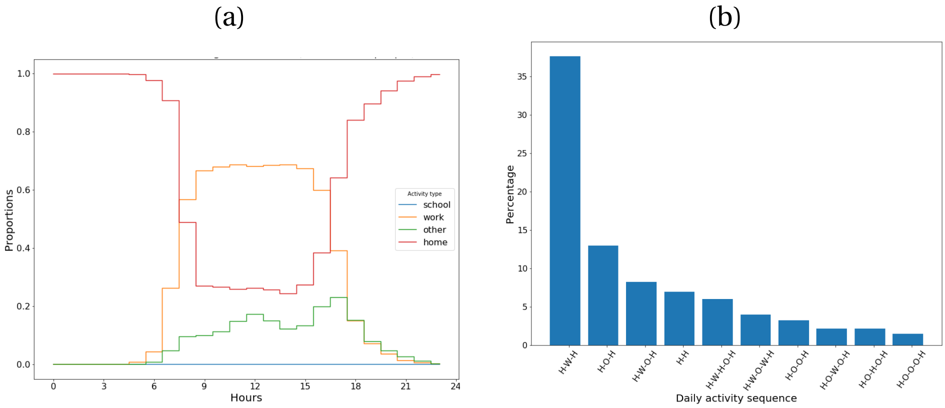


Figure 3.6: Activity pattern of the synthetic agents; aged 40-45, male, married, employee, in high-income class, no children $\leq$ 6 years old in household, and no car in household, residing in Stockholm. (a): Aggregated activity pattern of the sub-population by activity type, (b): Percentage of 10 most frequent daily activity sequences in the sub-population (26 thousand agents in total).

We also present model results demonstrating the heterogeneity in sub-populations' activity patterns. Agents' activity schedules are generated using a probabilistic methodology influenced by their socio-demographic characteristics, allowing two agents with identical attributes to have different activity schedules. Figure 3.6 illustrates this heterogeneity by focusing on a specific sub-population: individuals aged 40-45, male, married, employed, with high income, no children under six years old, no car in the household, and residing in Stockholm. Figure 3.6a shows the individuals' activity participation pattern by activity types and time of the day. Individuals' activity patterns show variability in this sub-population. Figure 3.6b shows the ten most common daily activity sequences' frequencies in the sub-population.

While the H-W-H sequence is dominant, it accounts for less than half of the agents within the population.

3.2 Activity plan generation with emerging data sources

This section explores the integration of emerging data sources into activity-based models. The rise of big data sources, e.g., call detail records (CDRs), location-based social networks (LBSNs), smart-card transactions, and mobile phone application data, presents an alternative for examining mobility. These sources can offer extensive geographical and population coverage with an extended observation period (months to years) compared to traditional methods (one day to weeks). Numerous studies show their benefits in transportation studies, including the detection of commuting patterns [107, 108], flows between locations [109, 110], and travel modes [111, 112]. However, these data sources' diverse characteristics and formats necessitate the development of specialised model features, and the ongoing research aims to better utilise these data in transportation studies.

3.2.1 Mobile Phone Application Data for Activity Plan Generation (Paper III)

Highlights

- A generative model synthesising individuals' activity-travel plans from geographically and population-wise extensive but individually incomplete mobile phone application data.
- The model introduces a temporal-score approach to accurately infer primary activities, e.g., home and work, enhancing state-of-the-art methods.
- The model generates multiple weekday activity schedules, including activity sequences, types, start/end times, and locations, while incorporating daily variability.
- The generated activity-travel plans are assessed against an existing large-scale agent-based model of Sweden (SySMo) and a dummy model using only mobile application data.

- The generative model provides realistic activity-travel plans on key mobility metrics by careful assessment.

Introduction

Activity-based models in transport provide a comprehensive and realistic understanding of individuals' activity-travel patterns, improving demand forecasting, policy analysis, and land use and transportation planning. While travel surveys have long served to develop activity-based models with complete activity-travel plans, they are often costly to collect and have small sample sizes. Mobile phone application data, one example of emerging mobility data sources, offers an alternative with broader geographical and population coverage over extended periods. This data is collected by capturing phone users' geographical locations with their consent as they interact with various mobile applications. The challenges of using these data include sampling biases in the population coverage and data sparsity at the individual level due to intermittent and irregular data collection.

Mobile phone application data have been utilised in trip-based analysis and population models, such as creating origin-destination (OD) matrices to describe mobility flows within specific geographic areas [113, 114]. However, its application in activity-based models remains under-explored. A few studies [115–117] using similar data sources, like call detail records (CDRs), have shown these datasets' feasibility to simulate individuals' activity travel schedules. However, they often overlook data sparsity, which impedes accurately capturing representative activity-travel patterns. A notable gap is the lack of a generative model transferable to similar big data sources.

In this study, we propose a novel model that combines mobile phone application data with travel survey data to synthesise activity-travel plans, addressing the limitations of each data type while leveraging their strengths. Our generative model simulates multiple average weekday activity schedules for over 263,000 individuals living in Sweden, approximately 2.6% of Sweden's population. These schedules include activity sequences, types, start/end times, and locations, incorporating daily plan variability. In this model, we propose a temporal-score approach to improve the state-of-the-art home and work location identification approaches, among other designs, for synthesising realistic activity-travel plans. We assess the model's performance against an existing large-scale agent-based model of Sweden (SySMo)[105] and a dummy model using only mobile application data. The proposed model is adaptable to other regions with similar travel surveys and emerging data sources, like call detail records, advancing the use of these data for activity-based models in a cost-effective, easily updated manner.

Methodology

The proposed generative model includes five modules: mobile GPS records processing, identifying primary activities, population debiasing, twin travellers searching, and synthesising activity-travel plans (Figure 3.7).

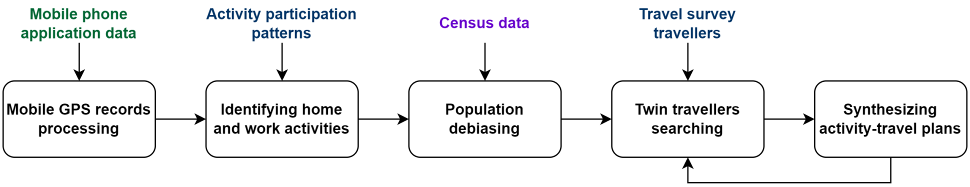


Figure 3.7: Generative model for synthesising activity-travel plans. Travel survey [118] inputs are marked with blue. Census data refer to the statistics related to Demographic statistics areas (DeSO zones) [119].

The **data preparation module** extracts building-level activity records from mobile phone GPS records, providing a glimpse into individuals' activity participation at various locations. The second module identifies individuals' **primary activity locations**, e.g., home and work, essential for generating daily activity schedules [91, 120]. We employ a temporal scoring approach that considers the varying likelihood of presence at an activity location across different hours of the day, unlike traditional methods [121, 122] that rely solely on temporal rules. This approach calculates scores for each location based on the duration of an individual's stay, with weights derived from the ratio of hourly primary activity participation to overall activity participation in the travel survey. We identify home activity locations with the highest score considering building types and being at least visited three times during the data-informed nighttime (6 pm to 7:59 am). A similar approach is used to identify work locations.

The **population debiasing module** addresses the inherent sampling bias and ensures a representative sample of the Swedish population. We calculate weights for each individual in the mobile phone dataset using population and employee size statistics from their residential zones.

In synthesising activity-travel plans, the **twin travellers searching module** establishes matching probabilities between individuals in the mobile data and travel survey participants based on their demographics, activity patterns, and trip attributes. For each individual in the mobile phone dataset, we sample one twin traveller from the travel survey candidates using matching probabilities. Using the selected twin travellers' daily activity sequence and individuals' activity locations in the mobile phone dataset, the final module **synthesises activity-travel schedules** for average weekdays. We assign the inferred primary activity locations to the corresponding activity schedules. We then determine other activities' locations in the schedule by sampling from individuals' activity locations in the mobile phone dataset, considering the distances to the individuals' home locations.

The proposed model can synthesise multiple activity-travel plans for each agent, producing daily variations in individuals' activity-travel patterns. By iteratively finding twin travellers and activity locations, the model generates multiple average weekday simulations for each individual in the mobile phone dataset.

Results

The proposed model generates over 263,000 people's daily activity-travel plans using geolocation data from their mobile devices. These plans include an average of 63,000 work activities, 612,000 home activities, and 436.000 other activities, each with unique and widely spread geolocations across Sweden.

We evaluate the performance of the generative model by comparing its produced activity-travel schedules with those from the SySMo model and a dummy model. As presented in Paper I, SySMo is a large-scale, agent-based transportation model for the Swedish population. To provide a baseline for comparison, we also developed the dummy model, utilising only mobile phone datasets to generate average weekday activity-travel schedules.

The performance evaluation focuses on the key aspects of activity-travel plans, e.g., daily activity sequence distribution, temporal activity engagement patterns, and trip distance distribution. The generative model produces an activity sequence distribution similar to the SySMo model and the underlying travel survey data. In contrast, the dummy model fails to capture the sequence distribution fully compared to these data. For example, Figure 3.8 shows the temporal distribution of activity participation by activity type during the day, comparing results across the models. While hourly activity participation patterns are quite similar in the generative and SySMo models, the dummy model deviates from a realistic pattern revealed by the other models.

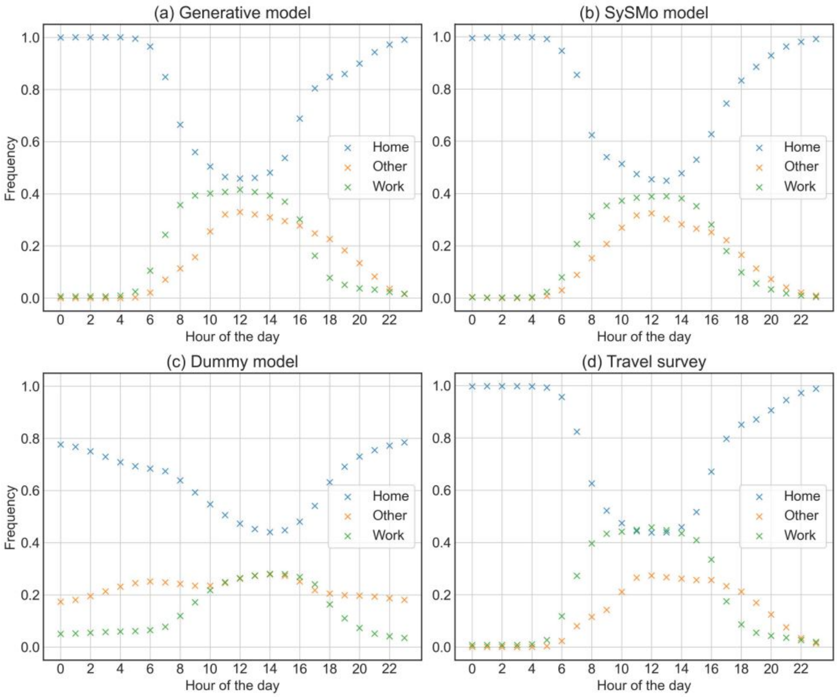


Figure 3.8: Aggregated temporal patterns of activity engagement by activity type: (a) Generative model, (b) SySMo model, (c) Dummy model, (d) Travel survey.

The comparison results highlight the generative model's capability in synthesising realistic activity sequences, whereas the dummy model fails to produce distinct temporal patterns between the three activity types. Moreover, the generative model's results align with SySMo and the underlying survey data. These findings suggest that the proposed model effectively addresses biases and sparsity in mobile phone application data, resulting in more realistic and reliable activity-travel plans than directly using the raw mobile data.

3.3 Model applications

Agent-based transportation modelling frameworks with activity-based travel demand models like the Synthetic Sweden Mobility (SySMo) model realistically simulate individual travel behaviours and activity patterns, supporting policymakers to develop and promote sustainable transportation systems. These models allow for evaluating transportation policies by simulating their potential impacts, such as assessing the necessary infrastructure to support electric vehicle adoption or exploring the feasibility of replacing traditional travel modes with sustainable alternatives. By providing detailed simulations, ABM contributes to informed decision-making on environmental sustainability goals, e.g., emission reduction and transportation efficiency improvement.

3.3.1 Potential of e-bikes to replace passenger car trips and reduce greenhouse gas emissions (Paper IV)

Highlights

- Proposes an innovative approach leveraging a synthetic population with detailed daily activity-travel schedules to evaluate the potential of e-bikes to substitute passenger car trips in Sweden's Västra Götaland region.
- Identifies replaceable car trips by incorporating socio-demographic factors such as age and gender, daily activity patterns, and the availability of biking infrastructure.
- Demonstrates that e-bikes could replace 57.6% of car trips, potentially reducing greenhouse gas emissions from passenger vehicles by up to 22.8%.
- Provides valuable guidance for policymakers and urban planners regarding the benefits of e-bike adoption in reducing emissions.

Introduction

In Sweden, like other developed economies, the transport sector is responsible for a large proportion of carbon emissions at 32%, and private cars within the transport sector contribute to 62% of these emissions [2]. In this context, electric bikes, commonly known as e-bikes, have emerged as a promising solution for reducing carbon emissions in the transportation sector. This paper explores the potential of e-bikes in substituting private car trips and reducing carbon emissions. To achieve this objective, we use a synthetic population with daily activity schedules in the Västra Götaland (VG) region, Sweden, developed in Paper I and Paper II. For assessing the potential for e-bike substitution, the current literature often uses trip-level data, which does not adequately consider people's daily trip chains [123, 124], resulting in an unrealistic estimation of replaceable trips and their carbon emissions reduction. Our simulation identifies potential car trips that can be replaced with e-bikes, considering individuals' activity travel plans, using the time geography concept explained in Subsection 2.1.2. The simulation results show that e-bikes can achieve a reduction of 22.8% in carbon emissions from car trips taken by the VG region residents. Furthermore, this paper also analyses population groups providing the highest substitution rates of e-bikes in their daily activity schedules and shows the spatial distribution of carbon emissions reduction from e-bike adoption. In areas with a high population density, substitutable car trips are more common than in rural areas. This research provides valuable insights into the potential of e-bikes in reducing carbon emissions. It contributes to the existing literature through its modelling approach that considers individuals' socio-demographic characteristics and daily activity schedules when assessing the substitution potential.

Methodology

This section explains the input data set and the key modelling concept to calculate the potentials of e-bikes (Figure 3.9). The study adopts a constraint-based modelling approach, as explained in Subsection 2.3.4. We also use the synthetic population with daily activity schedules generated by the Synthetic Sweden Mobility Model (SySMo) [105] and simulated using MATSim [97, 125] (refer to the agent-based simulation subsection 2.3.4 for details). The agents in the population perform trips between activities using different travel modes, e.g., car, car passenger, public transport, bike, and walking. However, we focus on individuals with car trips in their daily schedule and residing in the VG region, which includes Gothenburg, Sweden's second-largest city. The input data used in the study contains 284,000 car agents, representing 35% of all car users and 18% of the total population in the region.

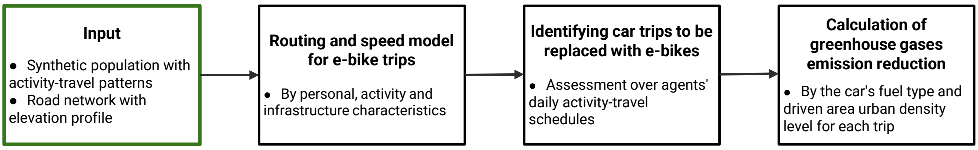


Figure 3.9: Methodology overview of the study showing carbon reduction capability of e-bikes through substituting to car trips in the Västra Götaland region.

Routing and speed model for e-bikes

The study's methodology starts by assigning an e-bike to all car trips in the agents' daily travel plans and simulating e-bike trips on the road network. We use a routing engine, open trip planner (OTP) [126] and integrate an e-bike cycling speed model into the simulation. OTP realistically performs routing of e-bikes on the given network while considering the bike-friendly facilities (e.g., separated bikeways, low-traffic streets, etc.). Using OTP, we extract the movement trajectories and travel distances of all e-bike trips. We then apply the e-bike cycling speed model to the e-bike movement trajectories to calculate the trips' travel duration. To develop the e-bike cycling speed model, we consider demographic characteristics (age and gender), activity type (work or other), and infrastructure characteristics (road gradient) using two relevant empirical studies [127, 128].

Substitution of car trips

To identify the car trips that e-bikes can potentially replace, we apply an algorithm on the agents' re-calculated activity-travel schedules using e-bikes. The algorithm incorporates constraints on individual trips, activities, and tours. We define tours as trip chains that initiate or terminate at home locations. The algorithm first checks all the trips of an individual against a defined threshold of maximum cycling distance and a maximum delay of the arrival activity. Next, the algorithm focuses on tours. For a car trip within a tour to be replaceable by an e-bike, each trip must separately satisfy the aforementioned criteria. Last, the algorithm assesses the cumulative impact of these substitutions on total daily travel duration and travel distance by e-bike, comparing them against specific thresholds.

Emission reduction

Following identifying potential e-bike trips, we calculate greenhouse gas emissions from passenger cars for both the baseline scenario and the scenario where e-bikes replace private cars. Our study focuses solely on tank-to-wheel (TTW) emissions. To determine reduced emissions, we first assign a passenger car with a specific fuel type to each agent using car fleet distribution data by fuel type and municipality. Given that diesel cars have higher mileage than other fuel types [129], our approach increases the probability of assigning diesel to vehicles exceeding the municipality's median travel distance. Subsequently, we calculate the average TTW emission for each trip. This calculation uses data that provides average emission factors for passenger cars, categorised by fuel type and urban density level. Finally, we calculate the reduction in passenger car emissions and present the spatial distribution of carbon emissions reduction for residents in the VG region.

Results

The study demonstrates the potential of e-bikes to replace passenger car trips using an algorithm that incorporates constraints at the individual trip, tour, and daily plan levels. At the individual trip level, our analysis reveals that e-bikes could replace up to 72.4% of all car trips. However, when evaluating the substitution at the daily plan level, the potential for e-bike substitution is reduced to 57.6% of all car trips. This level of substitution results in a significant reduction of greenhouse gas emissions, with the potential to lower emissions by up to 22.8% from passenger car trips in the study area.

In addition, we examined the spatial distribution of greenhouse gas emission reductions in the Västra Götaland (VG) region. Figure 3.10 shows the distribution of the reduction in greenhouse gas emissions by the agents' residential DeSO zones. The results indicate that the emission reduction potential is higher in urban areas compared to rural areas, reflecting the increased feasibility of replacing car trips

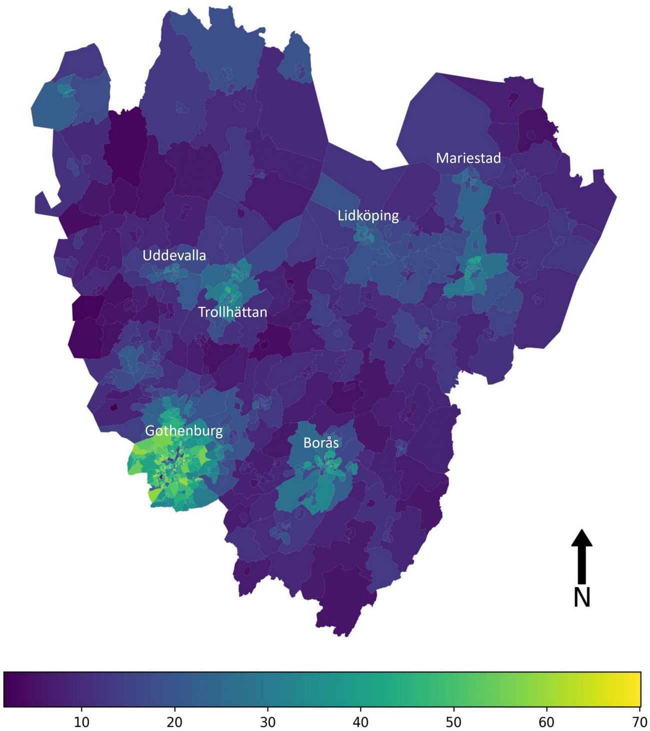


Figure 3.10: Reduction in greenhouse gas emissions from passenger cars by the agents' residential DeSO zones (in %).

with e-bikes in densely populated areas.

CHAPTER 4

Discussion and outlook

This thesis explores transportation modelling tools with a particular emphasis on individual-level approaches, introducing two transportation models: the Synthetic Sweden Mobility model (Paper I-Paper II) and the Mobile Phone Application Data for Activity Plan Generation (Paper III). The first model is abbreviated as SySMo, while the second model is called the Generative model. These models employ agent-based methodologies utilising activity-based travel demand generation to simulate realistic and detailed representations of individual mobility patterns.

The main contributions of this thesis are:

- **Development of large-scale agent-based transportation models for Sweden:** Paper I-Paper II present the SySMo model providing the Swedish synthetic population with various socio-demographic variables and activity-travel schedules across Sweden. Paper III proposes a generative model to synthesise activity-travel plans for individuals in mobile phone data. These models address Research Questions 1-2, introduced in Chapter 1.

- **Integration of advanced techniques in computer science and emerging big data sources:** The SySMo model uses a novel methodology to generate activity schedules by combining probabilistic machine learning techniques with sampling techniques, as demonstrated in Paper II. Additionally, Paper III explores the use of mobile phone application data, a relatively new big data source in activity-based travel demand generation. These methodologies address Research Question 3 in Chapter 1.

- **Evaluation of sustainable transportation systems:** The developed models serve as tools for assessing sustainable transportation vehicles. Paper IV exemplifies the model application by assessing the potential of e-bikes to reduce emissions. This study addresses the main research question in Chapter 1.

- **Creation of publicly available data repositories for transportation analysis:** The developed models are published in publicly available repositories containing model outputs, enabling researchers to evaluate sustainable transportation systems.

Building upon these contributions, this chapter provides an overview of the developed models, reviews my PhD research, and outlines potential future research directions. Section 4.1 discusses the potential and limitations of the developed models and compares them with each other. Section 4.2 reflects on the open science concept and emphasises its integration within my research. Section 4.3 presents various applications of the models, demonstrating their impact. Section 4.4 outlines future research directions. Finally, Section 4.5 includes my reflections on the research.

4.1 Strengths and limitations of the models

The models presented in Paper I-Paper II, and Paper III employ an agent-based approach to generate realistic and detailed individual mobility patterns. While these models demonstrate valuable capabilities, they are inherently simplified representations of reality, built on specific assumptions, and thus have their strengths and limitations. Recognising these strengths and limitations is crucial for effectively applying the models to transportation challenges. This section first compares the adopted methodology with other modelling approaches in the literature and then compares the models based on their main components outlined in Section 2.3 of Chapter 2, as well as their performance assessment and reproducibility to present the models' capabilities. Table 4.1 summarises the developed models.

Models compared to other modelling frameworks

The developed models are well-suited for sustainable transportation scenario analysis, providing individual-level analysis with a high spatio-temporal resolution. Modern transportation systems involve complex, large-scale interactions among travellers and their environments which include infrastructures, modes, services, and technologies [130]. Within the system, individuals' activities and trip decisions are influenced by various spatial, social, and economic factors, making individuals' behaviour heterogeneous and challenging to predict.

To address these challenges, the developed models employ an agent-based approach combined with an activity-based travel demand generation approach. ABMs are well suited to address the complexities of transportation systems through individual-level modelling (as detailed in Section 2.3 of Chapter 2). Activity-based travel demand generation complements this approach by integrating the behavioural dynamics that drive travel demand (as detailed in Section 2.3.3 of Chapter 2). In contrast, other modelling frameworks, like four-step transportation models, often fail to capture the complexities of transportation systems and to provide behavioural insights. By focusing on aggregated travel behaviours, these models inadequately reflect individual decision-making processes or the interrelationships among individuals' activities and trip chains. This inability to simulate individuals' daily travel behaviours and interactions limits their effectiveness in accurately assessing sustainable technologies. For example, Paper IV, using SySMo outputs, evaluates e-bikes' potential by positioning e-bikes within individuals' daily activity-travel plans through an agent-based approach and highlights the limitations of trip-level analyses.

However, the employed framework also presents certain limitations despite its significant advantages in capturing individuals' patterns. Capturing the details of human decision-making necessitates data on individual travel behaviours, preferences, and interactions, which can be difficult and costly to obtain. Simulating numerous agents requires significant computational power, especially for large-scale networks or extended simulation periods. Furthermore, ensuring the model's accuracy in representing both micro-level behaviours and macro-level patterns can be challenging due to the complexity of calibration and validation processes.

Population synthesis

The SySMo model generates an advanced synthetic population with various socio-demographic attributes using statistical data. This syn-

Table 4.1: Summary of Synthetic Sweden Mobility (SySMo) and the Mobile Phone Application Data for Activity Plan Generation (Generative) models.

	SySMo model	**Generative model**
Population synthesis	Generates over 10 million agents with diverse socio-demographic attributes using census and statistical data.	Captures over 263,000 individuals with their daily mobility from mobile data.
Travel demand generation	Assigns heterogeneous average weekday activity-travel plans by combining machine learning (ML), iterative proportional fitting (IPF), and probabilistic sampling.	Synthesises multiple average weekday activity-travel plans with daily variations by combining massive mobile phone GPS records with the national travel survey data.
Simulation	Uses fixed activity and travel plans without feedback loops.	Uses fixed activity and travel plans without feedback loops.
Performance assessment	Model outputs were compared with the underlying survey data.	Model outputs were compared with the underlying survey data and similar activity-based models.
Reproducibility	Limited reproducibility with comprehensive documentation.	High reproducibility with software available on GitHub and detailed methodology published.

thetic population is one of its key strengths, providing valuable inputs to simulation models, especially ABMs, in research areas such as transportation, land use, economics, and epidemiology. Unlike previous studies that are often limited to smaller regions, such as Stockholm [131], SySMo offers a statistically accurate representation of the entire Swedish population at the DeSO zone level. This large, attribute-rich population enables flexible scenario generation. For example, researchers can compare scenarios where electric vehicles are assigned only to individuals with income levels above a certain threshold against scenarios where electric cars are distributed evenly across all income levels.

The proposed generative model in Paper III, on the other hand, captures over 263,000 individuals with their daily mobility from mobile phone application data. However, these large datasets typically lack detailed socio-demographic information due to privacy concerns, presenting a significant limitation of utilising big data. Consequently, the model provides limited socio-demographic information inferred from the data, such as home locations and employment status. Moreover, these big data sources often exhibit sampling biases. To address these biases and ensure a representative sample of the Swedish population, the model calculates a weight for each individual based on their residential location and inferred employment status. These weights mitigate the inherent sampling bias arising from the spatial distribution of the individuals in the mobile phone dataset across Sweden. This weighting approach is one of the model's methodological contributions.

Travel demand generation

SySMo assigns activity-travel patterns to each individual in the synthetic population using a novel methodology that combines ML, IPF, and probabilistic sampling methods. The proposed methodology trains ML models using the Swedish national travel survey and maintains the correlation between an agent's attributes and mobility patterns. The probabilistic sampling approach introduces heterogeneity into activity-travel patterns, enhancing the realism of the simulations. Generating these realistic patterns enables more sophisticated analyses and targeted policy interventions. For instance, when designing and evaluating time-based congestion charges across different income levels, understanding the variability in travel patterns within each group can help assess both the equity implications and the expected outcomes of the intervention.

The generative model in Paper III synthesises activity-travel plans using abundant but incomplete activity records from mobile phone data. While previous studies using similar big data sources have highlighted their benefits in activity-based modelling studies, they have not adequately addressed data sparsity issues [115–117]. By combining mobile data with a travel survey, the models' methodology addresses mobile data's sparsity and generates average weekday activity schedules for 263,000 individuals in Sweden. Furthermore, the study advances the state-of-the-art primary activity identification approaches by introducing a temporal-score approach. In anonymised mobile phone data, these locations are typically inferred using temporal rules (e.g., the studies by Chen, Bian and Ma [121] and Sadeghinasr, Akhavan and Wang [122]), often failing to account for individual differences in activity participation patterns. This study considers variations in activity engagement at different observation times.

Incorporating a travel mode choices module could further enhance the models' methodologies. While the SySMo model assigns travel modes between activities based on the likelihood of mode usage, predicted by individuals' socio-demographic characteristics, the generative model does not contain travel modes in the individuals' activity-travel schedules. Choice models, particularly logit models, are commonly used to predict mode choices, offering valuable insights into travel behaviour [132] though they have limitations, as discussed under the econometric models in Chapter 2. Integrating a mode choice model into the models enables the prediction of individuals' mode preferences. Recent studies [133, 134] have demonstrated that using mobile phone data, either alone or combined with survey data, for mode choice predictions presents promising opportunities to further enhance the models' capabilities.

One common limitation of the presented models is their generation of average weekday mobility patterns for individuals, which overlooks the day-to-day variability inherent in human travel behaviour. Although the proposed generative model is capable of generating multiple average weekday activity-travel plans with daily variations, it still falls short in accounting for changes throughout the week or during different seasons. At the individual level, the generative model introduces variations across all elements of activity schedules, i.e., activity sequences, type, start/end times, and other activity locations, while maintaining the residential and workplace locations. However, the model does not capture shifts in travel demand on weekends, during vacations, or over longer time frames. Presenting only average weekdays may lead to an underrepresentation of long-distance trips, which are typically more common during weekends and vacation periods.

Extending the analysis to cover activity patterns over longer periods could provide valuable insights into infrastructure planning and reveal the stability or variability of travel behaviour [8]. Addressing this limitation, future research aims to enhance the model to simulate weekly activity patterns, offering a more comprehensive understanding of human mobility.

Agents' simulation

The presented models adopt a structure similar to Figure 2.3c in Chapter 2, where activity and travel plans generated by the demand model are fixed and serve as inputs to the agent-based simulation without a feedback loop. These models capture the agents' interactions and their learning from actions along with scenario analysis during the agent-based travel simulation. We utilise MATSim to simulate agents and capture their interactions by inputting their daily activity schedules and the transportation network. The agent-based simulation can lead to changes in agents' schedules. Although MATSim generates realistic activity-travel plans, simulating millions of agents demands substantial computational resources and requires specialised expertise in programming and modelling within the MATSim framework, posing barriers for some users. Incorporating a feedback mechanism into the developed models' travel demand generation component would further improve the models.

Scenario Analysis

The developed models provide a platform for conducting scenario analyses. Researchers can explore a wide range of hypothetical situations by changing certain aspects of individuals' activity-travel patterns and simulating the resulting changes in transportation systems. These analyses are essential for assessing the impacts of varying parameters, such as technological innovations and policy initiatives, on individual mobility patterns and the overall transportation system. One example could be evaluating the impact of autonomous vehicles (AVs) on conventional transport modes. By simulating scenarios where shared AVs fully replace conventional transport, researchers can estimate the required fleet size and understand potential effects on travel behaviour. Another example could involve evaluating vehicle-to-grid (V2G) systems, where vehicles serve as transportation modes and potential energy sources. By analysing individuals' activity-travel patterns, including parking durations, the model can explore the potential battery capacity available in the vehicle fleet.

However, these models are not capable of predicting individuals' preferences and their response to changes in certain factors like travel time, infrastructure, or mode availability. For instance, an increase in fuel prices or improvements in public transit might lead individuals to switch from private cars to public transport, or the introduction of AVs could significantly shift mode preferences. Incorporating choice models into the developed models would enhance the behavioural realism of the models, enabling predictions of how individuals might adjust their travel choices based on parameters such as convenience, travel cost, and travel time.

Another limitation is the models' applicability for future scenario analysis. The SySMo model generates a synthetic population based on 2018 data, while the generative model generates travel demand using mobile application data from June to December 2019. In a developed country like Sweden, it is reasonable to assume that infrastructure and mobility patterns remain relatively stable over short periods. However, this assumption raises concerns about the models' suitability for long-term evaluations. For example, Sweden's population grew by approximately 15% from 1990 to 2020, and it can be considered that the trend will continue in the following years. There are well-known methods (see more in the papers [13, 59, 135]) to project the population to certain years. Near-future mobility patterns can be predicted using choice models or advanced machine learning techniques with assumptions about economic growth, land use, etc. [136, 137]. However, predicting a substantial shift in the system based on past observations could pose challenges [138]. To improve the models' relevance for long-term planning, future model developments will include population and travel demand projections.

Model performance assessment

The performance of the developed transportation models is evaluated by comparing their results both with each other and with the underlying travel survey data[118]. The performance assessments include key components of the activity schedules, activity sequences, spatiotemporal activity engagement patterns and distance distribution across activity types. Despite using different methodologies and input data, both models generate activity-travel patterns that are similar to each other and closely align with the underlying survey data. However, one major challenge lies in evaluating the disaggregated activity-travel schedules. Many models in the literature face similar issues due to the lack of ground truth data for validation. Despite this study's promising findings, this gap limits our ability to test the model's reliability and accuracy in a precise way.

Reproducibility of the models

This section evaluates the developed models' reproducibility within the scientific computing context. Reproducibility, as defined by Goodman, Fanelli and Ioannidis [139], involves replicating a study's results under comparable conditions. In scientific computing, this means providing sufficient detail so anyone can reproduce the study, including access to all necessary digital artefacts such as source code, input files, and post-processing scripts [140]. Reproducibility is an essential principle of open science, enhancing the reliability and utility of scientific research. It enables other researchers to assess and improve research outcomes and exists on a spectrum, from studies that are difficult or impossible to reproduce to studies that are reproducible.

Reproducibility is closely related to transferability, which means that a model developed for one context, environment, or population can be used in other settings with only minor adjustments. Transferability is crucial for applying models to diverse scenarios but relies fundamentally on the model's reproducibility. Therefore, this section discusses the reproducibility of the models.

SySMo, a large-scale transportation model, employs a complex methodology to generate a synthetic population with an activity-travel plan. To enhance reproducibility, we developed comprehensive documentation [141] that details the methods, assumptions, and evaluation procedures. Additionally, an archive of the software and environmental configurations used in the study is maintained. While this archive ensures that others with the necessary resources can reproduce the study, it is not publicly accessible and is available upon request. Due to the complexity of its methodologies and the availability of the software, the SySMo model achieves a moderate level of reproducibility. This means that the model can be reproduced by its developers and others with guidance.

The model proposed in Paper III achieves a higher level of reproducibility. The study can be reproducible by anyone with internet access and having similar data. I published all model software in a publicly accessible GitHub repository [142] and provided a detailed explanation of the methodology in a journal article. These factors contribute to the model's high reproducibility. By offering both the source code and comprehensive methodological documentation, the generative model enables other researchers to independently assess and replicate the study's findings.

4.2 Open science practices

Open science aims to democratise access to knowledge, reduce disparities in science and technology, and support the human right to participate in and benefit from scientific advancements. It encourages active participation in scientific discovery for both researchers and the general public, offering cultural, social, and economic benefits. By adopting open practices, the scientific process becomes more efficient, transparent, and reliable, strengthening the evidence base for decision-making and fostering public trust in science. Open science practices include publishing open research and sharing data and code. Throughout my PhD studies, I have adopted open science principles in my research.

Access to model outputs through open data repositories allows researchers to build on existing work and further develop their studies. I made the developed models' results publicly available through dedicated data repositories. These repositories facilitate further transportation analysis and foster collaboration within the research community (see the model application SectionPaper sec:applications)). These datasets are valuable for assessing the potential impacts of transportation technologies, infrastructure changes, and policy interventions to enhance system sustainability. Additionally, these repositories strengthen the transparency of the research and encourage collaborative efforts to advance transportation modelling and policy-making.

Implementing open science practices presents certain challenges, particularly concerning data privacy. Big data sources, such as mobile phone data, inherently raise concerns about individual privacy. Data privacy is critical, and even anonymised data can risk revealing individual identities. In my research, I use GDPR-compliant data obtained from mobile applications with explicit user consent to share their locations. To mitigate privacy risks, the study strictly avoids publishing any trajectories that could be traced back to specific individuals.

The following subsections briefly summarise my contributions to open science practices, presenting the SySMo model documentation and data repositories.

4.2.1 SySMo model documentation

The documentation [141] for the Synthetic Sweden Mobility (SySMo) Model provides a comprehensive overview of its methodology, including population synthesis, activity generation, and location and mode assignment components. Additionally, it features detailed chapters on model evaluation and assessment to ensure the model's accuracy and reliability. To enhance accessibility, the documentation is published online, enabling researchers and practitioners to easily access, understand, and utilise the SySMo model for their transportation studies.

URL: `https://research.chalmers.se/en/publication/531094`

4.2.2 SySMo model data repositories

A Synthetic Population of Sweden: Datasets of Agents, Households, and Activity-Travel Patterns

This repository contains a synthetic replica of over 10 million Swedish individuals (agents), along with their household characteristics and activity-travel plans. The datasets are organised in a relational database format across three primary tables: Person, Household, and Activity-travel.

- Person Table: Includes socio-demographic attributes of the synthetic agents, such as age, gender, civil status, residential zone, personal income, car ownership, employment status, and more.
- Household Table: Stores household attributes, including household type, size, number of children, and number of cars.
- Activity-travel Table: Contains daily activity schedules of agents, detailing where and when they engage in activities (e.g., work, home, school) and their modes of transportation between these activities (e.g., walking, biking, car, public transport).

URL: `http://doi.org/10.17632/9n29p7rmn5`

Integrated Agent-Based Modelling and Simulation of Transportation Demand and Mobility Patterns in Sweden

This dataset encompasses 10.2 million simulated SySMo agents representing Sweden's population and corresponding travel trajectories. Utilising the MATSim agent-based traffic simulation platform, the agents' activity-travel plans were simulated across a multi-modal network that integrates road and public transit data for Sweden.

The open data repository includes four distinct datasets:

- Synthetic Agents: Detailed profiles of the synthetic population.
- Activity Plans of the Agents: Daily schedules outlining the activities and locations of each agent.
- Travel Trajectories of the Agents: Simulated paths agents take as they move between activities.
- Road Network: Geospatial data of Sweden's road infrastructure.

URL: `https://doi.org/10.5281/zenodo.10648078`

4.2.3 The generative model data repositories

Synthetic multi-day activity-travel schedules for Swedish residents

This dataset provides multi-day activity-travel schedules for over 263,000 individuals residing in Sweden, representing approximately 2.6% of the country's population. The individuals and their daily schedules are derived from mobile phone application data collected over seven months in 2019. Mobile phone application data, an example of emerging mobility data sources, offers an alternative to traditional data collection methods. This data is gathered by capturing users' geographical locations—with their consent—as they interact with various mobile applications.

The open data repository includes activity-travel schedules for each individual across five simulated average weekdays, capturing daily variability at the individual level. Each simulated day provides:

- Anonymised Identifiers: Unique IDs that link individuals across all simulation days.

- Activity Locations: Locations for home, work/school and other activities.
- Daily Activity-Travel Schedules: Detailed information on activity sequence, type, start and end times, and locations.

URL: `https://doi.org/10.5281/zenodo.14012139`

4.3 Applications of the models

The SySMo model has attracted attention (its results have been downloaded over 300 times) and was utilised by several research groups across universities and research institutes. Its applications include studies on battery electric vehicles, charging infrastructure, and energy systems modelling. Within our research group, the SySMo model has been applied in several studies. For example, Liao et al. [97] assess battery electric vehicles and their charging infrastructure, while Paper IV, included in this thesis, explores the potential of e-bikes to replace car trips and reduce emissions. As the model presented in Paper III is very recent, it has not been applied in any research.

This section provides an overview of the SySMo model's primary applications, which are not covered in the thesis.

4.3.1 Battery electric vehicle charging infrastructure

Journal article: Impacts of charging behaviour on BEV charging infrastructure needs and energy use

The SySMo model's one application evaluates battery electric vehicle (BEV) charging infrastructure needs and energy consumption, conducted by Liao et al. [97]. BEVs are crucial for the sustainable transformation of future transport systems. While increasing public transport use, cycling, and walking are essential, private vehicles likely remain significant to personal mobility. Achieving 100% BEV adoption necessitates careful infrastructure planning to accelerate uptake and meet user needs. Full adoption needs infrastructure to provide charging access for individuals across all dwelling types, including those without private parking.

The study utilises the SySMo model's advanced large-scale synthetic population, which preserves realistic socio-demographic attributes and heterogeneous activity plans. To provide valuable insights into charging infrastructure requirements, the study's methodology integrates detailed charging behaviours into the simulation. The findings reveal that the assumed charging strategies significantly influence spatiotemporal patterns of charging demand and the required number of charging points at homes, workplaces, and public spaces.

This research highlights the importance of employing detailed individual mobility patterns, incorporating socio-demographic characteristics and diverse activity-travel behaviours to inform infrastructure planning and policy-making. This approach enables decision-makers to more accurately assess infrastructure needs and support the transition to sustainable transportation systems.

Research project: Electricity for Even More (El för ännu fler)

The research project conducted by researchers from RISE research institutes employs the SySMo model outputs [143]. This study aims to develop a nationwide decision support tool for establishing public charging solutions in urban areas where private charging access is limited. Electric vehicles are essential for achieving climate goals and facilitating the transition to sustainable transportation. However, the rapid electrification of the transport sector faces challenges, particularly in providing charging access to approximately half of Swedish households residing in apartment buildings without private parking.

The project addresses this challenge by creating a tool that assists both public and private sectors in determining charging locations for expanding charging infrastructure. The tool leverages high-resolution individual mobility data from the SySMo model and incorporates detailed analyses of car owners in apartments, business models, and environmental impacts. This approach focuses on efficient and equitable electrification by providing accessible public charging options for households without private chargers.

The project aims to support the swift and economically viable expansion of public charging infrastructure, promoting a fossil-free society. The decision support tool will provide actionable insights for policymakers and stakeholders, facilitating informed decisions in infrastructure planning.

4.3.2 Energy systems modelling

Journal article: A comparison between urban and non-urban municipalities applying a participatory approach

Oliveira Laurin et al. [144] conduct a cost-optimisation comparison between urban and non-urban municipalities by testing different socio-technical and context-specific scenarios developed through a participatory approach. While local authorities play a pivotal role in shaping energy policies and implementing action plans, their involvement has been underexplored. Utilising a participatory energy systems modelling approach, the study examines the cost-optimal decarbonisation of road transport in four Swedish municipalities, encompassing both urban and non-urban areas.

In collaboration with local authorities, the research develops socio-technical scenarios incorporating climate actions, resource availability, infrastructure, and travel patterns. By leveraging data from the SySMo model, the study provides insights into travel behaviours and vehicle occupancy across different municipalities, enhancing the accuracy of infrastructure planning and policy-making. The findings highlight the importance of tailoring decarbonisation strategies to local contexts to maximise effectiveness, balancing national and regional climate goals with specific urban and non-urban challenges.

4.4 Outlook

This section outlines potential future directions for my research involving transport model developments (detailed in Papers I–III) to evaluate sustainable transportation modes. There are promising opportunities to further advance these models and contribute to the modelling field by incorporating big data sources and integrating energy system models into these frameworks.

Two key directions for future research are highlighted:

- **Exploring big data sources**

  This thesis presents research utilising big data sources to extract individuals' mobility behaviours. Big data enables large-scale observations over extended periods, which can be challenging to achieve with traditional data collection methods. Paper III explores the use of mobile phone application data collected from individuals' mobile devices for transportation studies. The mobile phone data contains abundant GPS records from individuals, offering detailed insights into individuals' movements and whereabouts.

This study proposes a generative model that synthesises individuals' daily activity-travel patterns from mobile data using individuals' stop points, where they spend a certain amount of time performing activities. However, the data also includes the individuals' movements to access the activities. The movement data is invaluable for modelling individuals' mode choices [133] and route choices [145]. By employing advanced techniques and fully leveraging the richness of the data, my future research aims to derive individual mobility patterns, providing a more comprehensive understanding of mobility patterns.

Moreover, big data sources often have incomplete observations about individuals' daily activities because they capture data when individuals interact with their phones. Paper III address this issue by incorporating a travel survey to complement the mobile phone data. Moving forward, I am interested in exploring methods to use mobile phone data as a standalone resource, removing the need for supplementary data and enhancing the transferability of models. This could involve developing algorithms to infer missing information or employing machine learning techniques to learn from complete activity-travel patterns from mobile phone data.

- **Integrating energy system models**

  This thesis primarily focuses on transportation systems and evaluating sustainable mobility; however, the transportation and energy sectors are undergoing simultaneous and interconnected transformations. Both systems are transitioning from fossil fuel dependence to renewable energy sources to achieve climate targets. In 2021, renewables accounted for 85% of new power capacity installed worldwide, predominantly driven by solar photovoltaics and wind power [146]. At the same time, the electrification of road transport is accelerating rapidly, with electric vehicles (EVs) increasing their market share in developed countries.

  The rapid growth in renewable energy generation and EV adoption presents unique challenges and opportunities. The increasing share of inflexible and variable power generation from renewables and the growing electricity demand from EV charging raises concerns about balancing energy supply and demand in the energy system with high renewable energy. Key questions include determining the necessary storage capacity, identifying flexibility mechanisms to maintain grid stability, and assessing the overall cost implications for energy systems. With their battery capacities and the potential for controlled charging and discharging, EVs offer a promising solution to these challenges. Techniques such as smart charging and vehicle-to-grid (V2G) services enable EVs to adjust their charging rates in response to grid conditions and supply energy back to the grid during energy production deficits.

To effectively assess this potential and guide policy decisions, it is essential to develop models that integrate power systems, transportation systems, and vehicle technologies [25]. Traditional transportation models, including the SySMo model, are limited because they often do not incorporate the dynamics of energy systems, particularly the variable nature of renewable energy supply. An integrated modelling approach can provide accurate insights into electrified transportation's temporal and spatial energy demand and supply.

Future research could focus on integrating power systems into the SySMo model. This integration enhances the model's capability to realistically simulate EV charging behaviours and their interactions with the energy grid. By incorporating energy system dynamics, the integrated model will enable a more comprehensive analysis of how transportation electrification interacts with and impacts the broader energy infrastructure.

While future studies include integrating big data sources and cross-disciplinary modelling approaches, my research interests are not limited to these areas and extend beyond to other dimensions of transportation systems and their sustainable transition. One area of particular interest is the evaluation of strategies to promote sustainability in freight transport. While passenger transport is rapidly evolving with the global growth in electric vehicle sales and the quick expansion of charging infrastructure [3], freight transport still requires significant improvements for de-carbonising the system. Exploring pathways to accelerate the adoption of battery electric trucks (BETs), which can potentially reduce tailpipe emissions from heavy-duty vehicles, presents an exciting opportunity for future research.

4.5 Concluding remarks and my reflections

This section places my research within a broader context and includes my reflections on my doctoral studies. During my PhD, I developed models to simulate transportation systems and evaluate sustainable transportation technologies. Recent technological advancements, computational power, and data have significantly influenced modelling studies, enhancing their relevance and impact on real-world applications. Today, we are surrounded by models assisting us in various aspects of daily life. For example, these models determine the best routes and times for our daily commutes, recommend movies and products based on our individual preferences, and manage energy consumption by analysing our usage patterns and predicting energy production. Therefore, I think it is essential to reflect on the role of models, assess their reliability, and consider the future roles of both models and scientists in the evolving scientific landscape.

Models are an integral part of science, serving as abstract representations of complex real-world phenomena that enable scientists to conceptualise and analyse underlying mechanisms. While prediction is often seen as the primary function of models, their utility is more than forecasting outcomes. Epstein [147] argues that models offer numerous benefits beyond prediction; models are invaluable for providing explanations of underlying mechanisms, guiding data collection through theoretical frameworks, illuminating the core dynamics of complex systems, suggesting analogies between different phenomena, and raising new research questions that drive scientific progress. In my research on transportation systems, I have found models' multifaceted roles particularly valuable in deepening my understanding of the intricate interactions within the systems. Developing realistic models has not only enabled me to address key research questions but also often raised additional questions about human mobility.

Beyond their scientific utility, models play a pivotal role in informing societal decision-making, serving as essential tools in infrastructure development, policy formulation, and urban planning. While my research has not yet been directly incorporated into public decision-making processes, I am eager to contribute to these discussions, recognising that well-designed models can greatly benefit society.

However, it is essential to interpret model outputs with caution. As the philosopher Korzybski [148] noted, "the map is not the territory", highlighting the distinction between models and reality. Models are simplifications of complex systems and cannot capture every detail. Overreliance on models without critical evaluation can lead to misguided decisions Thompson [149]. Furthermore, Oreskes, Shrader-Frechette and Belitz [150] argues that the concept of "validation" in modelling is flawed because it suggests that a model can be entirely accurate, overlooking the simplifications and uncertainties involved. Instead of pursuing absolute validation, models should be assessed based on their transparency, reproducibility, and their ability to be calibrated and aligned with empirical data. Recognising the limitations of models is essential to improve their accuracy, reliability, and overall value in both scientific research and decision-making processes.

Building upon these reflections on the models' roles and limitations, it is important to consider how the researchers' roles will evolve with advancements in artificial intelligence (AI). When I began my PhD, AI tools in science were limited, but recent years have seen their widespread adoption, particularly large language models like OpenAI's ChatGPT. AI enables researchers to accelerate their work, allowing them to focus on formulating innovative questions and designing experiments. However, AI still lacks the creative thinking needed to generate novel hypotheses or theories [151], making human expertise essential for interpreting AI outputs and integrating them into broader scientific contexts. The rise of machine intelligence also introduces new challenges, including ethical concerns, potential biases embedded within AI algorithms, and the transparency of decision-making processes [152]. Experts estimate a 50% chance of achieving high-level machine intelligence by 2040–2050, with a one-in-three likelihood of this development being uncontrolled and bringing unintended consequences [153]. Therefore, it is crucial to prioritise societal benefits while mitigating potential risks when developing future models. Additionally, open public dialogue is essential to build trust and ensure transparency in AI usage. By embracing these responsibilities, researchers can advance scientific knowledge while positively contributing to society's well-being.

In concluding this section, I would like to reflect on my PhD journey, which has been a long and transformative experience marked by both challenges and personal growth. Early in my PhD, one of the challenges I faced was determining the appropriate level of abstraction in modelling, particularly given the sophistication of human behaviour, while also considering time constraints and technical capabilities. The temptation to model every detail was strong, but I learned that clearly defining the study's boundaries is essential for creating a model that addresses the research question within practical constraints. This understanding has influenced not only my academic work but also my approach to problem-solving in everyday life, teaching me to focus on what truly matters. I am grateful to have had supportive colleagues and mentors who provided guidance whenever I found myself delving too deeply into details. One key insight I gained is the importance of transparency and openness in research. This perspective encouraged collaboration with researchers both within and outside my research group, enriching my understanding. The collaborative work culture in our division profoundly shaped my research approach, teaching me how to ask meaningful questions. As I move forward in my career, I am committed to applying these lessons to ensure my work contributes to both scientific progress and societal well-being.

References

[1] P. Jaramillo, S. K. Ribeiro, P. Newman, S. Dhar, O. Diemuodeke, T. Kajino, D. Lee, S. Nugroho, X. Ou, A. H. Strømman and J. Whitehead (2022). *Climate change 2022: mitigation of climate change. Contribution of working group III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, Chapter 10*. Tech. rep. Cambridge University Press, Cambridge, UK and New York, NY, USA. URL: `https://www.ipcc.ch/report/ar6/wg3/downloads/report/IPCC_AR6_WGIII_Chapter10.pdf`.

[2] SMHI (n.d.). *National emissions database*. URL: `https://nationellaemissionsdatabasen.smhi.se/`.

[3] I. E. Agency (2024). *Global ev outlook 2024*. Tech. rep.

[4] L. M. Fulton (2018). Three revolutions in urban passenger travel. *Joule* **2** (4), pp. 575–578.

[5] M. Matyas and M. Kamargianni (2019). The potential of mobility as a service bundles as a mobility management tool. *Transportation* **46** (5), pp. 1951–1968.

[6] J. de Dios Ortúzar and L. G. Willumsen (2011). *Modelling transport*. John wiley & sons.

[7] M. G. McNally and C. R. Rindt (2007). 'The activity-based approach'. In: *Handbook of transport modelling*. Emerald Group Publishing Limited.

[8] R. Moeckel, W.-C. Huang, J. Ji, C. Llorca, A. T. Moreno, C. Staves, Q. Zhang and G. D. Erhardt (2024). The activity-based model abit: modeling 24 hours, 7 days a week. *Transportation Research Procedia* **78**, pp. 499–506.

[9] T. Hilgert, M. Heilig, M. Kagerbauer and P. Vortisch (2017). Modeling week activity schedules for travel demand models. *Transportation Research Record* **2666** (1), pp. 69–77.

[10] P. M. Boesch, F. Ciari and K. W. Axhausen (2016). Autonomous vehicle fleet sizes required to serve different levels of demand. *Transportation Research Record* **2542** (1), pp. 111–119.

[11] R. Basu, A. Araldo, A. P. Akkinepally, B. H. Nahmias Biran, K. Basak, R. Seshadri, N. Deshmukh, N. Kumar, C. L. Azevedo and M. Ben-Akiva (2018). Automated mobility-on-demand vs. mass transit: a multi-modal activity-driven agent-based simulation approach. *Transportation Research Record* **2672** (8), pp. 608–618.

[12] K. Mueller and K. W. Axhausen (2011). Hierarchical ipf: generating a synthetic population for switzerland. *Arbeitsberichte Verkehrs-und Raumplanung* **718**.

[13] J. Rich (2018). Large-scale spatial population synthesis for denmark. *European Transport Research Review* **10** (2), p. 63.

[14] S. Hörl and M. Balac (2021). Synthetic population and travel demand for paris and île-de-france based on open and publicly available data. *Transportation Research Part C: Emerging Technologies* **130**, p. 103291.

[15] D. A. Hensher, K. J. Button, K. E. Haynes and P. R. Stopher (2004). *Handbook of transport geography and spatial systems.* Vol. 5. Emerald Group Publishing Limited, pp. 13–26.

[16] J.-P. Rodrigue and T. Notteboom (2020). Ports and economic development. *T. Nottleboom, P. Athanasios, & J.-P. Rodrigue, Port Economics, Management and Policy. New York: Routledge.*

[17] J.-P. Rodrigue (2024a). 8.1–transportation and the urban form. *The Geography of Transport Systems; Routledge: New York, NY, USA.*

[18] E. Blumenberg (2017). Social equity and urban transportation. *The geography of urban transportation* **332**.

[19] E. Weiner (1997). *Urban transportation planning in the united states: an historical overview.* US Department of Transportation.

[20] S. Wise, A. Crooks and M. Batty (2016). 'Transportation in agent-based urban modelling'. In: *International workshop on agent based modelling of urban systems*. Springer, pp. 129–148.

[21] R. van Nes and G. de Jong (2020). 'Transport models'. In: *Advances in transport policy and planning*. Vol. 6. Elsevier, pp. 101–128.

[22] Y. Xiong, J. Gan, B. An, C. Miao and Y. C. Soh (2016). 'Optimal pricing for efficient electric vehicle charging station management'. In: *Proceedings of the 2016 international conference on autonomous agents & multiagent systems*. International Foundation for Autonomous Agents and Multiagent Systems, pp. 749–757.

[23] D. Dias, O. Tchepel and A. P. Antunes (2016). Integrated modelling approach for the evaluation of low emission zones. *Journal of environmental management* **177**, pp. 253–263.

[24] F. J. Márquez-Fernández, J. Bischoff, G. Domingues-Olavarría and M. Alaküla (2021). Assessment of future ev charging infrastructure scenarios for long-distance transport in Sweden. *IEEE Transactions on Transportation Electrification*.

[25] M. D. Galus, R. A. Waraich, F. Noembrini, K. Steurs, G. Georges, K. Boulouchos, K. W. Axhausen and G. Andersson (2012). Integrating power systems, transport systems and vehicle technology for electric mobility impact assessment and efficient control. *IEEE Transactions on Smart Grid* **3** (2), pp. 934–949.

[26] *The Australian Transport Assessment and Planning (ATAP) Guidelines* (2019). Tech. rep. Transport, Australian. URL: `https://www.atap.gov.au/tools-techniques/travel-demand-modelling/2-overview`.

[27] E. Cascetta (2013). *Transportation systems engineering: theory and methods*. Vol. 49. Springer Science & Business Media.

[28] T. Litman (2017). *Understanding transport demands and elasticities*. Victoria Transport Policy Institute Victoria, BC, Canada.

[29] J.-P. Rodrigue (2024b). *Geography of transport systems*. Routledge.

[30] P. L. Mokhtarian and I. Salomon (2001). How derived is the demand for travel? some conceptual and measurement con-

siderations. *Transportation research part A: Policy and practice* **35** (8), pp. 695–719.

[31] H. Barbosa, M. Barthelemy, G. Ghoshal, C. R. James, M. Lenormand, T. Louail, R. Menezes, J. J. Ramasco, F. Simini and M. Tomasini (2018). Human mobility: models and applications. *Physics Reports* **734**, pp. 1–74.

[32] T. Adler and M. Ben-Akiva (1979). A theoretical and empirical model of trip chaining behavior. *Transportation Research Part B: Methodological* **13** (3), pp. 243–257.

[33] S. ALGERS, J. ELIASSON and J. KOEHLER (2005). Microsimulating the stockholm integrated model system (sims). *PROCEEDINGS OF ETC 2005, STRASBOURG, FRANCE 18-20 SEPTEMBER 2005-RESEARCH TO INFORM DECISION-MAKING IN TRANSPORT APPLIED METHODS/INNOVATIVE METHODS-ACTIVITY BASED MODELS.*

[34] Y. Shiftan (1998). Practical approach to model trip chaining. *Transportation research record* **1645** (1), pp. 17–23.

[35] T. F. Rossi and Y. Shiftan (1997). Tour based travel demand modeling in the us. *IFAC Proceedings Volumes* **30** (8), pp. 381–386.

[36] H. W. David Boyce (2015). *Forecasting urban travel: past, present and future.* Edward Elgar Press.

[37] D. Damm and S. R. Lerman (1981). A theory of activity scheduling behavior. *Environment and Planning A* **13** (6), pp. 703–718.

[38] R. Kitamura (1984a). A model of daily time allocation to discretionary out-of-home activities and trips. *Transportation Research Part B: Methodological* **18** (3), pp. 255–266.

[39] P. M. Jones (2021). 'New approaches to understanding travel behaviour: the human activity approach'. In: *Behavioural travel modelling.* Routledge, pp. 55–80.

[40] T. Hägerstrand (1970). 'What about people in regional science?' In: *Papers of the regional science association.* Vol. 24.

[41] E. I. Pas (1984). The effect of selected sociodemographic characteristics on daily travel-activity behavior. *Environment and Planning A* **16** (5), pp. 571–581.

[42] S. Hanson (1982). The determinants of daily travel-activity patterns: relative location and sociodemographic factors. *Urban Geography* **3** (3), pp. 179–202.

[43] R. Kitamura (1984b). Incorporating trip chaining into analysis of destination choice. *Transportation Research Part B: Methodological* **18** (1), pp. 67–81.

[44] T. F. Golob and M. G. McNally (1997). A model of activity participation and travel interactions between household heads. *Transportation Research Part B: Methodological* **31** (3), pp. 177–194.

[45] C. G. Pooley, D. Horton, G. Scheldeman, M. Tight, T. Jones, A. Chisholm, H. Harwatt and A. Jopson (2011). Household decision-making for everyday travel: a case study of walking and cycling in lancaster (uk). *Journal of Transport Geography* **19** (6), pp. 1601–1607.

[46] R. N. Buliung and P. S. Kanaroglou (2007). Activity–travel behaviour research: conceptual issues, state of the art, and emerging perspectives on behavioural analysis and simulation modelling. *Transport Reviews* **27** (2), pp. 151–187.

[47] R. Kitamura (1996). 'Applications of models of activity behavior for activity based demand forecasting'. In: *Activity-based travel forecasting conference, new orleans, louisiana.*

[48] A. R. Pinjari and C. R. Bhat (2011). 'Activity-based travel demand analysis'. In: *A handbook of transport economics.* Edward Elgar Publishing, pp. 213–248.

[49] D. Gopalakrishna, E. Schreffler, D. Vary, D. Friedenfeld, B. Kuhn, C. Dusza, R. Klein and A. Rosas (2012). *Integrating demand management into the transportation planning process: a desk reference.* Tech. rep., pp. 146–154.

[50] D. Chen (2009). 'A grid aware large scale agent-based simulation system'. In: *Quantitative quality of service for grid computing: applications for heterogeneity, large-scale distribution, and dynamic environments.* IGI Global, pp. 299–319.

[51] E. Guerci, M. A. Rastegar and S. Cincotti (2010). 'Agent-based modeling and simulation of competitive wholesale electricity

markets'. In: *Handbook of power systems ii*. Springer, pp. 241–286.

[52] D. Sornette (2004). A complex system view of why stock markets crash. *New Thesis* **1** (1), pp. 5–18.

[53] J. M. Epstein and R. Axtell (1996). *Growing artificial societies: social science from the bottom up*. Brookings Institution Press.

[54] J. H. Miller, S. E. Page and B. LeBaron (2008). Complex adaptive systems: an introduction to computational models of social life. *Journal of Economic Literature* **46** (2), pp. 427–428.

[55] J. Castiglione, M. Bradley and J. Gliebe (2015). *Activity-based travel demand models: a primer*. SHRP 2 Report S2-C46-RR-1. Transportation Research Board.

[56] D. for Transport (DfT) (2024). *Transport analysis guidance (tag) unit m5.4 - agent-based methods and activity-based demand modelling*. Tech. rep. Department for Transport (DfT).

[57] K. Müller and K. W. Axhausen (2010). Population synthesis for microsimulation: state of the art. *Arbeitsberichte Verkehrs-und Raumplanung* **638**.

[58] D. Ballas, G. Clarke, D. Dorling and D. Rossiter (2007). Using simbritain to model the geographical impact of national government policies. *Geographical Analysis* **39** (1), pp. 44–77.

[59] R. Tanton et al. (2014). A review of spatial microsimulation methods. *International Journal of Microsimulation* **7** (1), pp. 4–25.

[60] S. Feygin and A. Pozdnoukhov (2018). Peer pressure enables actuation of mobility lifestyles. *Transportation Research Part C: Emerging Technologies* **87**, pp. 26–45.

[61] I. Saadi, A. Mustafa, J. Teller, B. Farooq and M. Cools (2016). Hidden markov model-based population synthesis. *Transportation Research Part B: Methodological* **90**, pp. 1–21.

[62] S. S. Borysov, J. Rich and F. C. Pereira (2019). How to generate micro-agents? a deep generative modeling approach to population synthesis. *Transportation Research Part C: Emerging Technologies* **106**, pp. 73–97.

[63] L. Smith, R. Beckman and K. Baggerly (1995). *Transims: transportation analysis and simulation system.* Tech. rep. Los Alamos National Lab.(LANL), Los Alamos, NM (United States).

[64] M. Frick (2004). Generating synthetic populations using ipf and monte carlo techniques: some new results. *Arbeitsberichte Verkehrs-und Raumplanung* **225**.

[65] T. Arentze, H. Timmermans and F. Hofman (2007). Creating synthetic household populations: problems and approach. *Transportation Research Record* **2014** (1), pp. 85–91.

[66] J. Y. Guo and C. R. Bhat (2007). Population synthesis for microsimulating travel behavior. *Transportation Research Record* **2014** (1), pp. 92–101.

[67] D. Charypar (2008). 'Efficient algorithms for the microsimulation of travel behavior in very large scenarios'. PhD thesis. ETH Zurich.

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

[69] E. J. Miller and M. J. Roorda (2003). Prototype model of household activity-travel scheduling. *Transportation Research Record* **1831** (1), pp. 114–121.

[70] J. Auld and A. K. Mohammadian (2012). Activity planning processes in the agent-based dynamic activity planning and travel scheduling (adapts) model. *Transportation Research Part A: Policy and Practice* **46** (8), pp. 1386–1403.

[71] B. Lenntorp (1977). Paths in space-time environments: a time-geographic study of movement possibilities of individuals. *Environment and Planning A* **9** (8), pp. 961–972.

[72] P. M. Jones, M. C. Dix, M. I. Clarke and I. G. Heggie (1983). *Understanding travel behaviour.* Gower Publishing.

[73] P. P. Huigen (1986). *Binnen of buiten bereik?: een sociaal-geografisch onderzoek in zuidwest-friesland.* Vol. 7. Koninklijk nederlands aardrijkskundig genootschap.

[74] M. Dijst and V. Vidakovic (1997). Individual action space in the city. *Activity-based approaches to travel analysis.*

[75] M.-P. Kwan (1997). Gisicas: an activity-based travel decision support system using a gis-interfaced computational-process model. *Activity-based approaches to travel analysis.*

[76] N. A. Khan (2020). Modelling and microsimulation of activity generation, activity scheduling and mobility assignment.

[77] T. A. Arentze and H. J. Timmermans (2004). A learning-based transportation oriented simulation system. *Transportation Research Part B: Methodological* **38** (7), pp. 613–633.

[78] J. A. Downs and M. W. Horner (2012). Probabilistic potential path trees for visualizing and analyzing vehicle tracking data. *Journal of Transport Geography* **23**, pp. 72–80.

[79] D. Ettema and H. Timmermans (2007). Space–time accessibility under conditions of uncertain travel times: theory and numerical simulations. *Geographical Analysis* **39** (2), pp. 217–240.

[80] B. Y. Chen, Q. Li, D. Wang, S.-L. Shaw, W. H. Lam, H. Yuan and Z. Fang (2013). Reliable space–time prisms under travel time uncertainty. *Annals of the Association of American Geographers* **103** (6), pp. 1502–1521.

[81] M. H. Hafezi, H. Millward and L. Liu (2018). 'Activity-based travel demand modeling: progress and possibilities'. In: *International conference on transportation and development 2018: planning, sustainability, and infrastructure systems.* American Society of Civil Engineers Reston, VA, pp. 138–147.

[82] M. E. Ben-Akiva and J. L. Bowman (1998). 'Activity based travel demand model systems'. In: *Equilibrium and advanced transportation modelling.* Springer, pp. 27–46.

[83] J. L. Bowman and M. E. Ben-Akiva (2001a). Activity-based disaggregate travel demand model system with activity schedules. *Transportation Research Part A: Policy and Practice* **35** (1), pp. 1–28.

[84] C. R. Bhat, J. Y. Guo, S. Srinivasan and A. Sivakumar (2004). Comprehensive econometric microsimulator for daily activity-travel patterns. *Transportation Research Record* **1894** (1), pp. 57–66.

[85] R. Kitamura, S. Fujii et al. (1998). Two computational process models of activity-travel behavior. *Theoretical foundations of travel choice modeling*, pp. 251–279.

[86] X. Di and H. X. Liu (2016). Boundedly rational route choice behavior: a review of models and methodologies. *Transportation Research Part B: Methodological* **85**, pp. 142–179.

[87] T. Gärling, M.-p. Kwan and R. G. Golledge (1994). Computational-process modelling of household activity scheduling. *Transportation Research Part B: Methodological* **28** (5), pp. 355–364.

[88] J. Auld and A. Mohammadian (2009). Framework for the development of the agent-based dynamic activity planning and travel scheduling (adapts) model. *Transportation Letters* **1** (3), pp. 245–255.

[89] A. N. Koushik, M. Manoj and N. Nezamuddin (2020). Machine learning applications in activity-travel behaviour research: a review. *Transport reviews* **40** (3), pp. 288–311.

[90] M. H. Hafezi, L. Liu and H. Millward (2019). A time-use activity-pattern recognition model for activity-based travel demand modeling. *Transportation* **46** (4), pp. 1369–1394.

[91] T. Arentze and H. Timmermans (2000). *Albatross: a learning based transportation oriented simulation system.* EIRASS.

[92] M. Čertickỳ, J. Drchal, M. Cuchỳ and M. Jakob (2015). 'Fully agent-based simulation model of multimodal mobility in european cities'. In: *2015 international conference on models and technologies for intelligent transportation systems (mt-its).* IEEE, pp. 229–236.

[93] M. H. Hafezi, L. Liu and H. Millward (2018). Learning daily activity sequences of population groups using random forest theory. *Transportation research record* **2672** (47), pp. 194–207.

[94] M. Allahviranloo and W. Recker (2013). Daily activity pattern recognition by using support vector machines with multiple classes. *Transportation Research Part B: Methodological* **58**, pp. 16–43.

[95] J. Drchal, M. Čertický and M. Jakob (2019). Data-driven activity scheduler for agent-based mobility models. *Transportation Research Part C: Emerging Technologies* **98**, pp. 370–390.

[96] A. Horni, K. Nagel and K. Axhausen, eds. (2016). *Multi-agent transport simulation matsim.* London: Ubiquity Press, p. 618. ISBN: 978-1-909188-75-4, 978-1-909188-76-1, 978-1-909188-77-8, 978-1-909188-78-5. DOI: `10.5334/baw`.

[97] Y. Liao, Ç. Tozluoğlu, F. Sprei, S. Yeh and S. Dhamal (2023). Impacts of charging behavior on BEV charging infrastructure needs and energy use. *Transportation Research Part D: Transport and Environment* **116**, p. 103645.

[98] J. Hamadneh and D. Esztergár-Kiss (2021). The influence of introducing autonomous vehicles on conventional transport modes and travel time. *Energies* **14** (14), p. 4163.

[99] L. M. Martínez, G. H. d. A. Correia, F. Moura and M. Mendes Lopes (2017). Insights into carsharing demand dynamics: outputs of an agent-based model application to lisbon, portugal. *International Journal of Sustainable Transportation* **11** (2), pp. 148–159.

[100] K. Nagel and F. Marchal (2003). Computational methods for multi-agent simulations of travel behavior. *Proceedings of International Association for Travel Behavior Research (IATBR), Lucerne, Switzerland.*

[101] K. Nagel, B. Kickhöfer, A. Horni and D. Charypar (2016). *A closer look at scoring.*

[102] *Statistics Sweden* (2020). `https://www.statistikdatabasen.scb.se/pxweb/en/ssd/`.

[103] M. Beser and S. Algers (2002). ‘SAMPERS—The new Swedish national travel demand forecasting tool’. In: *National transport models.* Springer, pp. 101–118.

[104] *Passenger and goods transport report* (2021). URL: `https://www.trafa.se/ovrig/transportarbete/` (Retrieved: 2021-11-10).

[105] Ç. Tozluoğlu, S. Dhamal, S. Yeh, F. Sprei, Y. Liao, M. Marathe, C. L. Barrett and D. Dubhashi (2023). A synthetic population

of Sweden: datasets of agents, households, and activity-travel patterns. *Data in Brief* **48**, p. 109209.

[106] D. M. Endres and J. E. Schindelin (2003). A new metric for probability distributions. *IEEE Transactions on Information theory* **49** (7), pp. 1858–1860.

[107] S. A. O. Medina (2018). Inferring weekly primary activity patterns using public transport smart card data and a household travel survey. *Travel Behaviour and Society* **12**, pp. 93–101.

[108] A. Hadachi, M. Pourmoradnasseri and K. Khoshkhah (2020). Unveiling large-scale commuting patterns based on mobile phone cellular network data. *Journal of Transport Geography* **89**, p. 102871.

[109] P. Wang, T. Hunter, A. M. Bayen, K. Schechtner and M. C. González (2012). Understanding road usage patterns in urban areas. *Scientific reports* **2** (1), p. 1001.

[110] M. S. Iqbal, C. F. Choudhury, P. Wang and M. C. González (2014). Development of origin-destination matrices using mobile phone call data. *Transportation Research Part C: Emerging Technologies* **40**, pp. 63–74.

[111] Z. Peng, G. Bai, H. Wu, L. Liu and Y. Yu (2021). Travel mode recognition of urban residents using mobile phone data and mapapi. *Environment and Planning B: Urban Analytics and City Science* **48** (9), pp. 2574–2589.

[112] M. Yang, Y. Pan, A. Darzi, S. Ghader, C. Xiong and L. Zhang (2022). A data-driven travel mode share estimation framework based on mobile device location data. *Transportation* **49** (5), pp. 1339–1383.

[113] H. Zhang, X. Song, Y. Long, T. Xia, K. Fang, J. Zheng, D. Huang, R. Shibasaki and Y. Liang (2019). Mobile phone gps data in urban bicycle-sharing: layout optimization and emissions reduction analysis. *Applied Energy* **242**, pp. 138–147.

[114] Q. Yu, H. Zhang, W. Li, Y. Sui, X. Song, D. Yang, R. Shibasaki and W. Jiang (2020). Mobile phone data in urban bicycle-sharing: market-oriented sub-area division and spatial analysis on emission reduction potentials. *Journal of Cleaner Production* **254**, p. 119974.

[115] L. Alexander, S. Jiang, M. Murga and M. C. González (2015). Origin–destination trips by purpose and time of day inferred from mobile phone data. *Transportation research part c: emerging technologies* **58**, pp. 240–250.

[116] S. Jiang, J. Ferreira and M. C. Gonzalez (2017). Activity-based human mobility patterns inferred from mobile phone data: a case study of singapore. *IEEE Transactions on Big Data* **3** (2), pp. 208–219.

[117] N. Grujić, S. Brdar, S. Osinga, G. J. Hofstede, I. N. Athanasiadis, M. Pljakić, N. Obrenović, M. Govedarica and V. Crnojević (2022). Combining telecom data with heterogeneous data sources for traffic and emission assessments—an agent-based approach. *ISPRS International Journal of Geo-Information* **11** (7), p. 366.

[118] *The Swedish national travel survey* (2021). `https://www.trafa.se/en/travel-survey/travel-survey/`.

[119] *Demographic Statistical Areas (DeSO)* (2020). `https://www.scb.se/en/services/open-data-api/open-geodata/deso--demographic-statistical-areas/`.

[120] J. L. Bowman and M. E. Ben-Akiva (2001b). Activity-based disaggregate travel demand model system with activity schedules. *Transportation Research Part A: Policy and Practice* **35** (1), pp. 1–28. DOI: `10.1016/S0965-8564(99)00043-9`.

[121] C. Chen, L. Bian and J. Ma (2014). From traces to trajectories: how well can we guess activity locations from mobile phone traces? *Transportation Research Part C: Emerging Technologies* **46**, pp. 326–337.

[122] B. Sadeghinasr, A. Akhavan and Q. Wang (2019). 'Estimating commuting patterns from high resolution phone gps data'. In: *Asce international conference on computing in civil engineering 2019*. American Society of Civil Engineers Reston, VA, pp. 9–16.

[123] I. Philips, J. Anable and T. Chatterton (2022). E-bikes and their capability to reduce car co2 emissions. *Transport Policy* **116**, pp. 11–23.

[124] D. Bucher, R. Buffat, A. Froemelt and M. Raubal (2019). Energy and greenhouse gas emission reduction potentials resulting from different commuter electric bicycle adoption scenarios

in switzerland. *Renewable and sustainable energy reviews* **114**, p. 109298.

[125] K. W Axhausen, A. Horni and K. Nagel (2016). *The multi-agent transport simulation MATSim*. Ubiquity Press.

[126] OTP (n.d.). *Open trip planner*. URL: https://www.opentripplanner.org/.

[127] S. Flügel, N. Hulleberg, A. Fyhri, C. Weber and G. Ævarsson (2019). Empirical speed models for cycling in the oslo road network. *Transportation* **46**, pp. 1395–1419.

[128] K. Schleinitz, T. Petzoldt, L. Franke-Bartholdt, J. Krems and T. Gehlert (2017). The german naturalistic cycling study–comparing cycling speed of riders of different e-bikes and conventional bicycles. *Safety science* **92**, pp. 290–297.

[129] *Trafikanalys. Körsträckor 2021.* (N.d.). URL: https://www.trafa.se/globalassets/statistik/vagtrafik/korstrackor/2021/korstrackor-2021.pdf.

[130] F. F. Bastarianto, T. O. Hancock, C. F. Choudhury and E. Manley (2023). Agent-based models in urban transportation: review, challenges, and opportunities. *European Transport Research Review* **15** (1), p. 19.

[131] O. Canella, G. Flötteröd, D. Johnsson, I. Kristoffersson, P. Larek and J. Thelin (2016). *Flexible coupling of disaggregate travel demand models and network simulation packages (ihop2)*. Tech. rep. Technical report, KTH, Sweco, WSP.

[132] M. Fosgerau, M. Łukawska, M. Paulsen and T. K. Rasmussen (2023). Bikeability and the induced demand for cycling. *Proceedings of the National Academy of Sciences* **120** (16), e2220515120.

[133] A. Andersson, L. Engelson, M. Börjesson, A. Daly and I. Kristoffersson (2022). Long-distance mode choice model estimation using mobile phone network data. *Journal of choice modelling* **42**, p. 100337.

[134] A. Andersson (2024). 'Modelling long-distance travel demand by combining mobile phone and survey data'. PhD thesis. Linköping University Electronic Press.

[135] J. Li, C. O'Donoghue et al. (2013). A survey of dynamic microsimulation models: uses, model structure and methodology. *International Journal of microsimulation* **6** (2), pp. 3–55.

[136] C. Livingston, S. Penazzi, G. Scozzaro and C. Buire (2022). *Transit modelling and simulation framework*. Tech. rep. TRANSIT Consortium. URL: `https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e5f0b4eaff&appId=PPGMS`.

[137] M. Luca, G. Barlacchi, B. Lepri and L. Pappalardo (2021). A survey on deep learning for human mobility. *ACM Computing Surveys (CSUR)* **55** (1), pp. 1–44.

[138] C. J. Haboucha, R. Ishaq and Y. Shiftan (2017). User preferences regarding autonomous vehicles. *Transportation research part C: emerging technologies* **78**, pp. 37–49.

[139] S. N. Goodman, D. Fanelli and J. P. Ioannidis (2016). What does research reproducibility mean? *Science translational medicine* **8** (341), 341ps12–341ps12.

[140] P. V. Coveney, D. Groen and A. G. Hoekstra (2021). *Reliability and reproducibility in computational science: implementing validation, verification and uncertainty quantification in silico*.

[141] Ç. Tozluoğlu, S. Dhamal, Y. Liao, S. Yeh, F. Sprei, M. Marathe, C. Barrett and D. Dubhashi (2022). Synthetic Sweden Mobility (SySMo) model documentation. URL: `https://research.chalmers.se/publication/531094/file/531094_Fulltext.pdf`.

[142] Ç. Tozluoğlu (2024). *Mobile application data for activity generation (MAD4AG)*. URL: `https://github.com/tozlucaglar/MAD4AG`.

[143] J. Hagman, H. Alfredsson and J. Rogstadius (n.d.). *El för ännu fler*. URL: `https://www.ri.se/sv/vad-vi-gor/projekt/el-for-annu-fler`.

[144] M. de Oliveira Laurin, V. Aryanpur, H. Farabi-Asl, M. Grahn, M. Taljegard and K. Vilén (2024). 'A comparison between urban and non-urban municipalities applying a participatory approach'.

[145] J. Burrieza-Galán, I. Martín, F. Nicolai, J. Monroig, J. Torres and M. Picornell (2023). Identification of long-distance tour purpose through ml-based data fusion techniques for combining mobile network and survey data. *Transportation Research Procedia* **71**, pp. 93–100.

[146] D. Bogdanov and C. Breyer (2024). Role of smart charging of electric vehicles and vehicle-to-grid in integrated renewables-based energy systems on country level. *Energy* **301**, p. 131635.

[147] J. M. Epstein (2008). Why model? *Journal of artificial societies and social simulation* **11** (4), p. 12.

[148] A. Korzybski (1958). *Science and sanity: an introduction to non-aristotelian systems and general semantics.* Institute of GS.

[149] E. Thompson (2022). *Escape from model land.* Hachette UK Distribution.

[150] N. Oreskes, K. Shrader-Frechette and K. Belitz (1994). Verification, validation, and confirmation of numerical models in the earth sciences. *Science* **263** (5147), pp. 641–646.

[151] G. Marcus and E. Davis (2019). *Rebooting ai: building artificial intelligence we can trust.* Vintage.

[152] E. Commission (2024). *Living guidelines on the responsible use of generative ai in research.* Tech. rep.

[153] V. C. Müller and N. Bostrom (2016). Future progress in artificial intelligence: a survey of expert opinion. *Fundamental issues of artificial intelligence*, pp. 555–572.