Exploring the Spatial and Temporal Patterns in Travel Demand
A Data-Driven Approach

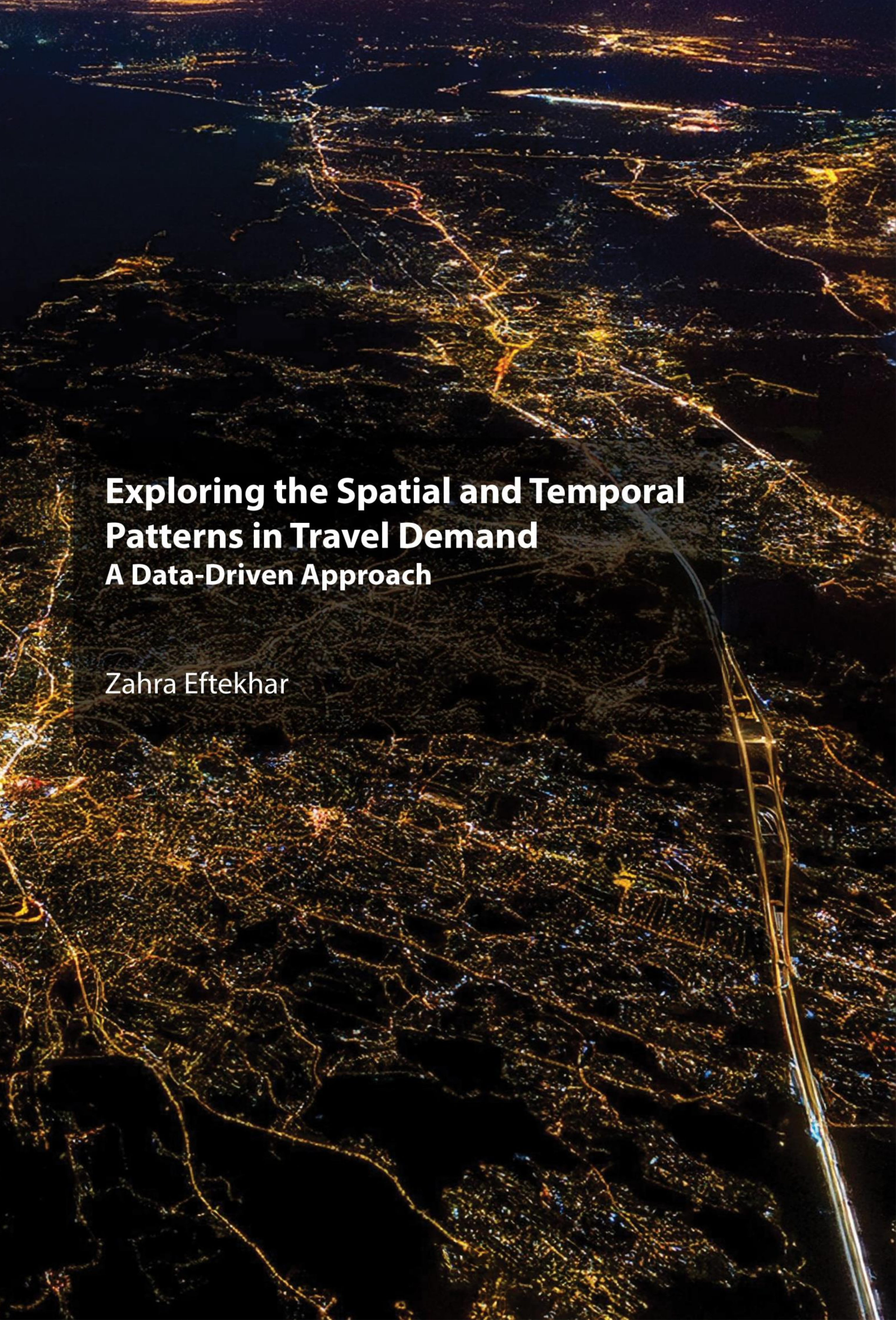
Eftekhar, Z.

DOI
10.4233/uuid:a6708496-3bee-49ae-8b67-9b268e06630c

Publication date
2025

Document Version
Final published version

Citation (APA)
Eftekhar, Z. (2025). *Exploring the Spatial and Temporal Patterns in Travel Demand: A Data-Driven Approach*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:a6708496-3bee-49ae-8b67-9b268e06630c



Exploring the Spatial and Temporal Patterns in Travel Demand

A Data-Driven Approach

Zahra Eftekhar

Exploring the Spatial and Temporal Patterns in Travel Demand: A Data-Driven Approach

Zahra EFTEKHAR

Exploring the Spatial and Temporal Patterns in Travel Demand: A Data-Driven Approach

Dissertation

for the purpose of obtaining the degree of doctor
at Delft University of Technology
by the authority of the Rector Magnificus, Prof. dr. ir. T.H.J.J. van der Hagen
chair of the Board for Doctorates
to be defended publicly on
Monday 2nd, June 2025 at 17:30 o'clock

by

Zahra EFTEKHAR
Master of Science in Civil Engineering
Sharif University of Technology, Iran
born in Lar, Iran

This dissertation has been approved by the promotors.

Composition of the doctoral committee:

Rector Magnificus	Chairperson
Prof.dr.ir. J.W.C. van Lint	Technische Universiteit Delft, promotor
Dr.ir. A.J. Pel	Technische Universiteit Delft, promotor

Independent members:

Prof.dr. C. Antoniou	Technical University of Munich, Germany
Prof.dr. F. Viti	University of Luxembourg, Luxembourg
Prof.dr.ir. L. Leclercq	Université Gustave Eiffel, France
Dr. S. Sharif Azadeh	Technische Universiteit Delft
Prof.dr.ir. S.P. Hoogendoorn	Technische Universiteit Delft

Reserve member:

Prof.dr.ir. L.A. Tavasszy	Technische Universiteit Delft

This doctoral dissertation was funded by the Dutch National Data Warehouse of Traffic Information (NDW). It was also part of the MiRRORS project (with project number 16720) within the Open Technology Program, which is (partly) financed by the Applied Sciences Division of the Dutch Research Council (NWO/TTW)

TRAIL Thesis Series no. T2025/09, the Netherlands Research School TRAIL

TRAIL
P.O. BOX 5017
2600 GA Delft
The Netherlands
E-mail: info@rsTRAIL.nl

ISBN: 978-90-5584-364-0

Printed in the Netherlands

To my loving parents, whose support turns the impossible into reality.

Acknowledgement

Embarking on the journey of pursuing a PhD has been one of the most significant and transformative experiences of my life. It was a path filled with aspirations, challenges, and invaluable learning, which I navigated with the support of many remarkable individuals. As I reflect on this journey, I am filled with gratitude for every person who played a role in making this dream a reality, especially in the vibrant and inspiring city of Delft.

First and foremost, I extend my deepest gratitude to my promotor, **Prof. Hans van Lint**. His unwavering support, guidance, and incredible confidence in me over the years have been nothing short of extraordinary. Prof. van Lint is not just a teacher but a true inspiration, embodying the essence of the saying, “good teachers teach, great teachers inspire.” One of my favorite pieces of advice from him—delivered with a signature grin—was: “Don’t always follow the rules; just apologize when you get caught.” That spirit of bold curiosity (and strategic charm) made every conversation with him both enlightening and unforgettable. I am immensely fortunate to have had him as my supervisor.

My heartfelt thanks also go to my co-promotor, **Dr. Adam Pel**, whose patience and insightful ideas significantly shaped my research and personal growth. One thing I truly admired about him was his ability to play with ideas—gently stretching, flipping, and testing them from all angles—until either a great insight emerged or it gracefully collapsed on its own, without anyone needing to declare it dead. Watching him work through ideas made even the dead ends feel like progress. Dr. Pel taught me the importance of being calm, analytical, and thoughtful before dismissing any notion—a lesson that will stay with me far beyond academia. His support during the challenging times of my PhD was invaluable, for which I am eternally grateful.

I would like to express my sincere gratitude to the members of my doctoral committee—Prof. Francesco Viti, Prof. Constantinos Antoniou, Prof. Ludovic Leclercq, Dr. Shadi Sharif Azadeh, Prof. Serge Hoogendoorn, and Prof. Lóri Tavasszy—for generously dedicating their time and expertise to review this thesis. Their thoughtful questions and remarks have been invaluable in refining my work. It is a great honor to have had such esteemed scholars as part of this journey, and I am truly thankful for their support.

I am also thankful for the wonderful memories created with my dear friends at Dittlab, including Panchamy & Jerry, Ding, Tin & Huong, Sanmay, Guopeng, Yiru, Kristel, Muriel, Ehab & Shahad, Saeed, Saman, Xue, Lukas, Kexin, and Ali. Your companionship made my time at TU Delft truly special, providing a warm and welcoming environment that felt like a second home.

Special thanks to the Iranian community of TU Delft and our close friends, Ali & Razieh, Peyman & Atefeh, Sarah, Mahsa, Saeed, Saman, Kayhan, Saghar, Behzad & Negar, Sepehr, Mostafa, Golnaz, Ava & Behrouz, Sadaf & Arash, Farinaz & Mousa, Shekoufeh & Milad, Mostafa & Zahra, Mahsa & Saeed, Sepideh & Farshid, Delaram & Nima. Finding each other brought a sense of belonging and support that enriched my experience in Delft, making every moment memorable and enjoyable.

I also want to express my gratitude to my Delft friends, including Hoda, Yosra & Ramin, Leyla, AmirMasoud, and AhmadReza who made the early days of my life in the Netherlands comfortable and welcoming. Your friendship helped me adjust and thrive in a new environment.

To my family—**Maman, Baba, Saleheh, Saeed**, and my younger sister **Fatemeh**—thank you for your constant love, emotional support, and unwavering faith in me throughout this journey. Ever since I was little, you believed in my academic path with such pride—already imagining me with a PhD and maybe even standing in front of a classroom as “Professor Zahra.” I hope this thesis brings a smile to your faces and a quiet sense of joy when you talk about it with others. Having Fatemeh, Saeed, Mozhdeh, Ali, and Laleh close by has been a source of comfort and warmth that I will always hold dear.

Lastly—and most importantly—I want to thank my husband, **Parviz**, the love of my life and my favourite partner-in-crime. We are each other’s best friend, resident stand-up comedian, and personal growth coach all rolled into one. Whenever my motivation meter hit empty, you refilled it with espresso runs, dad jokes, and an unwavering belief that I could conquer anything life (or academia) thew at me. Your energy turned stressful days into mini dance parties, and your calm turned pandemic lockdowns into gourmet kitchen adventures. I am unbelievably lucky to have you by my side.

This acknowledgment would be incomplete without expressing my gratitude to everyone who contributed to my journey. Each one of you has left an indelible mark on my heart, and for that, I am forever grateful. Thank you for being a part of my PhD adventure.

Zahra Eftekhar,
Utrecht, May 2025.

Contents

Chapter 1

Introduction

Travel demand modeling has been a key topic in transportation science and practice for many decades. Estimating and predicting travel demand is fundamentally challenging because many aspects of travel demand, especially detailed origin-destination flows, are rarely measured directly. Instead, travel demand models often rely on other types of data, such as counts of vehicles at specific locations, surveys of travel behavior, and information about economic and demographic factors that can influence travel demand.

Despite this challenge, reliable estimates and predictions of travel demand are essential for many transportation policies and measures to operate effectively. For example, Transportation Demand Management (TDM) strategies aim to reduce congestion and environmental impacts by encouraging the use of alternative modes of transportation, such as public transit, cycling, or walking (Ferguson, 1990). Accurate travel demand estimates help identify areas and times where interventions like increasing public transit frequency, adding new transit routes, or extending operation hours for public transit, as well as improving cycling infrastructure, will be most effective.

Similarly, infrastructure improvements designed to reduce congestion require precise knowledge of where and when travel demand exceeds capacity (Antipova & Wilmot, 2012). This ensures that investments in new roads or transit lines are made where they will have the greatest impact, thereby optimizing resource allocation and avoiding unnecessary costs.

Land use planning uses travel demand data to promote sustainable urban development (Cervero, 1996). By understanding how land use patterns influence travel behavior, planners can design mixed-use developments that reduce the need for long-distance travel and encourage the use of alternative transportation modes.

The deployment of vehicle technologies, such as electric or low-emission vehicles, depends on understanding travel patterns to strategically place charging stations and assess potential environmental benefits (van der Vooren & Alkemade, 2012).

Finally, Intelligent Transportation Systems (ITS), including traffic signal coordination and real-time traffic information systems, rely on accurate and timely travel demand data to manage traffic flow efficiently (Srinivasan et al., 2006). By predicting where congestion is likely to occur, ITS can dynamically adjust traffic signals or provide route recommendations to drivers, reducing delays and improving overall network performance.

In all these cases, having a clear understanding of people's movements is essential for these tools to be effective. As such, this thesis aims to contribute to a better understanding of the spatial and temporal patterns of travel demand. *Spatial and temporal patterns* of travel demand refer to the patterns, trends, and variations in travel demand over time and space. Many factors influence these patterns: the timing of travel, spatial aggregation level and features, and the characteristics of the individuals making the trip. Studying these spatial and temporal patterns of travel demand is essential in understanding the complex interactions between transportation systems, land use, and population demographics.

While traditional travel demand models often focus on long-term average patterns or rely on static socio-demographic characteristics derived from infrequent surveys, these methods may not adequately capture the dynamic nature of travel behavior influenced by real-time factors such as time of day, day of the week, special events, or unexpected disruptions. Traditional models might miss short-term variations and spatial heterogeneity in travel demand. Understanding detailed spatial and temporal patterns provides insights into how travel demand fluctuates throughout the day, week, or year, and across different geographic areas. This knowledge allows for more responsive and adaptive transportation planning, enabling strategies like real-time traffic management, dynamic transit scheduling, and targeted demand management interventions.

In addition to spatial and temporal dimensions, travel demand is influenced by other factors such as travel purpose, mode choice, and individual preferences. Recent advancements in travel demand modeling, such as activity-based models and agent-based simulations, have focused on capturing individual-level behaviors and inter-personal differences. These models aim to provide a more detailed and behaviorally realistic representation of travel demand by considering the sequence of activities and trips made by individuals, as well as their preferences and constraints.

However, while disaggregate models offer valuable insights into individual travel behavior, they often require extensive data and computational resources, which may not be readily available for all contexts. In contrast, analyzing aggregate patterns at the population level can provide practical insights for transportation planning and policy-making, especially when leveraging large-scale data sources such as GSM mobile phone data. By focusing on spatial and temporal patterns at an aggregate level, we can identify general trends and variations in travel demand that are essential for optimizing transportation systems.

Our research aims to contribute to the understanding of travel demand by exploring spatial and temporal patterns in aggregate travel data. By focusing on these dimensions, we can capture the collective behavior of travelers, which is critical for infrastructure planning, service scheduling, and policy interventions. Additionally, spatial and temporal analyses are well-suited for leveraging big data sources, enabling us to uncover patterns that may not be evident at the individual level.

For instance, by analyzing temporal patterns, planners can identify peak travel times and implement measures such as staggered work hours or increased public transit frequency during those periods. Spatial pattern analysis helps in recognizing areas with high demand, enabling targeted infrastructure development or service provision. Furthermore, utilizing GSM mobile phone data offers real-time, large-scale data collection, which surpasses the limitations of traditional survey methods in terms of sample size and temporal granularity. This approach facilitates a more detailed and timely understanding of travel demand, essential for effective policy-making and transportation system optimization.

In this introductory chapter, we first provide the background information in 1.1 to identify knowledge gaps in 1.3. Next, we formulate our overarching research question in 1.4 followed by the research overall approach (1.5) and sub-questions (1.6). Thereafter, we introduce the scientific contributions and societal relevance in 1.7 and 1.8. Finally, we present an overview of the book's structure in in 1.9. This doctoral research is a part of the MiRRORS project (Multi-scale integrated tRaffic obseRvatory fOr large Road networkS), funded by the Dutch Research Council (Nederlandse Organisatie voor Weteschappelijk Onderzoek or NWO). The overall objectives of MiRRORS are "To develop a new hybrid multi-scale travel demand and supply estimation and prediction methods". The objective of this thesis, as a component of the MiRRORS project, is to develop a new data-driven framework for exploring the spatial and temporal patterns inside travel demand. Moreover, this research explores the sensitivity of travel demand patterns to spatiotemporal scales and aggregated socio-spatial characteristics of the population.

1.1 Background

Travel demand shows people's travel and activity patterns in time and space. By knowing how much travel to expect at different times of day, on different roads, and in different areas of the city, planners can design and operate transportation systems that are efficient, effective, and safe. This can help reduce congestion, improve road safety, and make transportation more accessible for everyone. A reliable estimation and prediction of travel demand can lead to optimization in network use, which also benefits users in their daily trips. Additionally, understanding travel demand patterns can help city officials make informed decisions about infrastructure development, public transit, and other transportation-related issues.

Travel demand estimation and prediction are used to reconstruct and forecast the demand for travel between different locations. Travel demand *estimation* refers to the process of determining the demand for travel based on available data. It is used to understand and quantify the past or current demand for travel between different locations. Travel demand estimation methods use various data such as census or travel survey data, and GPS or GSM data (e.g., Fekih et al. (2022)). Travel demand *Prediction*, on the other hand, refers to the process of forecasting future demand for travel based on historical data and other factors that may affect travel demand. It is used to anticipate and plan for future travel demand. Predictions can be made using statistical models or machine learning techniques that are trained on historical data. There are various factors that can influence the prediction of travel demand patterns, including the time of the day, day of the week, level of aggregation, population demographics, and spatial land use characteristics.

The travel demand is commonly aggregated into *origin-destination (OD) matrices*. An OD matrix is a table showing the number of trips that originates from each zone (or origin) and ends at each other zone (or destination) in the region. The conventional four-step model is a commonly used approach for demand modeling in transportation planning (Levin et al., 2019). It involves four steps of trip generation, trip distribution, mode choice, and route choice steps; the first two relate to travel demand. In the trip distribution step, a *gravity model* is commonly used to estimate (or predict) the OD matrix (Ortúzar & Willumsen, 2011c). These models estimate travel demand between two locations based on the size of the locations and the distance between them. The assumption posits that travel demand between two locations is directly proportional to the size of the locations (e.g., population or employment) and inversely proportional to the distance separating them.

Gravity models for OD matrix estimation rely on trip "production" and "attraction" to predict flows between locations. *Trip production* refers to the ability of a location to generate trips, while *trip attraction* refers to the ability of a location to attract trips. While the OD matrix represents the number of trips made between each origin and destination pair, it doesn't show the underlying factors that influence the decision to make a trip (i.e., trip production) and to select where to go (i.e., trip attraction). This can be done by analyzing the demand at a more aggregate level; first finding patterns in producing trips at origins (and attracting trips at destinations), as investigated in Chapter 3 , and then exploring the relationships between such demand patterns and various characteristics of the area, as studied in Chapter 5.

Trip production (and attraction) can sometimes be measured directly from observed data, such as traffic counts or surveys, but these measurements may be limited in scope, outdated, or not available for all areas of interest. However, the production and attraction of a location can vary over time due to changes in economic conditions, built environment, population demographics, and other factors. This variation makes it challenging to accurately predict travel demand between locations using gravity models, as they often rely on data that may not be readily available or may be outdated, which can affect the reliability of the travel demand estimates.

Therefore, to accurately predict travel production and attraction for a short-term period, it is crucial to collect data on the relevant factors that influence them and to take into account the specific spatial and temporal context in which the model is being used. The influential factors may involve population distribution and demographics, built environment, economic activity, accessibility, and other factors. Using statistical techniques to explore this data and identify patterns in the flow of people between locations will allow for more accurate estimation and short-term prediction of travel demand using a gravity model.

Spatial context affecting demand estimation and prediction relates to the spatial characteristics of the associated area. Spatial context also includes the spatial scale in OD matrix prediction, referring to the geographical area over which the modeled movements are defined. The spatial scale can significantly impact the accuracy and reliability of the OD matrix prediction, as different spatial scales may capture different patterns of movement and may be subject to different factors influencing these movements (Ortúzar & Willumsen, 2011a). For example, a prediction at a fine spatial scale (e.g., at the level of individual streets or buildings) may be more accurate in capturing local variations in movement patterns within a day, day-to-day, and between people. However, it may also be more sensitive to measurement error or other sources of uncertainty. On the other hand, a prediction at a coarser spatial scale (e.g., at the level of entire cities or regions) may be less sensitive to these sources of uncertainty but may also be less able to capture the fine-grained details of individual movement patterns. Therefore, it is essential to carefully consider the appropriate spatial scale for an OD matrix prediction based on the specific context and objectives of the modeling effort (as explained in Chapters 4 and 5).

The classic gravity model estimates travel demand between locations based on trip production and attraction, and a deterrence function typically related to distance or travel impedance. While the deterrence function can be calibrated using observed travel behavior and may incorporate factors like travel time or cost, the gravity model operates at an aggregate level and may not fully capture individual-level factors influencing travel behavior, such as personal preferences, income, or the specific attractiveness of destinations beyond general measures. To address these limitations, more advanced models, including disaggregate trip distribution models like destination choice models and agent-based models, have been developed (see Ortúzar & Willumsen (2011c,b)). These models simulate the decision-making processes of individual travelers or households, allowing for a richer representation of travel behavior by considering individual characteristics, preferences, and accessibility to different destinations. This approach enables the estimation of the impact of changes in land use and transportation infrastructure on travel behavior with greater detail.

Despite the advantages of disaggregate models in providing a more detailed and accurate representation of travel behavior by considering the characteristics of individual agents, aggregate models such as the gravity model are still widely used for large-scale travel demand modeling (Ortúzar & Willumsen, 2011a). This is because they are relatively simple and easy to implement, and require less data and computational resources than disaggregate models. The gravity model, for example, only requires data on the total number of trips between different zones or areas, which is often readily available, and does not require data on individual agents or households. Additionally, aggregate models can provide a good first approximation of the origin-destination flows and are often used as a benchmark for more complex models (e.g., Calabrese et al. (2011)).

Regarding the data sources for travel demand estimation and prediction, GSM mobile phone data has several potential advantages over traditional travel survey methods. It can provide a more comprehensive and accurate picture of travel patterns, as it captures information on trips made by a large number of individuals, rather than relying on self-reported data from a smaller sample of respondents. Additionally, GSM data can be collected in real-time, allowing for more up-to-date analysis of travel patterns. These advantages have led to the increasing use of GSM mobile phone data in transport planning (e.g., Calabrese et al. (2011); Tolouei et al. (2017)).

One can use GSM mobile phone data in the four-step model of transport as follows:

- In the *trip generation* step, by analyzing the location data of mobile phones to identify the trips originating from a specific location and the time of day to determine the frequency of such trips.
- In the *trip distribution* step, by analyzing the GSM data to identify the destinations of trips and develop the “prior” origin-destination matrix.
- In the *modal split* step, we can infer the mode of transportation used for each trip from GSM data by analyzing the travel distance, duration, and stay at specific locations.
- In the *traffic assignment* step, by analyzing the GSM traces to assign the trips to the transportation network by matching the location data to the network.

Estimating OD patterns using GSM mobile phone data can effectively show human mobility and behavior. However, the accuracy of the estimates will depend on several factors. One important factor is the availability and quality of the data. Mobile phone data can infer OD patterns by tracking the locations of devices as they move around. However, the accuracy of these estimates will depend on the tracking (i.e., polling) frequency and the location accuracy of the devices in the data. Another factor affecting the accuracy of OD estimates using mobile phone data is the method used to analyze the data. Different methods use different assumptions to infer OD patterns, and the choice of method can significantly impact the accuracy of the estimates. Overall, the accuracy of OD estimates using mobile phone data will depend on the availability and quality of the data and the method used to analyze it (as explained in Chapter 2).

It is worth noting that GSM data is not the only data that can be used in a four-step model, and it’s not always possible to have all the data for each step. Furthermore, GSM data is usually only available for a subset of the population that might not represent the whole population. To overcome such limitations, it is essential to use a combination of data sources and carefully consider the sample’s representativeness when interpreting the results.

1.2 Literature review

Figure 1.1 provides an overview of travel demand data and analysis across aggregate and disaggregate levels. The term *aggregate* refers to data or analysis conducted at broader scales, such

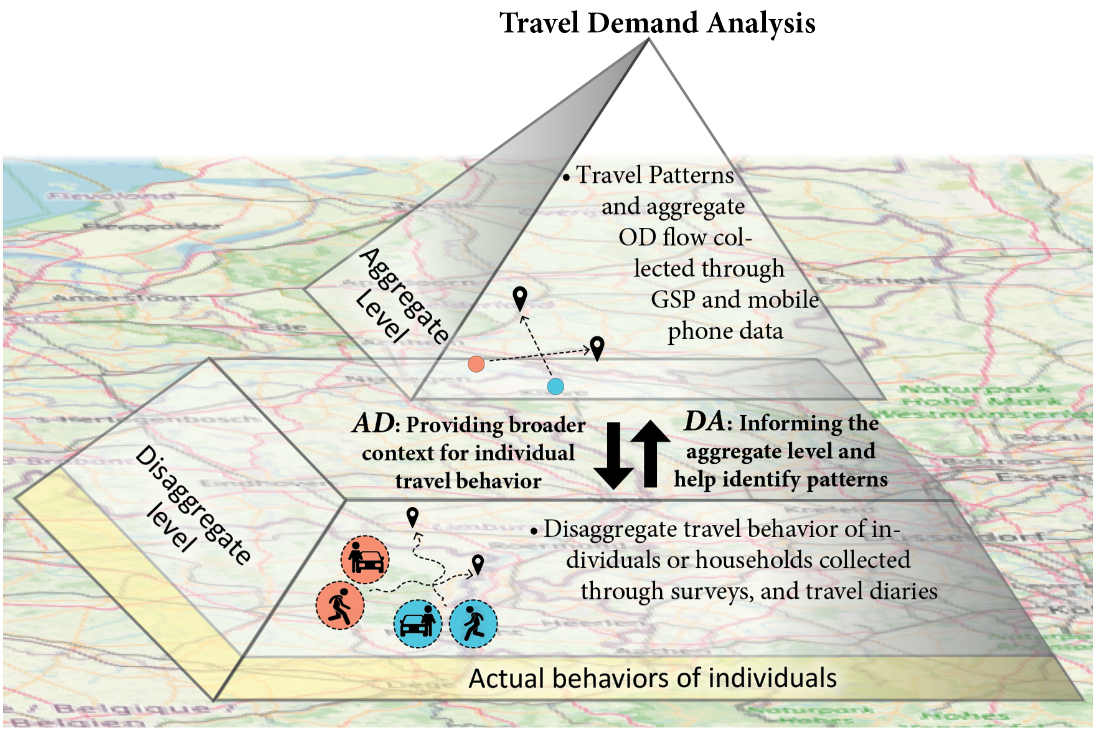


Figure 1.1: Conceptualization of travel demand at aggregate and disaggregate levels.

as zones, neighborhoods, cities, or provinces, while *disaggregate* pertains to data or analysis at the individual or household level. The complexity with which spatial urbanization, socio-economic factors, travel characteristics, and travel demand are represented varies significantly between these levels of data and analysis. According to Figure 1.1, individuals make trips at the lowest level of abstraction. Then on one end of the spectrum, at the disaggregate level, choice models and activity-based analyses center around individuals or households. While on the opposite end, aggregate studies employ average characteristics of cities or zones.The analytical techniques employed differ across these levels, spanning from basic correlation tests to the estimation of theoretically grounded behavioral models (Handy, 1996).

Traditional travel demand analysis often relies on survey data, which can be time-consuming to collect, limited in sample size, and may not accurately reflect real-time travel behavior due to recall bias or infrequent updates. In contrast, using GSM mobile phone data enables the collection of large-scale, real-time information on people's movements, capturing both spatial and temporal variations in travel demand. This approach allows for more accurate and timely analysis, which is crucial for designing responsive transportation policies and interventions.

By focusing on spatial and temporal patterns, our research addresses the limitations of traditional methods and provides a deeper understanding of travel demand dynamics. This enhanced understanding can inform more effective transportation planning and policy decisions, such as optimizing public transit schedules to match actual demand patterns, improving traffic management strategies, and guiding infrastructure investments to areas with growing travel needs.

Disaggregate travel behaviors are collected through surveys and travel diaries. The behaviors are observed in relation to contextual information such as land-use, socio-economic factors, and personal preferences. Models at this level hold significant policy implications since they enable estimation of the potential impact of minor policy changes on people's decision-making processes. Moreover, the application of disaggregate modeling techniques enables the connection of individual-level attributes to system-level characteristics, encompassing household and societal-level factors (Ghasri et al., 2017). Agent-based, activity-based and choice models are typically applied to study individual travel behavior at this level.

While disaggregate models offer the advantage of a more detailed and accurate representation of travel behavior by accounting for individual agent characteristics, aggregate models like the gravity model remain widely utilized for large-scale travel demand modeling (Ortúzar & Willumsen, 2011a). This preference arises from their simplicity, ease of implementation, and lower requirements for data and computational resources compared to disaggregate models. For instance, the gravity model only requires data on the total number of trips between different zones or areas—information that is often readily available—without necessitating data on individual agents or households. Furthermore, aggregate models can serve as a reliable first approximation of origin-destination flows and are frequently used as benchmarks for evaluating more sophisticated models (e.g., Calabrese et al. (2011)).

In this thesis, we focus on exploring travel demand patterns at the aggregate level using large-scale data sources, such as GSM mobile phone data. While we recognize the value of disaggregate models in capturing individual-level behaviors, our approach leverages the advantages of aggregate models to analyze population-level patterns in travel demand. This focus allows us to address practical challenges in transportation planning that require understanding collective travel behaviors across regions, especially when individual-level data may be limited or unavailable.

Despite significant advancements in computational capabilities, simulation cost remains a crucial issue associated with disaggregate models (Pendyala et al., 2010). Furthermore, the expense of substantial data gathering and model development is a challenging factor, particularly for small regions with limited funding resources (Stopher et al., 2003). In an effort to overcome these limitations, Ghasri et al. (2017) have presented a preliminary demand modelling framework that utilizes the concept of transferability demand models and evaluates its effectiveness. The key feature of their approach is their focus on replicating data structures at a highly disaggregated level, explicitly pertaining to individuals and households. By utilizing a bottom-up approach that captures individual behavior, the proposed framework allows for effective policy analysis. Furthermore, this framework is computationally efficient when compared to activity-based models.

At the aggregate level, collective travel patterns of individuals at a large scale are investigated. Analysis at this level provides insights for transport planners and policy-makers, but these insights are not behaviorally intensive as disaggregate models (Axhausen & Gärling, 1992). Travel demand data collected at this level is either collected at an aggregated level or has been collected at the disaggregate level but later aggregated. For example, GPS and mobile phone data are typically aggregated due to privacy concerns.

The travel demand field encompasses various levels of abstraction, ranging from individual trips to collective travel patterns at a large scale. The accuracy of models used at each level depend on various factors, such as data accuracy and resolution, the computational resources available, and the objective. Understanding the relationships and interplay between the different levels of analysis can inform transportation planning and policy-making.

In this regards, the literature has identified two primary perspective of travel demand analysis: *AD* and *DA*. In *AD*, the main input data and results are at the disaggregate level, investigating travel behavior features of individuals. However, data from the aggregate level has been used to provide a broader context for individual travel behavior.

For instance, the article by Zhou et al. (2022) explores the relationship between travel demand patterns and demographic information, and land-use characteristics in older adults. The study highlights that older adults have distinct travel patterns compared to younger adults, characterized by more out-of-home activities, longer trips for medical purposes, and frequent trips accompanying their grandchildren. The study also underlines the importance of considering land-use characteristics in understanding travel demand patterns. The built environment is found to significantly shape the travel patterns of older adults, particularly with regard to proximity to recreational facilities and active travel behavior. These findings emphasize the need to take demographic information and land-use characteristics into account in understanding travel demand patterns and improving accessibility and quality of life for older adults.

In contrast, *DA* approach studies the travel demand on a large scale. Analysis at this level primarily relies on aggregate data' although data from the disaggregate level, such as built environment and demographics, are used to help identify travel demand patterns and trends. Moreover, many studies use aggregate data sources, which have been available for less than two decades, to analyze large-scale travel demand estimation and prediction. The emergence of these novel data sources, such as GPS and mobile phone records, has facilitated the monitoring of the variation of urban travel demand, thus aiding in the development or adjustment of transportation planning and policies.

Leveraging mobile phone signaling data for travel demand modeling and transportation planning offers a cost-effective opportunity for practitioners, planners, and policymakers, especially given the increasing complexity of transport networks. Ensuring the quality and accuracy of data is essential for making reliable analyses that inform investment and transport policy decisions. In this regard, Fekih et al. (2021a) propose a cell phone activity indicator-based filtering pipeline to process signaling data to make it useful for mobility and transportation purposes, which could serve as a preliminary guideline that can be tailored to the specific case study. Additionally, the authors emphasize the significance of examining how the location accuracy of signaling data affects components like home location identification and trip detection, as well as analyzing travel patterns such as travel time and their consistency with travel survey data. The authors propose potential enhancements to the suggested pipeline, such as conducting a more in-depth analysis of the stationary time threshold and the assumptions underlying the trip expansion method, based on the identified home locations derived from the signaling data of a single operator. Finally, as mobile phone usage continues to rise, cell network-based traces are anticipated to generate higher-frequency data, encompassing a larger portion of the population. This advancement enables the estimation of movements with finer temporal granularity compared to traditional travel survey data. These emerging individual-based big data offer opportunities to deepen the understanding of less-explored mobility patterns, particularly during special periods and events.

Another study by Bonnel et al. (2015) focuses on the potential and limitations of using passive mobile phone data to construct origin-destination matrices. The study used data collected from the mobile phone network of Orange in the Ile-de-France region in 2009. Despite being able to generate matrices from the data collected from the Orange mobile phone network, the **validity** of these matrices has not been thoroughly investigated. Because they plot census commuting data against their home-based work mobile phone trips. Then they use only the R2 of the analyses as an indicator of quality. Later, the authors compare the matrices with the results from a household travel survey; however, they find limited results and a high degree of dispersion in the comparison. The authors note that a more detailed analysis may improve the comparability and that identifying home and work locations from mobile phone data is crucial in enhancing the matrices. The authors also acknowledge that further research is necessary to validate the use of mobile phone data in transport modeling and planning based on journey purpose split, distribution over time of day, or trip length distribution. They also find that the results are highly sensitive to the assumption of a minimum stationary time of one hour and suggest that this assumption should be refined and varied based on the characteristics of each location area.

In another study, Burkhard et al. (2020) investigate the impact of spatial accuracy and temporal granularity on transport mode detection using passively sensing mobile phone data. The study applies commonly used methods for mode classification and evaluates the results for different levels of spatial uncertainty and temporal granularity. The results are based on a dataset collected over half a year from 130 users who annotated their data. The authors found that the level of accuracy required for mode detection depends on the chosen classification scheme. The study also highlights that the reduction in classification quality is approximately linear with the standard deviation of the sampling error. Furthermore, contextual information can mitigate the effect of quality deterioration, while temporal sparsity is found to be more detrimental than spatial uncertainty. The authors suggest that improvements in passive data collection should prioritize increasing the sampling rate to a range of 30 seconds to a minute while addressing spatial accuracy with lower priority, particularly when temporal granularity is high.

Another area of active research in transportation using *DA* approach is focused on investigating the correlation between contextual information such as built environment, socioeconomic variables, special events, and specific times, as well as the analysis resolution on travel demand. For instance, the study of Liu et al. (2022) explores the impact of spatiotemporal granularity on the demand prediction of dynamic ride-hailing. The authors observe that travel demand exhibits a strong spatiotemporal correlation, with temporal factors such as rush hours, weekends, and holidays influencing demand in a highly periodic manner. Moreover, the demand for a specific region is shaped not only by variables within that region but also by factors across the entire network.

Zhao et al. (2022) focuses on predicting short-term bus travel demand, considering both the spatial-temporal influence of the built environment and graph deep learning. They conclude that given the ability of deep learning to capture spatial and temporal dependencies, the prediction performance of these models can be further improved by considering the dynamic influence of the built environment on bus travel demand. In this study, the authors used a time-dependent geographically weighted regression model to explicitly capture the dynamic influence of the built environment on bus travel demand. The results showed that this approach improved the prediction accuracy compared to a model that only considered the built environment as static features.

Finally, Xu et al. (2022) conduct a detailed analysis of how natural environmental and socioeconomic factors affect urban travel demand. They contend that relying solely on the spatial-temporal characteristics of traffic data is inadequate for accurately predicting urban travel demand. Instead, they advocate for a more comprehensive approach that incorporates the complex patterns shaped by environmental and socioeconomic influences. The study highlights the importance of considering these factors in predicting urban travel demand, which is crucial for effective transportation policies and strategies. The study highlights the need for continued research in exploring environmental and socioeconomic factors affecting urban travel demand.

1.3 Knowledge gaps

In this section, we will highlight several knowledge gaps related to the predictability of travel demand that this thesis aims to address:

1. *There is a significant knowledge gap in the standardized approaches for understanding the temporal* ***quality of GSM data*** *and how it impacts the accuracy of origin-destination (OD) estimates.* Traditionally, a "stay and pass-by" model is applied to GSM data for OD detection, which assumes that individuals either stay in a particular location for a certain amount of time or pass through it quickly without stopping. However, different studies have used different methods for determining when an individual is considered to be "staying" in a location, and there is currently no consensus on the most appropriate method. This lack of standardization encompasses not only the criteria used to define stays and trips (such as minimum duration thresholds, distance moved, or speed) but also the preprocessing steps, data filtering techniques, and validation methods employed. Without standardized methodologies, it's challenging to assess the quality and reliability of OD estimates derived from GSM data or to compare findings across different studies and contexts.

   Additionally, GSM data is typically collected at the level of individual phone calls or texts rather than continuously tracking an individual's movement over time, leading to a knowledge gap in understanding the temporal quality of this data and how it impacts OD estimates. The infrequency of records can make it challenging to accurately estimate the duration of an individual's stay in a particular location or to understand the spatio-temporal patterns of their movement.

2. *There is a major knowledge gap in our understanding of* ***spatial and temporal patterns*** *in the trip production* . This is especially relevant for accurately estimating and predicting OD matrices using trip distribution models. However, transportation systems are complex and dynamic, making it difficult to accurately model and predict travel demand, even with accurate data. This complexity and dynamics contribute to spatiotemporal heterogeneity in trip production (and attraction), which can impact the reliability of OD matrix prediction. In fact, when using the gravity model to estimate or predict the OD matrix, there is a significant knowledge gap related to trip production variability depending on the time of day, day of the week. For example, a region may have higher trip production during morning peak hours compared to afternoon peak hours, or certain zones may have higher production than others. People's travel patterns and behaviors can also change over time, further complicating the prediction of future trip production.

3. *One potential research gap in estimating and predicting travel demand is* ***understanding how spatial scale—specifically, the level of spatial aggregation, such as traffic analysis zones, cities, or regions—affects the accuracy and reliability of predictions****. There is a lack of studies that directly compare the prediction of travel demand across different spatial scales. This lack of knowledge makes it difficult to determine which spatial scale best suits a specific study objective or context.*. This lack of knowledge makes it difficult to determine which spatial scale best suits a specific study objective or context. Additionally, without such comparisons, it is challenging to make reliable generalizations about the predictability of travel demand across various spatial scales. To address this research gap

and improve the reliability of travel demand predictions, a direct comparison of travel demand predictions at different spatial scales is necessary. While there is a significant amount of research on travel demand prediction at various spatial scales, it is often difficult to compare the results of these studies directly due to differences in study design and the specific factors being considered.

4. *Another potential research gap in trip production prediction is the correlation between* ***dominant demographic and land-use information*** *(i.e., socio-spatial characteristics) and travel demand patterns at different spatial scales.* Collecting and exploring a large amount of data on factors such as travel times, population density, and economic activity can help to understand these complex relationships better and improve the accuracy of OD matrix predictions. While demographic data is commonly used to inform trip production models, there is a lack of understanding about how different demographic factors impact the accuracy of these models and how they vary at different spatial scales. For example, it is unclear how age, income, household size, or employment status affect the patterns of trip production or whether these factors have different impacts in different contexts. Therefore, further research is needed to understand these relationships better and to identify the most relevant factors for predicting trip production at different spatial scales.

1.4 Main research question

Building on the identified knowledge gaps in understanding the relationship between input data and estimated demand, as well as the necessity of developing a data-driven framework to investigate travel demand patterns, we can now articulate the primary research question:

What are the spatial and temporal patterns of travel demand considering the input data quality, spatio-temporal context, the objective spatial scale, and the socio-spatial characteristics?

The following section outlines the primary research approach that enables us to systematically address this critical research question.

1.5 Overall research approach

This thesis employs data-driven methods to explore and predict travel demand patterns. Demand pattern prediction methods can be broadly divided into two categories based on the balance of data-driven and theory-based models. **Mechanistic models**, which use a conceptualization of the system together with theoretical and parameterized relationships, require less data in their calibration as these are bound by theory. Therefore, they are well-suited for non-recurring scenarios where enough data may not be available. However, these models require a well-proven theory of the system we want to model. Otherwise, they may introduce invalid assumptions into the estimation (prediction) results.

Phenomenological models, also referred to as data-driven models, on the other hand, use agnostic and high flexible machine learning structures. Therefore, they require much more historical data on phenomena to calibrate the appropriate statistical inferences. These data for transport modeling include the distribution of activities, and modes of transport in time, as the input for machine learning methods. Extensive data availability and the minimum number of involved assumptions have resulted in more enthusiasm for these flexible methods when no proven theory is available. However, these methods treat the phenomena like a “black-box” — challenging to explain the outcome in terms of traffic theory. Besides, the nature of learning from historical data and having no traffic theory make future decisions constrained by what it learns from that data. In other words, it is not straightforward for the data-driven approach to predict traffic in case of unseen incidents or accidents that are not present in the data. Since these methods do not consider causality, they are not suitable for non-recurrent situations.

This thesis focuses on a data-driven (i.e., phenomenological) modeling approach to predict travel demand patterns. Data-driven models offer a high level of insight into the fundamental physical processes that shape transportation patterns, which can be helpful in situations where more detailed, context-specific theories are not available or appropriate. As a result, these approaches are generally more generalizable and potentially applicable to a broader range of contexts. Our approach differs from typical phenomenological models because we don’t treat phenomena as a black box. Instead, we perform sensitivity analyses to uncover the underlying system theory and relationships. By studying the underlying mechanisms driving transportation patterns, data-driven models can inform the development of targeted approaches better suited to specific contexts, i.e, the theory . Additionally, these models are often built from the ground up, starting with basic principles and adding complexity as needed, making them flexible and adaptable to different situations.

By employing a data-driven modeling approach that utilizes GSM mobile phone data, we can capture real-time, high-resolution travel behavior without relying solely on traditional survey methods, which may be limited in scope and frequency. This approach allows us to identify emerging trends and patterns in travel demand that are critical for adaptive transportation planning and policy-making.

While the machine-learning techniques adopted in this thesis are standard within data-science practice, their systematic tailoring to production-side demand modelling and their cross-comparisons at several spatial scales have, to our knowledge, not been documented before for the Netherlands. Emphasis is thus placed less on algorithmic novelty and more on (i) methodological transparency, (ii) reproducible benchmarking across scales, and (iii) domain-specific insights that can be transferred to demand modelling.

Furthermore, by integrating machine learning techniques with transportation analysis, we can develop models that are both flexible and capable of handling complex, nonlinear relationships inherent in travel behavior. This results in more accurate predictions and a better understanding of the factors influencing travel demand, enabling policymakers to design interventions that are responsive to current and future needs.

To unravel travel demand patterns, we employ various machine learning methods, including pattern recognition and clustering algorithms to identify patterns in the data, as well as deep learning models to predict future travel demand. Additionally, we use statistical methods such as hypothesis testing and analysis of variance to validate our findings and draw robust conclusions about travel demand patterns.

1.6 Research sub-questions

To achieve this thesis aim, we raise four research sub-questions. The following sub-questions are associated with the four mentioned elements of the main research question and the four knowledge gaps we identified.

1. *How do temporal aggregation and discretization of input data (indicating the input data quality) affect the demand estimation?*

2. *What are the dominant temporal patterns of trip production?*

3. *What is the relationship between spatial scale and prediction of trip production?*

4. *What is the relationship between the patterns of trip production and their associated overall socio-spatial features?*

1.7 Research contributions

Our research contributes to travel demand modeling by introducing a data-driven framework capable of quantifying the predictability of travel demand across multiple spatial and temporal scales. This framework utilizes advanced machine learning and statistical techniques to analyze available data sources, providing policymakers and traffic managers with valuable insights that can aid in the decision-making process related to the transport network. The development of this innovative modeling approach represents an advancement in the field and has the potential to improve our understanding of travel demand patterns, leading to more effective transport planning and management. Our study offers the following contributions categorized under four perspectives:

Spatio-temporal idiosyncrasies of input data for demand estimation

- A new data-driven framework based on Kernel density estimation and Bayesian classification for preprocessing and interpreting the raw GSM data for OD matrix estimation. This framework adopts the minimum number of assumptions on trip patterns which reduces bias toward longer-duration trips or longer-duration activities. This generalization makes it easy to compare results across studies with different contexts.

- New insight into the impact of temporal aggregation and discretization of input data on the OD matrix reconstruction. This understanding helps transport planners identify the appropriate context-specific polling frequency of mobile phone data and filter the more informative traces for more accurate OD matrix estimation.

These contributions help transport planners and analysts to process and interpret GSM data more effectively, leading to more accurate OD matrix estimations. Accurate OD matrices are foundational for developing reliable travel demand models, which in turn inform infrastructure planning, congestion management strategies, and service scheduling. By understanding how temporal aggregation affects OD estimation, planners can select appropriate data collection frequencies to capture essential travel behaviors, ensuring that models reflect actual travel patterns and support effective decision-making.

Distinguishing the spatial-temporal travel demand patterns

- A new data-driven framework that uses deep neural networks and K-means clustering to identify dominant temporal patterns of trip production. This framework automates the trip production pattern recognition, which reduces bias toward specific patterns and number of underlying clusters. This generalization makes it easy to compare results across different studies.

- A new analytical data-driven framework that makes use of a Gradient Boosting algorithm and hierarchical clustering to find the association between dominant temporal patterns and spatial land-use features.

- A new validation method of data consistency without ground truth which help get some insight into the validity of processed aggregated OD data using unknown assumptions or methods without having the ground truth.

By identifying dominant temporal patterns of trip production and their association with land-use features, planners can understand how different areas generate travel demand throughout the day and week. This information is critical for optimizing public transit schedules, designing demand-responsive transport services, and planning infrastructure improvements that align with actual usage patterns. For example, areas with significant afternoon peaks in trip production may require increased transit services during those times, while areas with consistent demand throughout the day might benefit from different strategies.

Prediction of travel demand at multiple spatial scales

- A new data-driven framework using a graph-based deep neural network to incorporate spatial adjacency of traffic analysis zones into the trip production prediction. This insight makes it possible to compare the results of demand prediction under multiple spatial scales.

Developing a predictive framework that operates across multiple spatial scales enables planners to make informed decisions at various administrative levels, from local neighborhoods to entire regions. By incorporating spatial adjacency into demand prediction, the models can capture the influence of neighboring zones, leading to more accurate predictions. This is particularly valuable for regional planning efforts where coordination between adjacent areas is essential for effective transportation system development.

Sensitivity of travel demand patterns to socio-spatial characteristics

- A new data-driven computer vision framework to identify dominant temporal patterns of the prediction error of trip production. This framework helps improve the prediction of travel demand using the gravity model.

- New insight into the possible spatial features to reduce the prediction error of trip production. As such, this insight improves the prediction of spatiotemporal patterns inside travel demand.

Understanding how socio-spatial characteristics influence prediction errors allows planners to refine their models and address specific factors that may affect travel demand. By identifying the key features that reduce prediction errors, such as certain land-use types or demographic variables, planners can enhance model accuracy and reliability. This leads to better-informed policies that are tailored to the unique needs of different communities, ultimately improving transportation system performance and user satisfaction.

1.8 Practical relevance

This research is particularly relevant for policymakers and planners seeking to optimize transportation infrastructure to develop more effective strategies for planning and managing transportation systems that better serve the needs of their communities. By providing a data-driven framework that leverages GSM mobile phone data, this research offers advanced tools for understanding and predicting aggregate travel demand patterns in space and time.

In contrast to traditional travel demand models, which often rely on infrequent surveys and historical averages, our research leverages large-scale, timely GSM mobile phone data to capture dynamic and fine-grained travel demand patterns. This enables transportation planners to identify areas and times with significant fluctuations in overall travel demand, allowing for more informed decisions regarding resource allocation and strategic planning. For example, planners can prioritize infrastructure maintenance in high-demand areas or adjust development plans based on observed demand trends.

Moreover, the ability to analyze travel demand at multiple spatial scales supports coordinated planning efforts across different administrative levels. This ensures that local, regional, and national transportation policies are aligned and effective, facilitating integrated transportation and land-use planning.

By identifying the key factors that influence travel demand prediction, researchers and practitioners can develop more robust models that accurately forecast travel demand. This enables the implementation of targeted transportation policies and measures, such as travel demand management strategies, promotion of flexible work arrangements, and informed land-use policies, ultimately leading to more efficient, sustainable, and user-centric transportation systems.

Furthermore, by refining travel demand models through error pattern identification and integrating socio-spatial features, planners can enhance model accuracy and reliability. This leads to better-informed policies that are closely aligned with community needs, promoting efficiency and user satisfaction. Recognizing error patterns also highlights areas where additional data collection or research is needed, guiding future efforts to enhance model performance.

Overall, the practical contributions of this research are geared towards enhancing transport planning and policy-making, anticipating and responding to actual travel behaviors, and contributing to a more sustainable, efficient, and user-centric transportation system.

1.9 Thesis outline

This thesis is divided into several chapters that address the research questions posed in the introduction. These chapters are based on articles written during the Ph.D. process. For some chapters, the text is identical to these articles in order to avoid potential citing conflicts, which may result in some repetition among chapters. However, in other instances, chapters are adapted from articles and may combine or expand upon them to better fit the thesis structure.

Specifically, Chapters 3, 4, and 5 are based on two papers but are presented as three chapters in this thesis. This division aligns better with the research sub-questions outlined in the introduction. At the beginning of each chapter, the corresponding article(s) are identified along with their publication status. The final chapter discusses the conclusion and implications of this research.

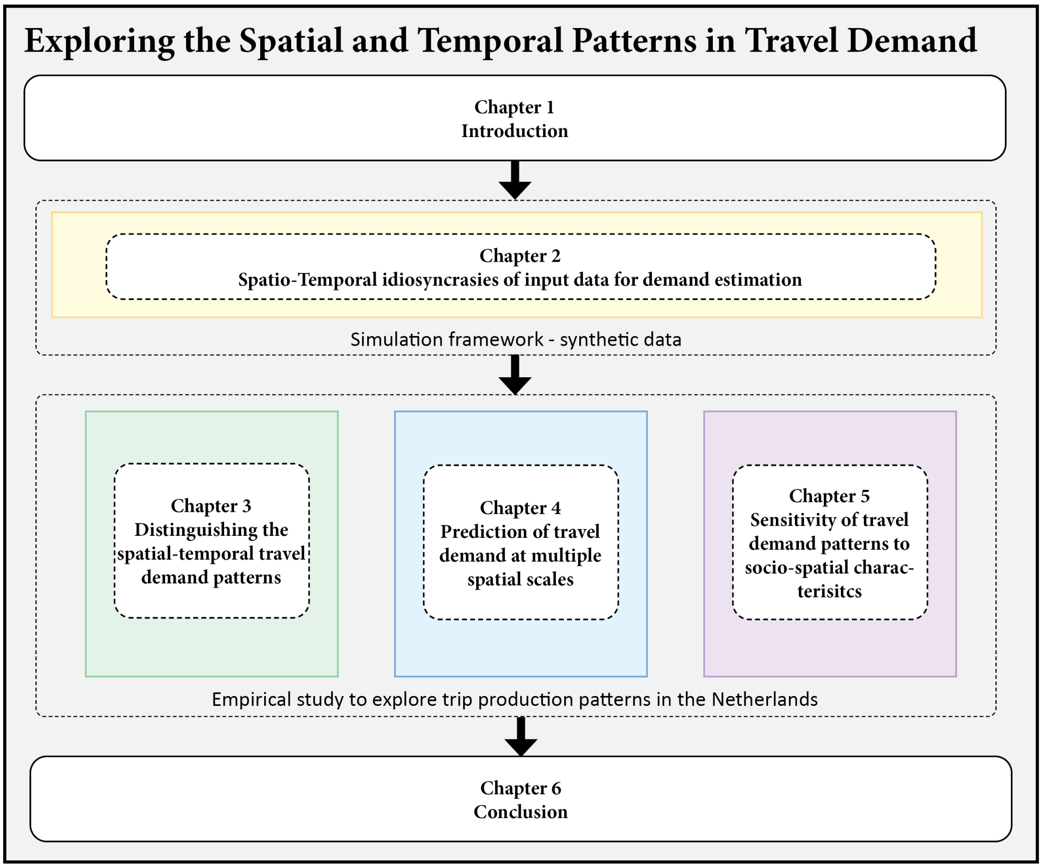


Figure 1.2: Structure of the thesis.

Figure 1.2 presents an overview of the thesis structure. Accordingly, Chapter 2 of this study involves the use of a simulation test bed to evaluate the impact of mobile phone data quality on the estimated origin-destination (OD) matrix. The remainder of the thesis focuses on an empirical study of trip production in the Netherlands.

Chapter 3 identifies spatial-temporal patterns in trip production using a data-driven pattern recognition framework. In this chapter demand patterns of traffic analysis zones are distinguished based on the dominant within-day and day-to-day trip production. These patterns show correlation with spatial characteristics which are investigated in Chapter 5. Chapter 4 gives insight into how spatial scale affects the prediction of trip production using graph-based neural networks. Moreover, the effect of considering zone adjacency for the prediction at various spatial scales are studied. Chapter 5 evaluates the sensitivity of trip production prediction to socio-spatial characteristics under different spatial scales. To achieve all these objectives, real-world data sources will be analyzed using machine learning and statistical techniques. By exploring demand patterns, this research aims to improve the accuracy of trip production prediction in the Netherlands and inform transportation planning and policy decisions.

This research offers a novel approach for exploring travel demand patterns. A new data-driven framework is developed to find patterns of travel demand through the combination of machine learning methods from various disciplines. The study begins by synthesizing a test bed to demonstrate the effects of temporal aggregation and discretization of raw data on travel demand in Amsterdam. This is followed by an empirical analysis of dominant patterns of trip production in time and space in the Netherlands. Another empirical study explores the trip production patterns at different spatial scales in the Netherlands, considering the adjacency of traffic analysis zones. Finally, another empirical study identifies dominant patterns of prediction errors in trip production and investigates their relationship with socio-spatial characteristics.

In summary, this thesis aims to enhance our understanding of spatial and temporal travel demand patterns through a data-driven approach utilizing GSM mobile phone data. By addressing the identified knowledge gaps and developing innovative analytical frameworks, we contribute to the advancement of travel demand modeling. The practical implications of this research offer policymakers and planners the tools and insights needed to develop transportation systems that are efficient, responsive, and aligned with the actual needs of communities. This work supports the creation of transportation policies that are data-driven and adaptable, ultimately improving mobility, reducing congestion, and promoting sustainable urban development.

Chapter 2

Spatio-Temporal Idiosyncrasies of Input data for Demand Estimation

Building upon the foundational insights introduced in the first chapter of this thesis, this chapter delves deeper into the nuanced intricacies of GSM data in the realm of transportation planning. This chapter is a pivotal extension of our exploration, aiming to bridge a critical knowledge gap: comprehending the temporal quality of raw GSM data and its consequential impact on the accuracy of origin-destination (OD) estimates.

In pursuit of a more robust method for interpreting GSM data, this chapter proposes a novel, data-driven approach. This initiative addresses the prevalent ambiguity in existing studies regarding the criteria for determining an individual's 'stay' duration in a location. By critically examining the challenges and inherent limitations of GSM data, particularly concerning the infrequency of records, we seek to provide a clearer picture of its implications on the accuracy of the OD matrix. A key aspect of our investigation is analyzing how varying polling frequencies influence the integrity of the reconstructed OD matrix, a question of significant importance for effective transport planning.

This chapter is based on the following published paper:

Eftekhar, Z., Pel, A., & van Lint, H. (2023). Effects of Periodic Location Update Polling Interval on the Reconstructed Origin–Destination Matrix: A Dutch Case Study Using a Data-Driven Method. Transportation Research Record, 2677(9), 292–313.

Abstract

GSM data is a valuable source to understand travel demand patterns as these contain traces of people's consecutive locations. A major challenge however is how the polling interval (PI; the time between consecutive location updates) affects the accuracy in reconstructing the spatio-temporal travel patterns. Longer PIs will lead to lower accuracy, and may even miss shorter activities or trips, when not properly accounted for. In this chapter, we analyse the effects of the PI on the ability to reconstruct an origin-destination (OD) matrix, as well as propose and validate a new data-driven method that improves accuracy also in case of longer PIs. The new method first learns temporal patterns in activities and trips, based on travel diaries, that are then used to infer activity-travel patterns from the (sparse) GSM traces. Both steps are data-driven thus avoiding any a-priori (behavioral, temporal) assumptions. To validate the method we use synthetic data generated from a calibrated agent-based transport model. This gives us ground-truth OD patterns and full experimental control. The analysis results show that with our method it is possible to reliably reconstruct OD matrices even from very small data samples (i.e., travel diaries from a small segment of the population) that contain as little as one percent of the population's movements. This is promising for real-life applications where the amount of empirical data is also limited.

2.1 Introduction

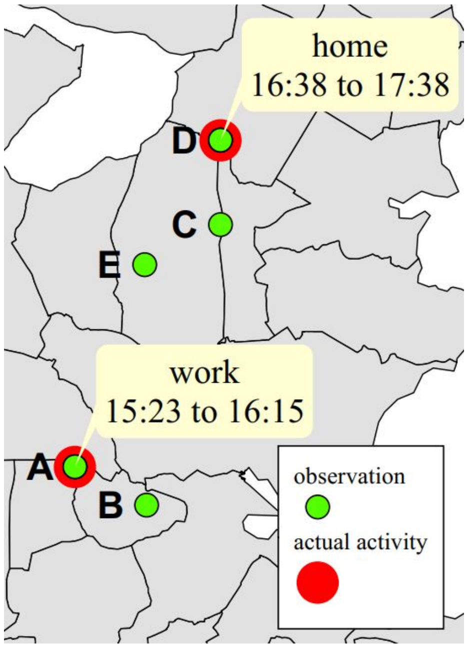
The design of transport infrastructure, services, policies and technology all start with an understanding of travel demand. Travel demand relates to people's spatial and temporal patterns of activity locations and associated trips from one location to the next, and are commonly aggregated into origin-destination (OD) matrices. One data source in this is GSM data as it allows to trace people (carrying the mobile phone). The time between consecutive location updates is called polling interval, and evidently affects the accuracy with which we can reconstruct people's spatio-temporal travel patterns.

Traditionally, traffic planners use direct methods, including roadside and household surveys, conducted every 5-10 years (Wang et al., 2013) for estimating the OD matrix. While these methods are making essential contributions to the traffic demand field, exclusively using survey data makes the estimations liable to sampling bias and reporting errors (Hajek, 1977; Kuwahara & Sullivan, 1987; Groves, 2006). Because travel surveys provide a high level of detail (LOD) in terms of activity and movement behavior with minimal sampling ratios. On the other hand, mobile phones have generated a wealth of low-cost GSM data on people's movements. These movement traces are often of reasonable sample size but contain (much) less detail than survey data. Generally, GSM data contain discretized traces of users without precise indicators of time and location of the underlying activities or activity types. Therefore, only after analysis one can use it to estimate the OD matrix. Combining travel diaries in such GSM analyses could potentially lead to the best of both worlds; that is, the high sampling ratio of mobile phone data combined with the high LOD (in terms of spatial and activity patterns) in travel diaries.

Fundamentally, two types of synthetic and real-world data sets are used in demand estimation research. Real-world GSM data sets can offer voluminous information about millions of mobile phone users. However, the problem with them is the privacy issue and difficulty of carrying out reliability and validation experiments. In fact, in the early analysis phase of research,

time of observation	location
15:35	A
15:50	A
16:05	A
16:20	B
16:35	C
16:50	D
17:05	D
17:20	D
17:35	D
17:50	E

(a) A prototype of GSM traces



(b) Visual representation of traces in (a)

Figure 2.1: An example of the GSM records of a user

using synthetic data for assisting in the operational tests and evaluation has been strongly advocated (Zilske & Nagel, 2014). Conducting experiments using such data helps us to evaluate the effects of various potential components in our models. Therefore, in this research we used synthetic (travel diary and GSM) data to validate our methods. Moreover, it enabled us to produce many ground truth training sets for statistical learning and reliability testing.

Our method addresses some key questions regarding the accuracy and robustness of OD matrices estimated from GSM data related to the *temporal* discretization. Clearly, such discretization errors could be reduced by choosing smaller polling intervals between records. However, this is often not possible since, by their very nature GSM records are spatially coarse and temporally infrequent (e.g., Becker et al. (2013); Burkhard et al. (2017); Chen et al. (2018)). To date, few studies have investigated the relationship between GSM quality and mobility pattern detection (e.g., Calabrese et al. (2013); Chen et al. (2014)). To the best of our knowledge, no study has looked specifically at the effects of GSM data temporal frequency on the estimated OD matrix. This research is the first to examine these effects. To do this, we quantify the temporal quality of the data using **polling interval** (PI), defined as the time interval between consecutive records, where a record is an update on the current location of the mobile device. (Note that the inverse of the polling interval is referred to as **polling frequency**.)

The fact that our temporal record of events is discretized (with the PI) has two consequences. The first effect is that the recorded start and end times of each underlying event (an activity or travel between activities) occur later than the actual start and end times. This discretization error can vary from *zero* to the *PI* value. For instance, if we poll the user epsilon before the actual event time, the interpreted time is about one *PI* later.

To illustrate, Figure 2.1(a) shows several GSM records of a user. Note that while the reported locations in the empirical GSM data belong to the antenna that receives the cell phone signal, in this example, we assume that these records are the user's exact locations. The spatial changes and errors could be investigated in separate research—here, we focus on temporal effects only.

Therefore, the person has been observed in five locations (Figure 2.1(b)) named $\{A, B, C, D, E\}$ with a constant PI of fifteen minutes. However, the actual travel diary implies that only in two of the reported locations (*A* and *D*) activities occur (*work* and *home*); thus, understanding whether the user is traveling or staying (engaging in an activity) in each record requires developing a separate procedure. To decide on the event type (travel or stay), one could for instance use each event's starting time and duration. Moreover, based on Figure 2.1(a), the user was observed in *A* from *15:35* to *16:20*, but the actual traces imply that the user stayed at the mentioned location from *15:23* to *16:15* to *work*; hence, due to time discretization, delay in observing the start and end of each event is inevitable. How PIs affect the resulted OD matrix (under different conditions) is part of this study.

The second—and related—consequence of time discretization is the discrepancy between observed and actual activity (or travel) *duration*. For instance, in Figure 2.1, the perceived duration of *A* is 45 minutes; whereas, the actual duration is 52 minutes. This duration discretization error ranges from *-PI* to *PI*. In fact, we may even lose a fraction of OD trips (i.e., observed with zero duration) because the data might not capture specific trips with duration less than the PI. For example, activities that last less than one minute could easily be missed from the GSM data with high PI. In many cases, detecting such short-time activities is not very useful from a travel demand perspective. However, if activities are longer (e.g., more than 15 minutes), it might be insightful to configure them. Consider for example three activity categories: *home*, *work*, and *other*, representing staying at home, working, and engaging in other types of activities, respectively. One could argue that for OD matrix estimation, the distinction between stay (on a specific location) and travel (between locations) is sufficient. *Home* and *work* are major activities from a traffic planner's perspective since they account for a large part of travel diaries. Additionally, they often have aggregated daily durations of a couple of hours, making them more likely to be captured, even with very coarse PIs. However, *other* encompass all activities made for less common purposes, such as shopping, socializing, and health. These normally last much shorter. The maximum PI (for having cell phone reception in case of no network connection) adopted by the telecommunication company is about two hours. Consequently, there are interactions between the mix of activity durations and PIs, whose effects on the reconstructed OD matrix are not fully understood. It is our aim to gain a better understanding of how the reconstructed OD matrix deteriorates by testing a range of duration-PI combinations (from one minute to two hours).

To this end, our approach is threefold:

- **Pre-processing and ground-truth analysis:** First, we generate synthetic GSM data directly from a detailed set of ground-truth travel diaries. Furthermore, to train our GSM interpretation method (for event-activity type detection), called **Kernel-based approach (KA)**, we select a random one percent sub-sample from the ground-truth travel diaries.

- **Developing and applying KA and OD matrix determination:** Second, we develop and validate the KA algorithm by which we reconstruct the travel diary of each person for determining the OD matrix from the interpreted GSM data.

- **Comparison of OD matrices:** Third, we compare the reconstructed and actual OD matrices derived from the interpreted GSM data and ground-truth, respectively, where we use multiple evaluation metrics.

This way, our analysis studies the mixed effect of the polling interval and temporal criterion. Our method adopts this interaction (derived from the training data) to discern activities from trips on the reconstructed OD matrix. Hence, our results show the causes that affect the accuracy and robustness of the estimated OD matrix using GSM traces.

In this research, the following contributions are made:

1. We propose and evaluate a data-driven method for interpreting GSM data, which does not rely on a-priori assumptions of activity-travel behavior, and hence is applicable for both synthetic-experimental analyses (as in this research) and empirical-practical implementation.

2. We show that even a small portion of the population could train our method for location-activity type detection of GSM records. This method could further be trained when more observations are available.

3. We provide an overview of the effect of the underlying PI on the resulting OD matrix. Our analysis results also imply that the shorter the activity duration, the less its possibility to be identified correctly.

4. We show when randomness in the OD matrices spike relatively to how frequently we poll the users.

The research is performed with, as a case study, the data of the Amsterdam region in The Netherlands. We assume to have GSM data of a given population (i.e., we do not deal with the second-part problem of scaling from GSM-sample towards full population) as well as a limited amount of travel dairies (i.e. 1% of the given population).

The remainder of this chapter is organized as follows: Section 2.2 explains fundamental characteristics of GSM data, that need to be accounted for. Section 2.3 describes the research data and the implemented method. In Section 2.4, we evaluate the proposed method, present the results of applying the method on the GSM data for reconstructing the OD matrix and comparing it with the ground-truth OD matrix. Finally, Section 2.5 concludes the chapter.

2.2 GSM data in general and as used in this study

Basically, three main types of GSM data are generated by telecommunication companies:

- The first type is **Call Detail Record (CDR)**, which constitutes a majority of GSM data in transport research (e.g., Calabrese et al. (2013); Chen et al. (2014)). It includes event-based history information on the communication of a specific device, which consists of calls, SMS (short message service), internet connections. CDRs consist of the timestamp, call duration, type of events (voice call, SMS, data), and the cell's code in which the event occurs. Consequently, recording phone positions is dependent on the users' communication behaviour. Therefore, we need to assume on how to generate synthetic CDR. For instance, a random communication rate derived from a Poisson distribution between a minimum and maximum rate can be assumed for each user. A more mature and complex scenario is using discrete choice theory, which is based on utility maximization, i.e, it couples agent's decisions to attributes of the alternatives and agent's environment.

- The second type is named **Signaling Data** which informs us of the location area (LA) of the mobile phone on a permanent basis. Nonetheless, its spatial resolution is much lower than CDR because each LA includes more than a hundred base stations (Bonnel et al., 2015). Therefore, this type of data does not seem suitable for demand estimation and activity analysis for transportation purposes.

- The third type which is called **Periodic Location Update (PLU)** contains anonymous user ID code, time of the day, and location coordinates. Unlike CDR, PLU does not involve mobile phone users for storing their records. In fact, the GSM operating system decides on when to collect all users' data. Additionally, the spatial resolution is the same as CDR. Moreover, the PI is constant among all users independently from their behaviors. Thus, the random errors of the data has been partially eradicated due to the fixed interval of records. As a result, activity locations would be detected efficiently and by shorter data collection time. However, compared to CDR, accessing such data is an arduous task. Nonetheless, some research has already used them in mobility demand estimation (e.g., Zhang et al. (2010)).

In this study, we used PLU instead of CDR because it does not rely on a-priori assumptions between GSM data events and activity-travel behavior (which may otherwise introduce behavioral biases if not adequately addressed). Indeed PLU hence is applicable for both synthetic-experimental analyses (as in this research) and empirical-practical implementation.

As all types of GSM data capture the movement of vehicles and people, they could be used in estimating the travel patterns. However, one needs to deal with the new challenges of developing, and validating of models adopted for estimating the OD matrix. In fact, despite great opportunity of using GSM traces for OD matrix estimation, several drawbacks cause obstacles when it comes to practice:

1. Mobile phone data only observes the user's presence at a certain point in time in a particular mobile phone cell. Whether the person was traveling then or attending an activity cannot be directly concluded (Zilske & Nagel, 2014). Therefore, one must interpret the GSM traces to reconstruct the travel patterns. A number of previous research has specified a certain duration (or speed) to make a distinction between *stay* and *pass-by* locations in the GSM records (e.g., (Iqbal et al., 2014; Alexander et al., 2015; Demissie et al., 2019)). For instance, (Iqbal et al., 2014), Alexander et al., and (Demissie et al., 2019) assumed that a trip is recorded if in the CDR, subsequent entries of the same user indicate location change with a time difference more than 10 minutes but less than 1 hour. Wang et al. (2020), on the other hand, assumed that if the duration between two consecutive records be more than the sum of assumed minimum activity duration (e.g., 2h) plus time needed to get from previous location plus time needed to get to the next location, the current location was identified as a *stay* location. In another study, (Bachir et al., 2019) grouped stay points according to a speed threshold $\Delta v < 10$ km/h and a duration threshold $\Delta t > 15$ minutes; hence, a device was stationary if the duration between the first and last stay points lasted several minutes, with a low speed. Records not fulfilling this condition were considered as *pass-by* points. However, all the mentioned studies, raise a question of how to select and validate the clear-cut duration or speed.

2. Studies that intend using GSM traces are hindered by privacy protection regulations. A conventional procedure obligates the researcher to use only the minimum of information needed for the study in the form of aggregated results that do not focus on individual phones (Becker et al., 2013). This kind of data is regularly achieved by decreasing time

resolution and increasing space granularity (e.g., Bianchi et al. (2016)); Hence, the available data are spatially coarse and temporally sparse (Becker et al., 2013; Burkhard et al., 2017; Chen et al., 2018).

3. Another challenge which only results from using CDR is that mobility analyses based on such data could be biased (Zhao et al., 2016) since recording phone positions is based merely on the users' communication activities, which are unevenly distributed in space and time.

This research adopts three strategies to avoid each of the indicated issues:

- **simulated test bed:** Using simulation in the initial research analysis phase has been vigorously promoted (Zilske & Nagel, 2014). This environment allowed us to set up a coherent synthetic testbed to evaluate the effects of various potential components in our models. This environment promises solutions to the first deficiency regarding real-world GSM traces since users' actual activity locations are available. Therefore, it is possible to verify our methods of interpreting GSM data.

- **synthetic instead of empirical data:** Synthetic data is the preferable solution for developing a new method and comparing its performance with various methods to initially decide which models and methods to use on real-world data. Using synthetic data, no privacy concern is involved. Moreover, to the best of our knowledge, such highly detailed empirical data is hardly available, at least in the Netherlands. Therefore, using the synthetic data is not only advised but also necessary.There is no doubt that the proposed methods' final evaluations and performance measurements have to be fulfilled using real data.

- **PLU instead of CDR:** The dependency of the demand estimation on user's communication activities could be dealt with using the second type of cell phone data, PLU (Zhang et al., 2010). Since PLU has constant PF for the whole participants, dependently from their behaviours, OD estimation's bias towards more active GSM users, specific periods, and areas would be partially eradicated.

2.3 Material and Method

2.3.1 Experimental framework

As discussed in the previous section, to use GSM data in mobility pattern detection, we need to deal with three issues (i.e. not reporting actual activity locations, privacy-related aggregation, and dependency of CDR on the user's GSM activities). Considering these three strategies mentioned in the previous section, figure 2.2 represents the overall experimental design of this research in six steps:

1. pre-processing and selecting the intended users; For example, due to computational reasons, we only considered users with the *car* mode. In this step, data cleaning, reduction, and transformation take place.

2. analysing the ground-truth which is twofold:

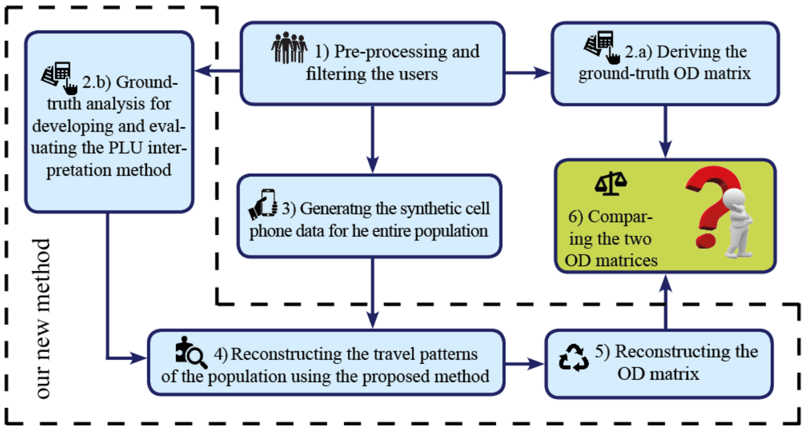


Figure 2.2: Overall experimental framework.

(a) aggregating the travel patterns of the entire population to form the ground-truth OD matrix;

(b) developing and evaluating our KA method in reconstructing users' traveling patterns from the PLU; Regarding method developing, we assumed to have PLU of all users. Then, travel diaries of one percent of the users trained our KA method. This prior knowledge specified the temporal routines of travels and activities, which allowed the Bayesian Classifier in our method to make distinction between *stay* and *pass-by*. The same way, the overall type of each activity (*home*, *work*, or *other*) could be detected.

3. generating the synthetic PLU directly from the ground-truth. This is done by applying the PF to the activity-travel patterns, in order to derive their locations at a given interval.

4. applying the KA method on the synthetic PLU to reconstruct the travel diaries,

5. determining the OD matrix from the reconstructed patterns,

6. comparing the actual and reconstructed OD matrices using several measures which are introduced in the next part.

2.3.2 Kernel-based approach of estimating the OD matrix

In this section, the features of the proposed KA method are discussed which mainly focus on the steps *2.b*, *4*, and *5* of Figure 2.2. In part 2.3.2 the fundamental concepts of Bayesian approach is discussed. The next section 2.3.2 explains Kernel density estimation which we used for extracting temporal routines from the training set. Section 2.3.2 explain the applied spatial aggregation on the synthetic GSM data. The comparison measures used in step *6* are also introduced in section 2.3.2.

As mentioned in section 2.1, PLU periodically observes each cell phone at certain places. However, figuring out whether the person engages in an activity or simply passes by needs further investigations. To identify location type (*stay* or *pass-by*) and activity category (*home* or *work* or *other*), we use a Bayesian classifier. Bayesian inference is regularly applied to estimate distribution parameters from data. In our research, a random one percent sub-sample from the ground-truth data is selected to train the Bayesian classifier. Furthermore, using Bayesian inference, it is possible to update conclusions based on this training set by incorporating new observations (Zhao et al., 2020).

As indicated in section 2.2, a number of studies in the literature have addressed the problem of location type identification in GSM data. Additionally, they have mostly used a time boundary to discern *stay* from *pass-by* (e.g.,Iqbal et al. (2014); Alexander et al. (2015); Demissie et al. (2019)), i.e., they impose a specific duration as a clear-cut distinction between *stay* and *pass-by*. However, in this research, the Bayesian classifier only learned from a training sub-set, randomly selected from the entire travel-activity patterns. In fact, the classifier infers the time boundary from the training set's temporal routines and applies it to the PLU to distinguish each user's stationary locations. Temporal routines allude to the distribution of duration and start time of records that are intended to be classified. The major merit of Bayesian inference is that the data in the training set are allowed to "speak for themselves" in determining location type; much more than in the case when the location type would be detected using pre-specified duration boundaries.

The primary logic behind selecting the duration and starting time of events as explanatory variables is that people's location and activity types are correlated with their temporal patterns.

For instance, trips in urban areas often have duration of less than an hour, whereas stays (i.e., activities) usually last for a couple of hours. Additionally, activity category of *home* mainly starts in the afternoons or early evenings and takes more than six hours. *Work* also largely takes more than five hours, however they normally start from the morning. Systematically considering these differences in the distributions enabled us to recognize each event or activity from others.

To understand how we measure the observed duration and start time, consider the example in Figure 2.1. The first record of each event (i.e *stay* or *pass-by*) is labeled as the start of it. Accordingly, the start of *A*, *B*, *C*, *D*, and *E* are *15:35*, *16:20*, *16:35*, *16:50*, and *17:50*, respectively. The first record of the next event is labeled as the end of each event, thereby the end times *A, B, C,* and *D* are *16:20, 16:35, 16:50,* and *17:35*, respectively. Since the duration is the difference between the end and start time, With a constant PI (here is 15 minutes), the durations would be multiples of PI.

Bayesian classifier

The proposed method, considering pre-specified temporal patterns, estimates the probability of *stay* and *pass-by* or activity categories. This approach is general in that it can be applied to diverse mobility patterns and data sources. We use the fact that people's location and activity types are correlated with duration and start time, i.e., events' classes are causes that affect their temporal patterns. Given the cause (event class), the duration and starting time are conditionally independent (see Russell & Norvig (2016)). Hence, the full joint distribution can be written as

$$P(Y, X_1, \ldots, X_n) = P(Y) \prod_i P(X_i|Y), \tag{2.1}$$

where Y is the event class and can be either location type, with two classes: *stay* and *pass-by*; or activity type, with three classes: *home*, *work*, and *other*. Furthermore, X_i is the effect and consist of two temporal variables: duration and starting time. Such a probability distribution is called a **naive Bayes** model — "naive" because it does not account for cases where the effect variables are not truly conditionally independent given the event class. Practically, the naive Bayes method can work surprisingly well, even when the conditional independence assumption is violated (Russell & Norvig, 2016).

The Naive Bayes model is sometimes called a Bayesian classifier since often **Maximum a-posteriori (MAP)** estimation concept is used for classification (Demirbas, 1988), i.e., Bayesian classifier assigns the most probable class $\hat{y}$ to the observed data X. Defining $P(y|X)$ as the probability of class y given that $X = x_1, \ldots, x_n$ was observed, the Bayesian classifier evaluates the following maximization scheme (see Yair & Gersho (1990)):

$$\hat{y} = \underset{y \in \{1,\ldots,Y\}}{\operatorname{argmax}} \{P(y|X)\} = \underset{y \in \{1,\ldots,Y\}}{\operatorname{argmax}} P(y) \prod_{i=1}^{n} P(x_i|y). \tag{2.2}$$

The quantities $P(y|x)$ are known as the a-posteriori(or class) probabilities, and the Bayesian classifier supplies the MAP decision.

Kernel density estimation

The assessment of the a-posteriori probabilities using Bayes Rule requires an **a-priori** knowledge about the probability density functions of the priors. The probability density function of the priors, which are duration and start time of events (activities and trips), needs to be estimated. Assuming that the observed data points in the training set are a sample from an unknown probability density function, **density estimation** is the construction of an estimate of the density function from the training set. In this regard, **kernel density estimation** (KDE) is currently the most popular **non-parametric** approach for probability density estimation (Scott, 2012; Simonoff, 2012). Non-parametric density estimation is an alternative to the parametric approach in which a model with a small number of parameters is specified and, using maximum likelihood, the model is calibrated. However, non-parametric density estimator is aimed to estimate the density of a variable from a sample set without assuming any specific form for the density function (Silverman, 1986; Wand & Jones, 1994; Simonoff, 2012; Scott, 2012). As we generally happen to know very little about the given data, a general smoothness assumption is a reasonable choice. Accordingly, we selected the **Gaussian** kernel estimation (see Smola et al. (1998)). Hence, our proposed method is named **Kernel-based approach (KA)**.

Given N independent observations $\chi_N = X_1, \ldots, X_N$ from an unknown continuous probability density function f on X , the Gaussian kernel density estimator is defined as

$$\hat{f}_h(x) = \frac{1}{N} \sum_{i=1}^{N} \phi(x, X_i; h) \forall x \in \mathbb{R}, \tag{2.3}$$

where:

$$\phi(x, X_i; h) = \frac{1}{\sqrt{2\pi h}} \exp^{\frac{-(x-X_i)^2}{2h}}$$

is a Gaussian kernel with location Xi and bandwidth $\sqrt{h}$. Many researchers have focused on the optimal choice of h, since Bandwidth selection greatly affects the estimate obtained from the kernel density estimation (much more than the shape of the kernel) (e.g., Jones et al. (1992); Sheather & Jones (1991); Botev et al. (2010)). In this research, bandwidth selection was made by a "rule of thumb" following Scott's Rule (see Scott (2015)), which suggests that the optimal bandwidth is $n^{\frac{-1}{d+4}}$ in which n is the number of data points and d is the number of dimensions.

Spatial Aggregation of GSM Data

As already mentioned in section 2.1, the reported locations in the empirical GSM data generally belong to the antenna that receives the cell phone signal. This naturally means that the highest spatial resolution of the data is antennas' coverage areas. Therefore, to demonstrate such spatial aggregation effect, we used the associated OD zones instead of the exact location coordinates. This means that each OD zone represent one antenna. Also when users only travel inside a zone, their movements are not observed because it is like the person has stayed in one location representing that OD zone (as shown in figure 2.3) . Under this setting, the hypothetical antennas for each zone are shown in the figure.

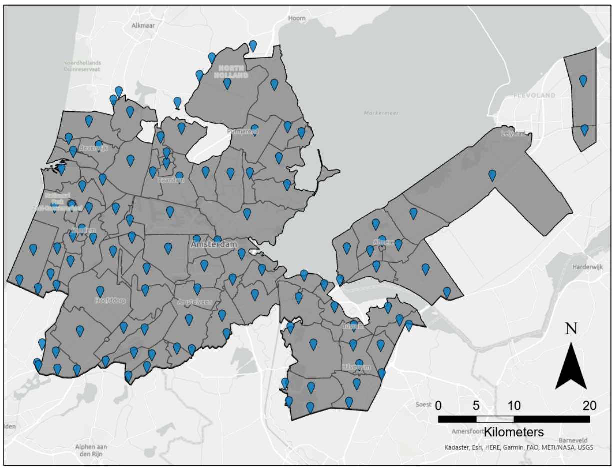


Figure 2.3: OD zones assumed to represent the antennas' coverage areas.

In addition to the *duration* and *start* time, the location of each activity provides information about the activity category. A hierarchical model is applied to the locations (i.e., OD zones) in the training set to estimate the a-priori for each spatial zone in the Bayesian model. This hierarchical model estimates the probability of each activity category in each OD zone. The activity category is drawn from a categorical distribution with the parameter specific to each zone. This parameter is drawn from a Dirichlet distribution with parameter α assumed to be unique among all the zones (we assumed it be a vector of one which is equivalent to a uniform distribution).

$$\theta_{d=1\ldots M} \sim Dirichlet_K(\alpha) \tag{2.4}$$

$$z_{d=1\ldots M,n=1\ldots N_d} \sim Categorical_K(\theta_d) \tag{2.5}$$

where:

z = the activity category of the observation,

α = the hyper-parameter vector of the Dirichlet prior,

M = the number of zones,

N = the number of records, and

K = the number of assumed activity categories (three here).

Section 2.4.3 investigates whether considering the spatial variable under the mentioned setting will improve the model accuracy for our data set.

OD matrix comparison

To evaluate the performance of the proposed method for OD matrix estimation, as well as evaluate the effect of polling frequency hereon, a metric is needed that measures the accuracy of the estimated OD matrix against the ground-truth OD matrix. OD matrices can be compared in two complementary ways. Firstly, the degree to which the estimated OD matrix correctly represents the absolute amount of demand between all individual OD pairs. Secondly, the degree to which the estimated OD matrix correctly represents the relative demand pattern seen across OD pairs (Behara et al., 2020a).

For the former absolute comparison, traditional measures such as mean absolute error (MAE) (see Ashok & Ben-Akiva (2002); Lo & Chan (2003)), and R-squared (see Tavassoli et al. (2016)) can be used. Here we use MAE to compare the OD pair values in the two matrices.

For the latter relative comparison, the geographical window-based structural similarity index (GSSI) (Behara et al., 2020b) is capable of distinguishing structural differences due to the geographical closeness of OD zones. Here we use GSSI to compare the correlation of OD pair values across geographical windows (where OD zones with geographical proximity belong to the same window).

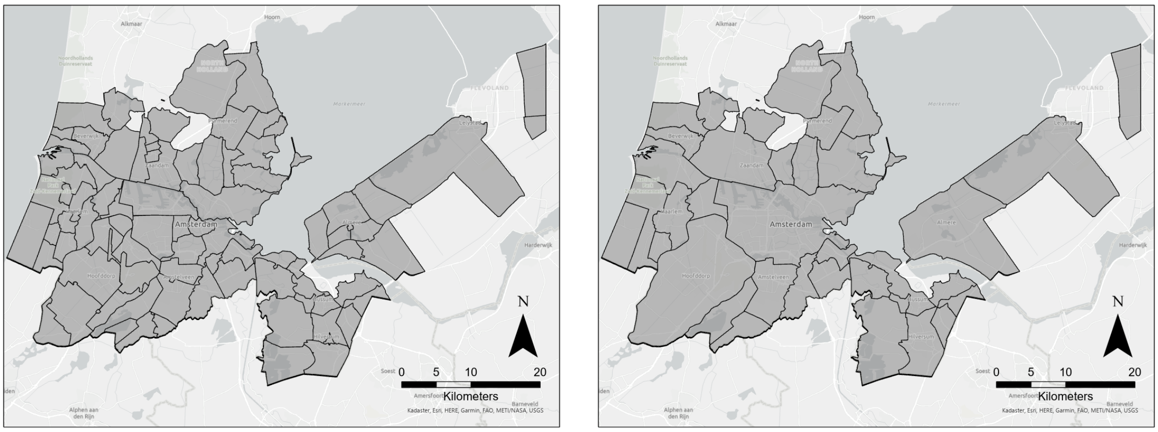
2.3.3 Method implementation

The experiments are done on daily activity plans of agents derived from the nationwide activity-based model ALBATROSS (Arentze & Timmermans, 2004). All agents are selected for which at least one household member use the mode car to perform at least one activity within the Amsterdam region (Winter et al., 2020). This leads to the activity plans of 22 thousand agents during a representative working day.

Note that we use synthetic (model-generated) activity plans in the KA step, our method estimates temporal patterns of location and activity types. Therefore, to ensure generalisability of our method towards empirical travel diary data, it is important to emphasize that the ALBATROSS model does not assume any a-priori (theoretical) distribution of activities, but instead uses decision trees that are directly calibrated from travel diaries. ALBATROSS implements a sequential decision-making process to generate an individual's schedule. Empirical demand data are employed to induce a decision tree for each step in the scheduling process (Timmermans & Arentze, 2011). Thus, the model framework (i.e., decision tree) does not restrict the activity's temporal pattern to a specific distribution. Decision trees can describe discontinuous impacts of discrete attribute variables on decision making. Hence the fact that our KA step can capture the temporal distribution of location and activity types is based on the underlying empirical travel diaries. This is important, because otherwise in case the synthetic data would be based on a model that adopts a (parameterized) distribution function to simulate activity patterns, then the temporal distributions might have been artificially imposed by the model structure (i.e., a model assumption, not a model result).

The agents' activity plans are simulated using **MATSim** (Zilske & Nagel, 2015), which is an open-source agent-based transport simulation model. The MATSim model output is used to generate synthetic PLU data as follows.

The **experienced plan** output contains basically the traces of each agent in our data set. This file represents the ground-truth OD matrix. One percent of these agent traces is sampled to be used as travel diaries in the KA step of our method (to estimate the Bayes model distinguishing *stay* and *pass-by* locations, and detecting the activity category).



(a) Amsterdam OD zones (b) High-level boundaries as geographical windows

Figure 2.4: Zone boundaries applied to this study.

The **snapshot** output contains the records of all agents per snapshot interval, which is conveniently used (with slight modifications regarding formatting) to generate our synthetic PLU traces. The resulting PLU format is shown in Figure 2.1(a) in which the associated OD zone represent the location.

The OD zoning system is an aggregated version of 4-digit postal codes in Amsterdam leading to 115 zones (Figure 2.4(a)). When computing the GSSI metric (i.e., structural similarity between OD matrices) these are aggregated to 50 geographical windows as displayed in Figure 2.4(b).

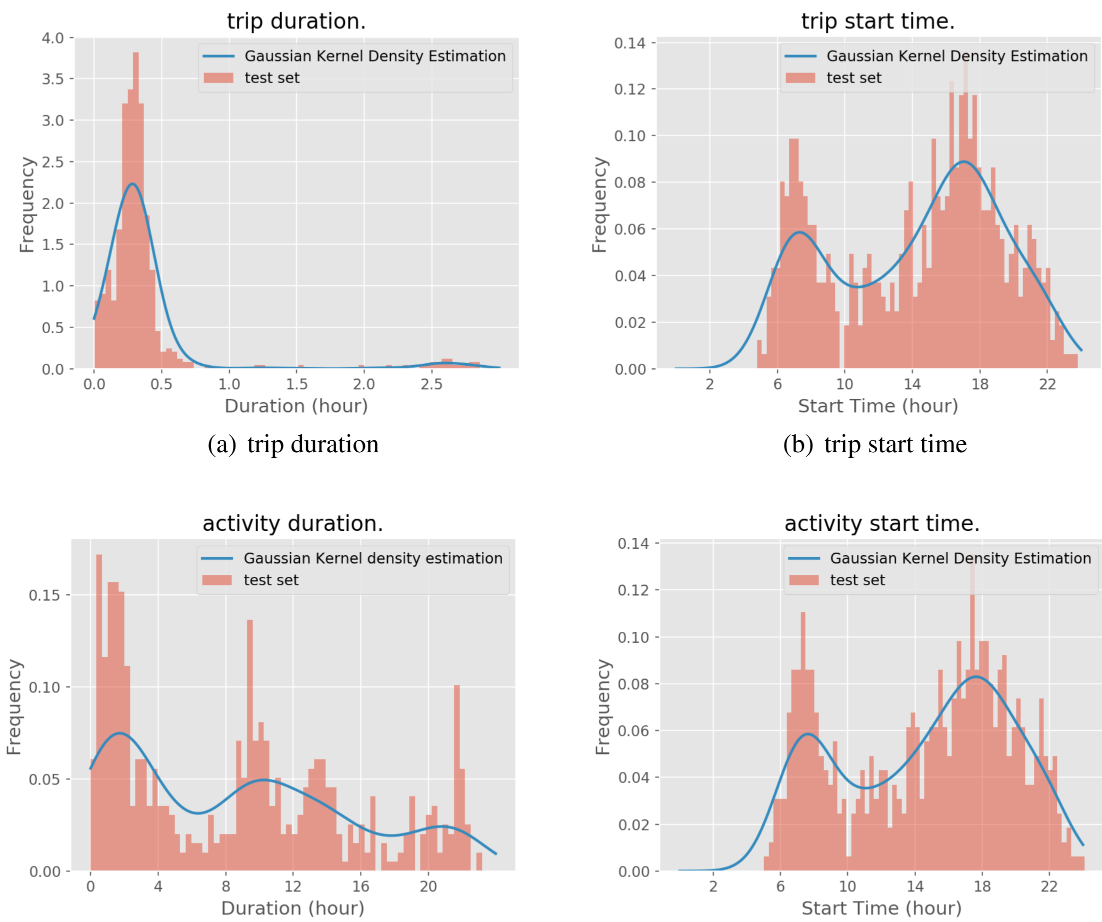
2.4 Results and Discussion

In the following, we present the KDE results (part 2.4.1) and evaluate the KA performance in detecting location types (part 2.4.2) and in detecting activity types(part 2.4.3). This is done for a random training set. Part 2.4.4 then shows how robust these results are against selecting the training set. Finally, part 2.4.5 shows the accuracy of the reconstructed OD matrices for a range of PIs.

2.4.1 KDE results

As mentioned in the previous part, the proposed methodology is trained on a one percent subsample of ground-truth, and its performance is tested on the entire ground-truth data set.

An example result of applying KDE on the training set (based on one example random seed for selecting the training set) is in figures 2.5 and 2.6 for location and activity category detection, respectively. For each location/activity type, two distribution of duration and starting time is calculated. Having assumed to have two location types of *stay* and *pass-by* and three activity categories of *home*, *work*, and *other*, ten distributions are fitted to the training data. The KDE simply fits a smooth curve to the data and introduces the likelihood required for the Bayesian classifier. As we usually happen to know very little (in our case 1% of the population) about the ground-truth, a smoothness assumption for the training set's density estimation is justifiable.The smoothness assumption prevents over-fitting caused by sparse sampling when the training data, like in our framework, comes from a small proportion of the population. It is worth mentioning that the smoothing assumption can be relaxed in case the training data represents the entire population more thoroughly.



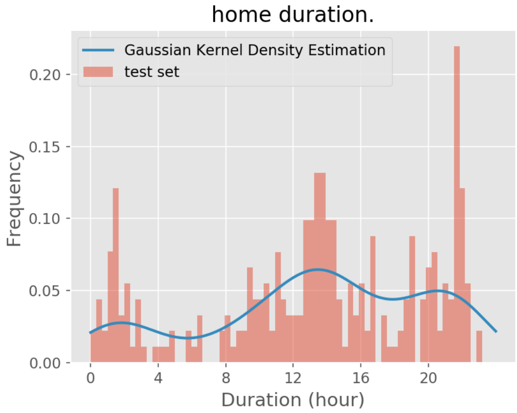
(a) trip duration (b) trip start time

(c) activity duration (d) activity start time

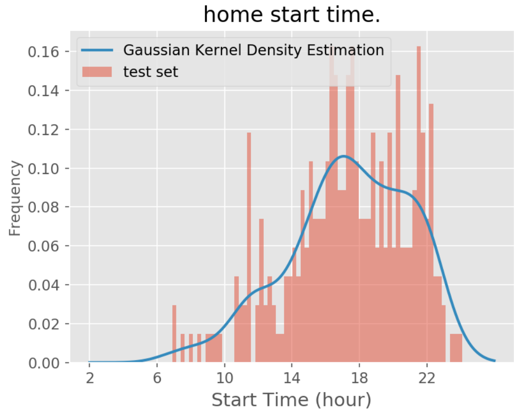
Figure 2.5: Application of KDE for location type detection.

It can be inferred from Figure 2.5(b) and 2.5(d) that the start time pattern of trips and activities follow a close distribution. Consequently, the KA distinguishes the event types merely based on the duration distributions which follow different pattern in each event type. As a result when duration patterns of activity and trip are similar, miss-prediction of location type takes place. Considering Figure 2.5(a) and 2.5(c), miss-predictions might happen when duration is less than 45 minutes.

For such cases, the location of the record plus temporal features might give a more precise prediction of the event and activity category. However, this is only true if the spatial resolution of the data is high enough to recognize different land-use types from each other. In fact, in dense urban environments like Amsterdam, assuming high spatial resolution for GSM data is not realistic. Because various events and activity categories take place in a small area and to estimate a reliable probability of each, the GSM antennas have to be unnecessarily closer to each other. Perhaps for sparsely developed urban environments, using spatial variables in the KA model become reasonable.

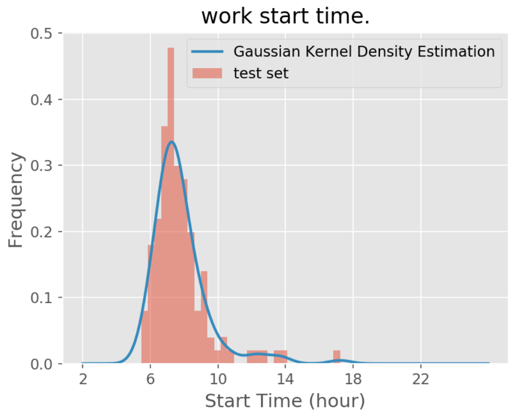


(a) *home* duration

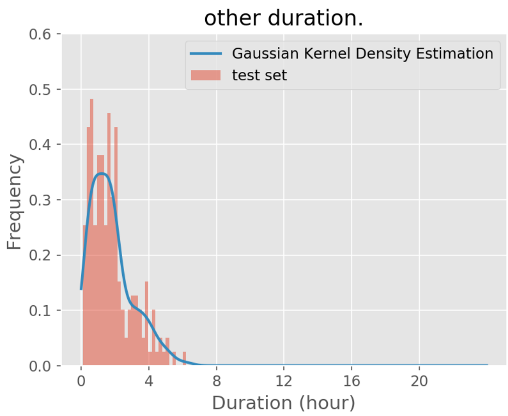


(b) *home* start time

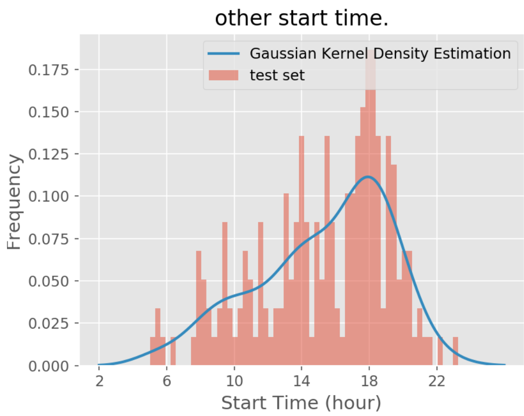
(c) *work* duration



(d) *work* start time



(e) *other* duration



(f) *other* start time

Figure 2.6: Application of KDE for activity type detection.

2.4.2 KA performance regarding location type

The performance of the KA classifier for location type detection is shown by the confusion matrix in Table 2.1. The results show that 92% of all *stay* and 97% of all *pass-by* locations were detected correctly, using the proposed methodology (overall, in 94.4% of the time, the location type was distinguished correctly). Based on Table 2.1, KA underestimated the *stay* locations. Accordingly, overestimation in *pass-by* locations occur at the same time. Therefore, in *stay* detection, we often had false-negative errors, and in *pass-by* detection, false-positive errors happened the most.

Table 2.1: Confusion matrix of applying the methodology on the entire ground-truth for location type recognition.

	observed *stay*	observed *pass-by*	total predicted
predicted *stay*	54231 (91.8%)	1799(3%)	56030(47.4%)
predicted *pass-by*	4822(8.2%)	57254(97%)	62076(52.6%)
total observed	59053(100%)	59053(100%)	118106(100%)

As already mentioned in section 2.4.1, due to closeness of *stay* and *pass-by* starts, duration of events has a more significant role in differentiating them. Observing the results showed that the minimum duration of correctly recognized *stay*s was about 44 minutes. Therefore, it seems that KA specifies a duration threshold to separate *stay*s from *pass-by*s. Although this threshold is selected by analysing the temporal distributions in the training set, it cannot entirely divide *stay*s and *pass-by*s due to overlaps around the threshold. In fact, our analysis shows that about 88% of activities have duration more than 44 minutes, whereas, 96% of trips endure less than 44 minutes. Thus, activities are less probable to be recognized. This justifies the underestimation in activity detection in Table 2.1.

2.4.3 Variable selection and model performance in activity category inference

In our model we included predictors that help maximize the activity category accuracy.To select the model's explanatory variables, we considered three states: First, only spatial variable; second, spatial and temporal variables together; third, only temporal variables.The associated confusion matrices are shown in Tables 2.3, 2.4, and 2.5. The overall performance of the classifier for activity category detection for each of these states is in Table 2.2. The **precision score** is a ratio that shows the quality of positive predictions made by the model and is defined by the number of true positives divided by the total number of positive predictions so that 100% precision implies no false positives. However, this typically coincides with a lower **recall score**, which is defined as the number of true positives divided by the sum of true positives and false negatives, so that 100% recall implies no false negatives. In an imbalanced classification, the **balanced accuracy** is the average of the recall score obtained in each class.

The results (Table 2.2) show that considering only OD zones leads to accuracy scores as low as 36.6%. Also, based on table 2.3, this model seems to be biased towards detecting *home* as more than 80% of other categories are incorrectly labeled as *home*. Figure 2.7 also shows that in most OD zones the probability of *home* is more than the other two.

Adding temporal variables in table 2.2 increases the overall accuracy to around 90%. However, due to the presence of the spatial variable, the results are slightly biased towards *home* (Table 2.4). Table 2.5 confirms this because as the accuracy of *home* detection is reduced, the accuracy of *work* and *other* increase. Overall, the highest accuracy is achieved by considering only temporal variables in Table 2.2. Therefore, using location does not improve the

Table 2.2: Performance metrics of the three states used for activity type recognition.

state:	only spatial variable	both spatial and temporal variables	only temporal variables
precision score	0.4211	0.9019	0.9058
recall score	0.4792	0.8931	0.8969
balanced accuracy	0.3659	0.9086	0.9144

Table 2.3: Confusion matrix of applying the method using only spatial variable on the entire ground-truth for activity category recognition.

	observed *home*	observed *work*	observed *other*	total predicted
predicted *home*	25434 (92.8%)	10952(80%)	15242(84.8%)	51628(87.4%)
predicted *work*	488(1.8%)	585(4.3%)	447(2.5%)	1520(2.6%)
predicted *other*	1476(5.4%)	2151(15.7%)	2278 (12.7%)	5905(10%)
total observed	27398(100%)	13688(100%)	17967(100%)	59053(100%)

Table 2.4: Confusion matrix of applying the method using both spatial and temporal variables on the entire ground-truth for activity category recognition.

	observed *home*	observed *work*	observed *other*	total predicted
predicted *home*	22659(82.7%)	111(0.8%)	568(3.2%)	23338(39.5%)
predicted *work*	223(0.8%)	12899(94.2%)	217(1.2%)	13339(22.6%)
predicted *other*	4516(16.5%)	678(5%)	17182 (95.6%)	22376(37.9%)
total observed	27398(100%)	13688(100%)	17967(100%)	59053(100%)

Table 2.5: Confusion matrix of applying the method using only temporal variables on the entire ground-truth for activity category recognition.

	observed *home*	observed *work*	observed *other*	total predicted
predicted *home*	22458(82%)	52(0.4%)	175(1%)	22685(38.4%)
predicted *work*	289(1.1%)	12963(94.7%)	248(1.4%)	13500(22.9%)
predicted *other*	4651(17%)	673(4.9%)	17544(97.6%)	22868(38.7%)
total observed	27398(100%)	13688(100%)	17967(100%)	59053(100%)

overall activity category detection under the considered spatial level of aggregation, and using only temporal variables suffices. If detecting *home* is of a higher priority, considering location alongside temporal variables becomes a preference.

The results in tables 2.5 and 2.4 also roughly show that false-negative errors occur mostly for *home* category. Moreover, most of the false-positives are revealed for *other*. Focusing on

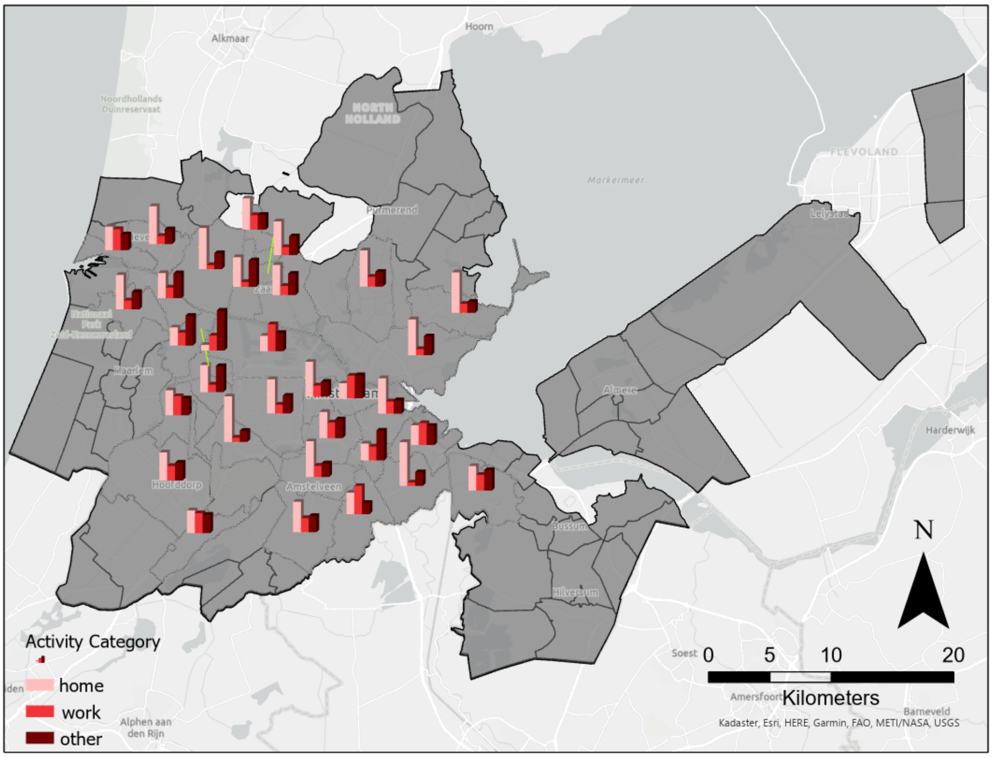


Figure 2.7: Average probability of each activity category based only on location OD zone.

when the method fails to detect the right activity type, Figure 2.8 presents the distribution of false-negative error in predictions over duration for *home*. Since the long duration of *home* discriminate this type of activity from other categories, False Negatives mostly occurs when it comes to shorter duration (less than 5 hours). Figure 2.9 shows the actual duration distribution of *other*. Similarity of the distributions in Figure 2.8 and 2.9 shed light on why the methodology might confuse *home* with *other*. Moreover, Figure 2.10 presents the distribution of false negative error in predictions over start time for *home*. Figure 2.11 displays the actual start time distribution of *other*. The afternoon peak in both figures give rise to confusion of *home* with *other* activities. Thus, activity type is estimated based on the start time and duration of stay, but under specific conditions such as short activity duration in the early evening the temporal distributions are non conclusive i.e., due to the sporadic closeness of the temporal pattern of

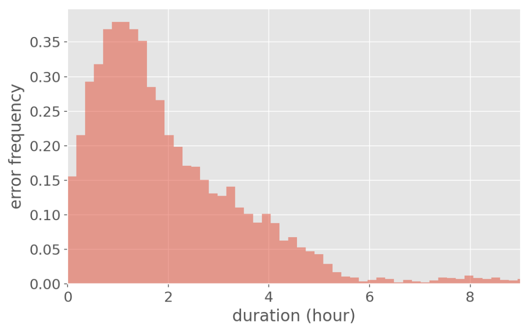


Figure 2.8: False negatives based on the duration of home.

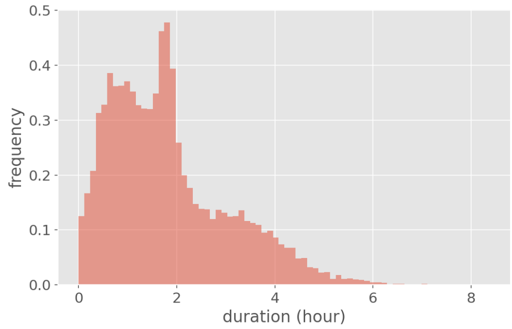


Figure 2.9: Actual duration distribution of other activities.

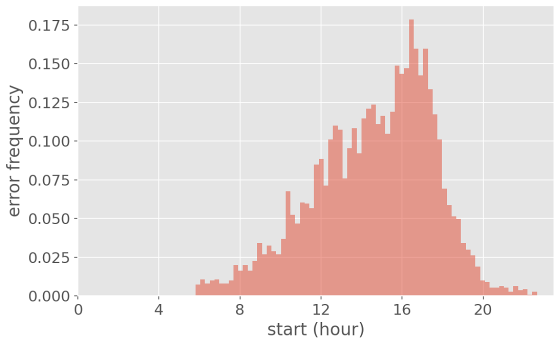


Figure 2.10: False negatives based on the start time of home.

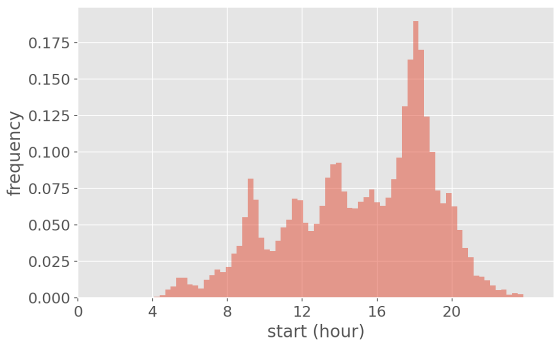


Figure 2.11: Actual start time distribution of other.

home and *other*, miss-prediction may occur.

2.4.4 Sensitivity of KA results to training set sampling

To evaluate the sensitivity of the randomly sampled training set (seed and size), the analyses in the previous sections were repeated using 50 different seeds. Figure 2.12 presents the KA performance (i.e. accuracy) in location-activity type detection across these 50 seeds.

The KA performance appears robust for different random seeds. This is further tested statistically using a Chi-square test on two random sub-samples each containing the performance for 25 different seeds (i.e. training sets). With a significance level of 0.05, the null-hypothesis

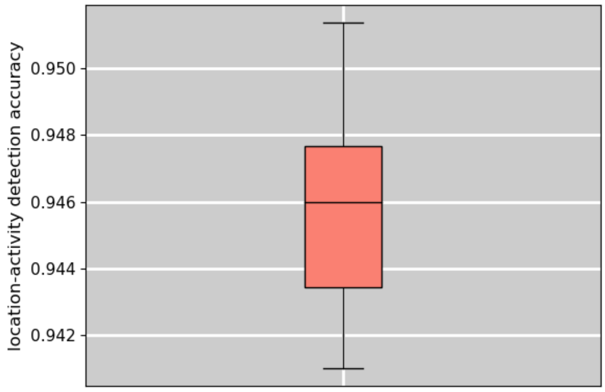


Figure 2.12: KA performance in location-activity detection on 50 different random seeds for selecting the training set.

that the two sub-samples belong to the same distribution could not be rejected (i.e. we cannot conclude that the KA performance derived from different sub-samples of random training sets would yield different distributions).

In the previous sections we evaluated the KA performance for a single random training set. The results of the Chi-square test indicate that a sample size of 25 is robust to evaluate the KA performance. Hence in the following section the accuracy of the reconstructed OD matrix is evaluated based on 25 experiment runs (i.e. across 25 different random seeds for training set sampling).

2.4.5 OD matrices comparison results

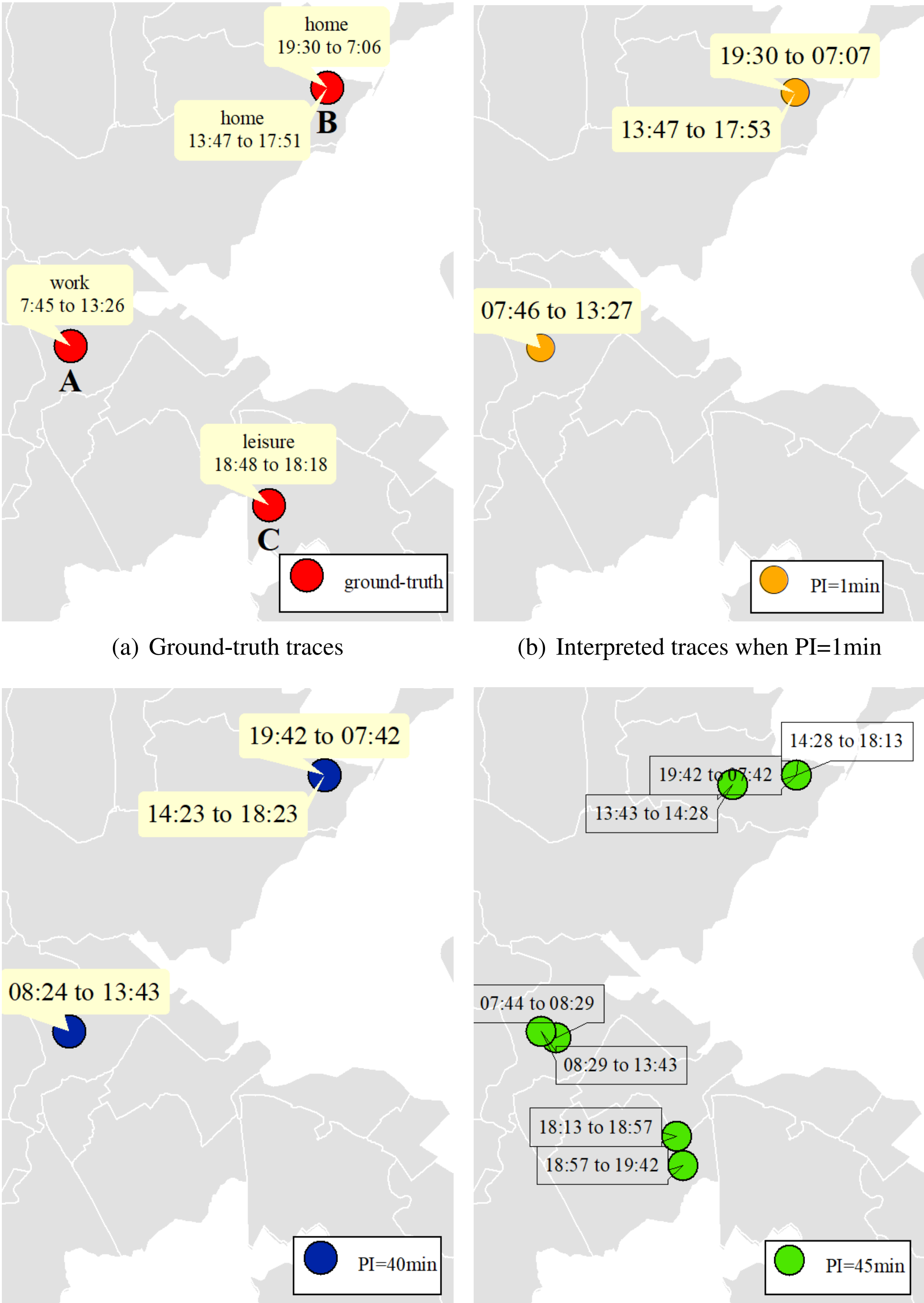
This part presents the results of applying KA on the generated PLU from the ground-truth using MATSim. We considered 18 various PIs in generating the PLU, which are 30 and 60 seconds, every 5 minutes from 5 to 60 minutes, and every 15 minutes from 1 to 2 hours. Since non-linearity and variation of results are high for PIs less than two hours, these interval were selected Moreover, for more clarification on the correlation of randomness of the results and the underlying PI, all results were calculated for 25 different random seeds (for selecting the training data from the ground-truth). Having derived the OD matrices for the morning peak (6:30 to 9:30) for both PLU and the ground-truth, we compared the actual and reconstructed outcomes using two performance indicators:

1. MAE between OD pairs of reconstructed and ground-truth OD matrices, which compares the OD pairs values, and

2. GSSI (for further information refer to Behara et al. (2020b)), which captures the structural similarity between the two matrices.

Figure 2.14 describes the MAEs resulted from comparing the ground-truth OD cells (as the observed values) and the reconstructed OD cells (as the predicted values) over a range of PIs. It is reasonable to see that the average value of MAE gradually increases by reducing the PF.

On the other hand, drops and jumps in the MAE values (from 45 to 120 minutes) might be due to the particular timing pattern of travels and activities (i.e. the context of the data result in such changes). In fact, it seems that the reliability of the method drops after *PI=45 min*.



(a) Ground-truth traces (b) Interpreted traces when PI=1min

(c) Interpreted traces when PI=40min (d) Interpreted traces when PI=45min

Figure 2.13: Example traces of a user.

To clarify the reliability fall, Figure 2.13 shows an example of a user's traces. Assuming that the letters show the O-D zone, the user's actual traces, in Figure 2.13(a), are $\{A,B,C,B\}$, in chronological order. However, the traces interpreted from the PLU with PI of 1 minute (Figure 2.13(b)) are $\{A,B,B\}$ which lacks the detection of C. This occur due to duration of staying in C (30min) which was less than the duration threshold (about 45min); thus, the record was interpreted as *pass-by*. Likewise, when *PI=40 min* (Figure 2.13(c)), the traces are $\{A,B,B\}$, but due to another reason-we lost activity *C* due to stay duration less than the PI value. Since travel demand derives from people's needs and desires to participate in activities, there could be a condition to compensate for such errors: When the interpreted travel pattern include similar (close in location coordinates) successive activities, a missed activity is assumed to be in between. It is located in the farthest record location between the two similar activities. The start time and duration could be derived using speed data and the distances between the locations.

Important deviations are seen in Figure 2.13(d). Since *PI=45 min* exceeded the duration threshold (44min), any record was considered as *stay* and the method's job was limited to only cluster the similar records. Therefore, significant number of imaginary *stay*s (i.e. activities) were generated; hence, the traces in Figure 2.13(d) were $\{A,A,B,B,C,C,B\}$ and instead of having four activities, we detected seven. The major issue here is that the travel time between all these activities is zero which is not feasible. Moreover, detecting and alleviating these massive errors is complicated due to lack of information. Consequently, it is suggested not to consider users or scenarios with PI values more than the duration threshold. Because the duration boundary is

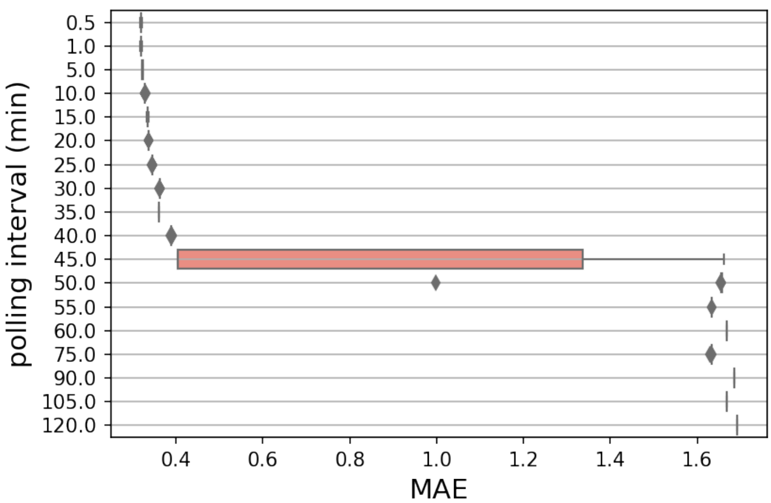


Figure 2.14: MAEs related to OD matrices resulted from 25 different random seeds over 18 different PIs.

a critical value, after which a sudden change in the performance of the KA for OD estimation happens. A little while (about 10 minutes) after this critical value, the variance seems to drop to values even less than before, which shows the more stability of the results against different random seeds underlying the training set. This phenomenon happens because in PIs more than 50 minutes,the KA cannot detect short-time activities anymore. In other words, the remaining activities are of long durations. For instance, *home* is more robust against random seeds since its duration is much higher than the PIs discussed here.

As another performance measure indicator, comparing the structure of OD matrices, we derived the GSSI for each of the PIs with 25 random seeds (Figure 2.15). Basically, GSSI is in the range of [0,1] and the higher GSSI indicate higher structural similarity of matrices. therefore, the gradually descending trend is perceivable (due to the inability to detect short time activities with durations less than PIs) before the duration threshold, but high variation is also noticeable. Regarding this, a higher PI increases the probability of missing the activity, even when the PI is still below the activity duration. In fact, higher PIs increase the delay

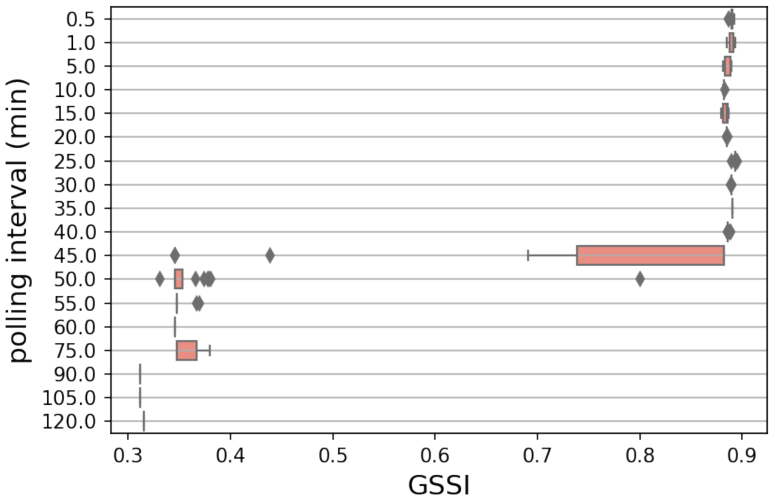


Figure 2.15: GSSI related to OD matrices resulted from 25 different random seeds over 18 different PIs.

range of detecting the start and end of activities. This increase the range of interpreted duration consequently variation jumps up and miss-identification is more likely to happen. Apart from the indicated rationale, errors due to the data characteristics might also produce variation in the results.

To understand the changes in Figure 2.15, note that the interpreted duration is a multiple of PI value. Furthermore, as mentioned in part 2.4.2, about 88% of activities have duration more than 44 minutes (the duration threshold), whereas, 96% of trips endure less than 44 minutes. Thus, activities are more probable to be miss-identified (i.e. *stay*s have more false negatives than *pass-by*s). However, under a special condition, activity's false negatives decrease.

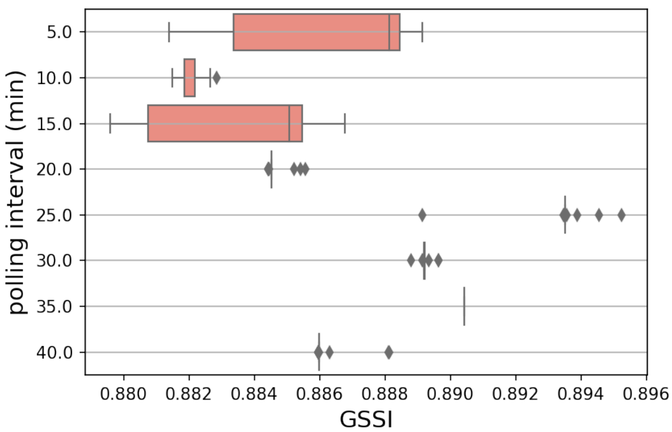


Figure 2.16: GSSI related to OD matrices resulted from 25 different random seeds over 8 different PIs.

To take a closer look at the results in figure 2.15, figure 2.16 shows the GSSI only for PIs less than the duration threshold. Since the input of KA is the interpreted duration, identification of a *stay* requires to have an interpreted duration more than the duration threshold (45 min) even if the actual duration is less than the threshold. For instance, when *PI=10min* and *interpreted duration=50min*, it is probable that the record be considered as *stay* even if its duration be in the range of (40,45) minutes. We call this range, the fortunate range (FR), which We define only to explain the results in figure 2.16. The length of FR causes the changes in GSSI value for PIs less than the threshold. The FRs for our other PIs are shown in Table 2.6. Notice that once the interpreted duration reaches the duration threshold, it is considered as *stay*. The longest FR

Table 2.6: FRs for PI = (5, 10, 15, 20, 25, 30, 35, 40).

PI (min)	interpreted duration (min)	FR (min)
5	45	(40, 45)
10	50	(40, 45)
15	45	(30, 45)
20	60	(40, 45)
25	50	(25, 45)
30	60	(30, 45)
35	70	(35, 45)
40	80	(40, 45)

belong to *PI=25min*. The same PI got the highest GSSI in Figure 2.16.

On the other hand, when the interpreted duration falls near the duration threshold, the variation spikes. Accordingly, *PI= 5,15 min* have high variances. In fact, this variation is higher for *PI=15 min* due to longer FR-more stochasticity. Naturally, between three PIs of 5, 10, 20, and 40 min that have the similar FRs, the lower PI gets the higher GSSI. PI values more than the duration threshold are not discussed due to poor performance of KA and a large number of unreal generated activities.

Overall, it seems that when false negatives of *stay*s are more than that of *pass-by*s, the longest FR with a minimum PI result in the highest GSSI; the most structurally similar reconstructed OD matrix to the ground-truth matrix. In our case, with duration threshold of about 45 minutes, $PI = 25min$ yields the highest GSSI. Basically, under the mentioned circumstances, with duration threshold of T, the ideal PI is $T/2+\varepsilon$ where ε is a very small value.

2.4.6 Research limitations

This research had several strengths: It certainly adds to our understanding of the effects of temporal characteristics of PLU data on the accuracy and robustness of the resulting OD matrix. It also proposes a data-driven method for interpreting the raw PLU data. Nonetheless, these findings must be interpreted with caution, and a number of limitations should be borne in mind as follows:

- **Limited synthetic data:** Real-world mobile phone datasets can offer ample information about millions of mobile phone users. Nevertheless, their limitation concerns the privacy issue and the difficulty of carrying out reliability and validation experiments (due to the unavailability of the ground-truth). In order to conduct operational tests and evaluations, we used synthetic (travel diary and GSM) data, thus including the ground-truth by design and without any privacy concerns. However, the question is whether our findings are generalizable toward empirical data. It is worth mentioning that the ALBATROSS model (based on which our data is generated) does not assume any a-priori (theoretical) distribution of activities but instead uses decision trees that are directly calibrated from travel diaries. Therefore, the model structure has not artificially imposed temporal distributions. However, a sampling bias in the original travel diary used to train ALBATROSS might be affecting the data's representativeness in describing the entire population.

On the other hand, the Bayesian model is naturally resistant to non-informative predictors. Nevertheless, due to the naive nature of our model, incorporating the location in addition to temporal variables in activity category detection slightly reduced the overall accuracy. This is because, in Naive Bayesian models, different prior variables are assumed to be conditionally independent. This assumption can be avoided by establishing the correlation of prior variables, which usually require a more extensive data set than the data available in this study. Therefore, we acknowledge a need for a future study (e.g., a longitudinal study) to consider a fully Bayesian model to comprehensively describe the relationship between the explanatory variables.

- **Mismatch between TAZs and base station coverage zones:** This research assumes that the base station coverage area is the same as its associated TAZ. However, there is a mismatch between the antenna's coverage zone and TAZ in practice. But we ignored this mismatch because the base station coverage area is typically much smaller than a TAZ. In fact, in the urban areas, the size of a typical TAZ is about 2-5 km, and that of a base station zone is about 50-200 m (Dong et al., 2015). To investigate further, Figure 2.17 shows the Cumulative density function (CDF) of TAZ sizes. Accordingly, the average TAZ size is about 15*km* which naturally contains tens of base stations.

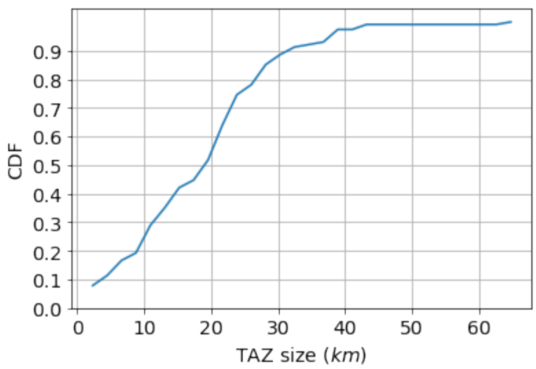


Figure 2.17: Cumulative density function (CDF) of the TAZ sizes in our study.

Therefore, the influence of the size of base station areas on the quality of TAZ division is not significant in urbanized areas (as is Amsterdam). In other areas and to generalize the analysis, one needs base station distribution and location data to account for the errors in the traffic activity analysis. It is worth pointing out that the TAZs seem too large to capture short-distance trips. Figure 2.18 shows the CDF of the trip distances in our study. Accordingly, more than 10% of the trip distances are less than 3*km* despite considering only *car* mode. Undoubtedly, adding active modes of transport to the analysis makes this proportion much higher. Despite the large extent, coarse-sized TAZs used for transport planning (as used here) will inevitably eliminate such short-distance trips.

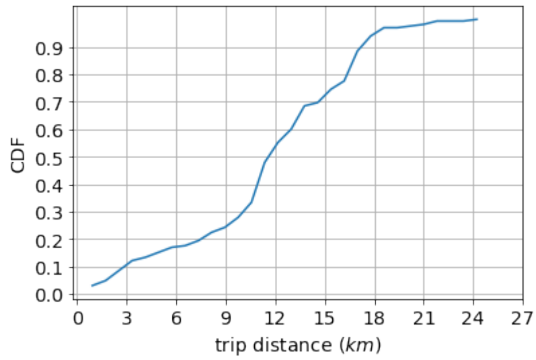


Figure 2.18: Cumulative density function (CDF) of the trip distances in our study.

- **Positioning accuracy and disturbance (ping-pong handover):** Overall, OD accuracy is affected by multiple factors, namely positioning interval, positioning accuracy, and positioning disturbance (Ping-Pong handover). This research only discusses the effects of positioning intervals because we assume this factor can be decoupled from the other two and investigated in separate studies. Future studies can empirically study positioning accuracy in the presence of positioning disturbances.

- **Disregarding up-scaling the sample OD matrix toward the OD matrix of the entire population:** The available GSM data typically pertains to a sample of the population. Hence the use of GSM data to estimate OD matrices is a two-part problem (e.g., Iqbal et al. (2014); Toole et al. (2015); Mohanty & Pozdnukhov (2020)). The first part is to go from GSM traces of the sample population to an OD matrix of that sample (usually using zonal and temporal aggregation). The second step is to go from the OD matrix of the sample to the OD matrix of the entire population (usually using weighted scaling). In this research, we focus on the first part of the problem - i.e., to derive an OD matrix for the (sampled) GSM traces - while addressing how the polling interval affects the task of zonal and temporal aggregation.

- **Disregarding the mode of transport:** Generally, the Bayesian classifier is sensitive in modeling the activity pattern with a low sampled mode. This study only uses the data belonging to users with the *car* mode. However, there might be some challenges to the approach's generalizability for practical implementation which can be a direction for further research in the future. Although it is worth mentioning that upon analyzing a raw set of GSM data, the mode of transport can be distinguished in two ways: 1) by adjusting our model to detect the mode of transport within the framework and 2) by considering the mode detection as a separate problem before our framework. The former way of addressing mode detection will increase the model errors significantly as the Kernel density estimator uses the features of all modes, with different spatiotemporal patterns, simultaneously. Different modes of transport have a different distribution of features, especially in terms of trip duration. Therefore, the Bayesian classifier adapts itself to put the cutting edge between stay and pass-by on an average value for all modes, producing high deviation relative to associated observed values.

  In this regard, Huang et al. (2019) reviews the literature on transport mode detection with mobile phone network data. Accordingly, mode detection should take place after location type detection because the trip properties like speed, duration, start time, and stay location help detect the mode. However, Some studies used geographic data to extract main transport modes based on proximity to main roads, shortest paths, or train stations with the public transport timetable. These map-matching methods can still be applied before location-type detection. A suggestion to improve the generalizability of the KA method is to adapt mode detection inside KA using an iterative process. The major steps are as follows: 1) location-type detection for the entire data (containing all the modes), 2) applying a similar method we used for activity-type detection to extract the main modes of transport, and 3) re-detection of location type separately for each detected mode. This iterative process potentially leads to simultaneously detecting location types and modes of transport. Validating the suggested iterative approach requires another research and data available on all modes of transport.

2.5 Conclusion and Outlook

GSM data allows observing the location of users over time, but the challenge remains on discerning activity (stay) locations. For this additional information is needed. We show that a Kernel-based approach can provide this location detection, when based on travel diaries of a sample of as little as one percent.

The results presented in this chapter describe how temporal characteristics (i.e., aggregation and discretization) of GSM data affect the accuracy and robustness of the reconstructed OD matrix. It seems that the PI and temporal criterion (i.e., the duration threshold to distinguish *stay* from *pass-by* locations in each user's traveling traces) jointly affect the OD matrix reconstruction.

As perhaps expected, we show that the accuracy of the reconstructed OD matrix gradually declines with higher PI. However, we also show that the reliability of the KA accuracy declines substantially when PI exceeds the duration threshold. Therefore, the combination of larger PI (in data collection) than duration threshold (in OD matrix reconstruction) are best avoided.

Depending on the data context (observable in the training set), fortunate ranges exist for activities with durations and PI less than the duration threshold. These ranges exist due to the data temporal discretization and different interpreted and actual durations. In fact, since the interpreted duration of events defines their type, if interpreted duration of a *stay* be more than the duration threshold (even if actual duration of it be less than the duration threshold), it would be recognized as a *stay*. Our results imply that an "optimal PI" seems to exist which brings about the most structurally similar OD matrix to the ground-truth one. This PI is the minimum value that results in the longest FR. If the duration threshold has the value of T (it can be assumed to be the minimum *stay* duration resulted from applying KA on the training set) the *optimal PI* is about $T/2+\varepsilon$, where ε is a small value. Moreover, the ε better be selected in a way that the interpreted durations not be close to the duration threshold. This increases the robustness of the results against the random seeds adopted for selecting the training set. For instance, when duration threshold is *45min*, we assume the PI to be *25min* for which the minimum interpreted duration (more than duration threshold) is *50min*.

Unavoidably, as it is inherent to GSM data, short-duration activities will remain difficult to detect. Importantly, this study has shown how non-detection or mis-identification mainly occurs in case of unusual activity-travel behavior, e.g. short-duration *home* activities during the afternoon peak are susceptible to be confused with *other* activities due to similar temporal patterns.

Many models, particularly those which are based on regression slopes and intercepts, will estimate parameters for every term in the model. Therefore, having non-informative variables can add uncertainty to the model performance. However, the Bayesian model is naturally resistant to non-informative predictors. Still, due to the naive nature of our model, one needs to select the explanatory variables carefully to avoid biased inferences. Depending on the spatial aggregation level, the location of records might help to infer the activity type. Spatial aggregation is associated with the density and distribution of antennas across the network. In our study, the location of the records does not provide much data on the activity categories. Therefore, incorporating the location in addition to temporal variables in activity category detection did not significantly change the accuracy. In fact, it even slightly reduced the overall accuracy. This may appear counter-intuitive, but is because, in Naive Bayesian models, different prior variables are assumed to be conditionally independent. This assumption can be avoided by establishing the correlation of prior variables, which usually require a more extensive data set than the data available in this study.

In addition to temporal characteristics, this method could use speed and distance between the records to decide on their type (*stay* or *pass-by*). For instance, initially, we assumed that the travel time is the Euclidean distance between two location coordinates divided by the average speed extracted from the training data. However, this assumption did not improve our estimations (i.e. OD matrix) due to two reasons: First, the actual speed has a variety of ranges depending on time, space, and user's behaviour (and mode of transport but here we only considered *car*s). Therefore, it seems that the average speed of the training set is not a proper approximate for speed at various times, spaces, and users. Second, the actual route that the user selects to get from origin to destination is generally longer than the Euclidean distance. Thus, due to the required extra efforts, we did not modify the travel times in the data. Whereas, depending on the data availability, future research could use either the speed with route data or directly the travel time approximates to enhance the O-D travel times.

Future research should be undertaken to explore how we can improve OD reconstruction accuracy through data-driven approaches. The authors suggest three ways to improve the performance of KA in location-type detection. One is to simultaneously evaluate the travel motifs of all users in training data to identify the different feature distributions for each primary activity tour across the network. This step can be performed using Kernel density estimation and the Bayesian modeling we used in this research. Later on, we can assign the most probable activity motif based on the features of each individual. Another way of improving location-type detection is to assess the repetition of visited locations at the network level (i.e., all users) and each user. Evidently, evaluating repetition patterns per user require data for a longer period.

Another way of improving location and activity type recognition is adding different features of TAZs to the analysis. These features include mixed land-use, population density, road density, and dominant demographic information. For instance, a TAZ with large commercial/industrial areas is more probable to be a *stay* locating for *work* activity. Adding the mentioned spatial features also helps increase the detail level and detect activity types more disaggregate. It goes without saying that higher spatiotemporal resolution might also be required to increase the level of detail in activity category detection. For instance, given that a particular location holds an entertainment event at a particular interval, observing a close-by record implies *attending an entertainment activity* more probable.

The experimental setup and newly developed method for OD matrix estimation provides more avenues for future research. Firstly, the framework can be extended towards CDR data where polling intervals are irregular and endogenous (instead of regular and exogenous as with PLU data). Secondly, the KA model can be adapted to address the effects of spatial characteristics of GSM data (i.e. spatial accuracy of data as well as OD matrix). Thirdly, the KA model can be adapted to address the problem of mode detection (i.e. distinguishing mode of transport for GSM traces, based on e.g. average speed, location and time-of-day). For the latter two studies again a sub-sample of travel diaries can be used as training set.

Chapter 3

Distinguishing the Spatial-Temporal Travel Demand Patterns

Building on the insights gained from the previous chapter, where we explored the nuances of GSM data in travel demand estimation, this chapter pivots from the methodological focus of interpreting raw data. We perform a deeper analysis of the spatial-temporal heterogeneity in trip production, particularly in the context of OD matrix estimation and prediction using gravity models. Transportation systems, with their inherent complexity and constant evolution, present substantial challenges in accurately modeling and predicting travel demand. This chapter particularly highlights how this complexity and heterogeneity contribute to spatial-temporal variations in trip production and attraction, influencing the predictability and reliability of OD matrix estimations. By examining the variability in trip production at different times of day and days of the week, we shed light on the dominant patterns that emerge among traffic analysis zones (TAZ), providing valuable insights for more effective transportation planning and policy-making.
This chapter is based on the following papers:

Eftekhar, Z., Pel, A., & van Lint, H. . (2023). A Cluster Analysis of Temporal Patterns of Travel Production in the Netherlands: Dominant within-day and day-to-day patterns and their association with Urbanization Levels. European Journal of Transport and Infrastructure Research, 23(3), 1–29. (published)

Eftekhar, Z., Behrouzi, S., Krishnakumari, P., Pel, A., & van Lint, H. .The Role of Spatial Features and Adjacency in Data-driven Short-term Prediction of Trip Production:An Exploratory Study in the Netherlands (under review)

Abstract

This research explores temporal patterns in travel production using a full month of production data from traffic analysis zones (TAZ) in the (entire) Netherlands. The mentioned data is a processed aggregated derivative (due to privacy concerns) from GSM traces of a Dutch telecommunication company. This research thus also sheds light on whether such a processed data source is representative of both regular and non-regular patterns in travel production and how such data can be used for planning purposes. To this end, we construct normalized matrix (heatmap) representations of weekly hour-by-hour travel production patterns of over 1200 TAZs, which we cluster using K-means combined with deep convolutional neural networks (inception V3) to extract relevant features. A silhouette score shows that three dominant clusters of temporal patterns can be discerned (K=3). These three clusters have distinctly different within-day and day-to-day production patterns in terms of peak period intensity over different days of the week. Subsequently, a spatial analysis of these clusters reveals that the differences can be related to (easily observable) land-use features such as urbanization levels (i.e., Urban, Rural, and mixed-level). To substantiate this hypothesis and the usefulness of this clustering result, we apply an OVR-SMOTE-XGBoost ensemble classification model on the land-use features of the TAZs (i.e., to identify their cluster). The results of our clustering analysis show that given the land-use features, the overall production patterns are identifiable. These findings are directly useful for data-driven estimation and prediction of demand time series. Furthermore, this study provides further insights into people's mobility, relevant for transportation analysis and policies.

3.1 Introduction

Call detailed records (CDR) can show valuable insights in various contexts such as daily mobility motifs (Schneider et al., 2013; Jiang et al., 2017), population movements (Antoniou et al., 2020; Szocska et al., 2021), and disaster response (Yabe et al., 2022). While all the mentioned studies use raw trajectories, acquiring this type of data is difficult. Telecom operators are banned from providing ground truth or contextual information to protect people's privacy at the—inevitable—expense of data utility. Usually, what is available is processed aggregated OD matrices for which the methodology used for processing and up-scaling to the whole population is unclear. This data is a valuable source of information on people's mobility, but its suitability for transportation planning remains unclear. Are these OD matrices representative of the population demand variations and irregular patterns? In many cases, due to the limited market share of the telecom operator, only a small and possibly behaviorally biased sample of people is presented in the raw data; How well do they represent the whole population? Mamei et al. (2019) raises attention to the need for more research and experiments assessing and evaluating the OD matrices derived from these data sources.

To make these processed data useful for urban planning, we must clarify their potential biases even when facing limited data resources to test these against. In this regard, *consistency* of data deals with its representativeness and correspondence to the real world. Essentially, data consistency has two main aspects: uniformity across the dataset and alignment with expected patterns (Demchenko et al., 2013). The consistency of the data can be assessed by using multiple sources when available, checking data credibility using various quantitative tools, and evaluating the internal data structure (Rubin & Lukoianova, 2013). In this chapter, we evaluate the consistency of the OD matrices resulting from CDR both by assessing their associated travel demand patterns (representing internal data structure) and by comparison with data from an independent source, in our case, land-use characteristics.

Understanding travel demand patterns is essential for transportation planning and management. Firstly, travel demand analysis plays a significant role in identifying the current problems of transportation systems and helps in modeling the future traffic state (Thakuriah, 2001; Rich & Mabit, 2012). Secondly, the demand patterns help evaluate the impact of transportation infrastructure and management policies and strategies, such as flexible-time work schedules and congestion pricing (Gärling et al., 2002). Thirdly, understanding demand patterns is useful for developing better standards for evacuation plans, and responses (Pel et al., 2011; Xu et al., 2017).

Fekih et al. proposes a framework to extract spatiotemporal travel demand patterns from large-scale GSM traces. Their analysis focuses on within-day variations of travel demand. In this research, we build on this work and investigate both within-day and day-to-day production patterns of all the traffic analysis zones (TAZs) in the Netherlands. We tried to capture a more holistic picture of temporal production patterns through normalized heatmaps. After feeding these heatmaps to a deep convolutional neural network (DCNN) and K-means method, three main patterns were discerned. Analysis of these patterns is one of the two methods of evaluating data consistency. The performed temporal analysis of the underlying patterns is valuable for adjusting the demand models and prediction. Later we link these temporal patterns to spatial urbanization levels through their land-use characteristics, which is beneficial for urban development strategies and policymakers. In fact, this study proposes three urbanization levels for the Netherlands: urban, rural, and other, associated with their specific land-use characteristics and travel production patterns within the day and day-to-day, and this study explores the differences in these patterns. This effort is the second method of assessing the data consistency. Furthermore, an OVR-SMOTE-XGBoost ensemble classification method is proposed to investigate the relationship between land-use characteristics and temporal production patterns. Our findings suggest that given the land-use features of each TAZ, their most probable travel production temporal pattern is detectable. The results are beneficial for dynamic demand prediction models. Furthermore, we find indications for the data's ability to represent both dominant patterns and variations correctly in spite of the data's inherent bias. This study will help planners discern and assess the representativeness and suitability of using the processed, up-scaled derivative of GSM traces in traffic planning.

The remainder of this chapter is organized as follows: Section 3.2 describes the research data and the implemented method. In Section 3.3, we present and discuss the results of our analysis on the temporal patterns found in the Netherlands. Finally, Section 3.4 concludes the chapter.

3.2 METHODOLOGY

To understand the temporal patterns in travel production and how these might be associated with spatial urbanization levels, we can distinguish three main research components (as shown in the overall research framework in Figure 3.1: the data showing the spatiotemporal travel productions, the clustering method to discern temporal patterns, and the association analysis method to discern spatial relations. The first two components are explained one by one in the following subsections. The association analysis is included in section 5.2.

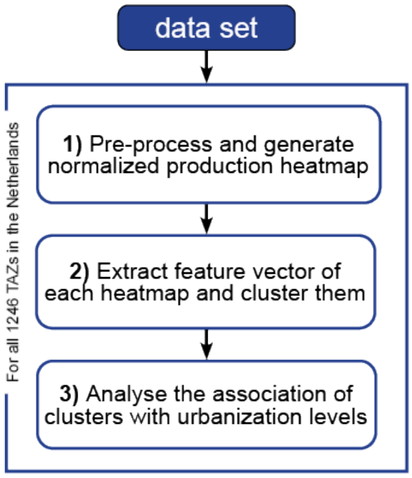


Figure 3.1: Overall research framework.

3.2.1 Travel production data

This research uses the hourly production of the 4-digit postal code zones in the entire Netherlands during March 2017. Using 4-digit postal code zones leads to 1246 TAZs in the Netherlands (see Figure 3.2). **Travel production** of TAZ i is defined as the number of inter-zonal trips

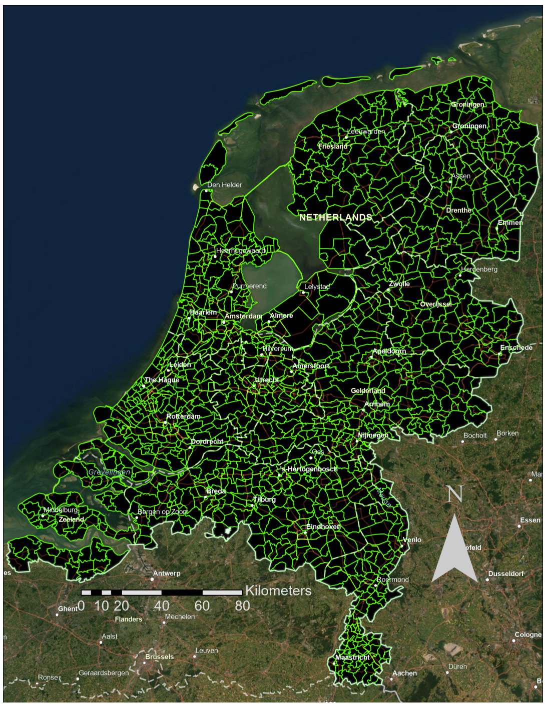


Figure 3.2: TAZs in the Netherlands.

made by motor vehicles starting at i. The production values are derived from the GSM traces of the Dutch telecommunication company Vodafone, whose market share is about one-third of the Dutch population. Another company performed the processing due to privacy concerns regarding the raw mobile phone data. Consequently, the available data for this study, instead of the mobile phone traces, consists of origin-destination (OD) matrices of the motor vehicles based on TAZs in the Netherlands. These OD matrices have been initially scaled up to the entire Dutch population. We refer the reader to Meppelink et al. (2020) for more detail on the scaling procedure.

To begin with, we collected the (processed) data. After pre-processing and reshaping, each zone had a heatmap of normalized production values. Normalizing the production values enables fast and stable pattern comparison of various zones. The technique we applied on each production value x for normalizing is Min-Max Scaling, i.e.,

$$x_{normalized} = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{3.1}$$

where $x_{normalized}$ is the normalized value, x_{min} and x_{max} are the minimum and maximum production values in the (month-long) time series of that particular TAZ. Thus the resulting normalized values range between 0 and 1.

The result is a production heatmap for each TAZ in which the horizontal and vertical axes represent the days of the month and the hours of a day, respectively. Figure 3.3 shows an example. This representation allows us to see the temporal patterns of production both within

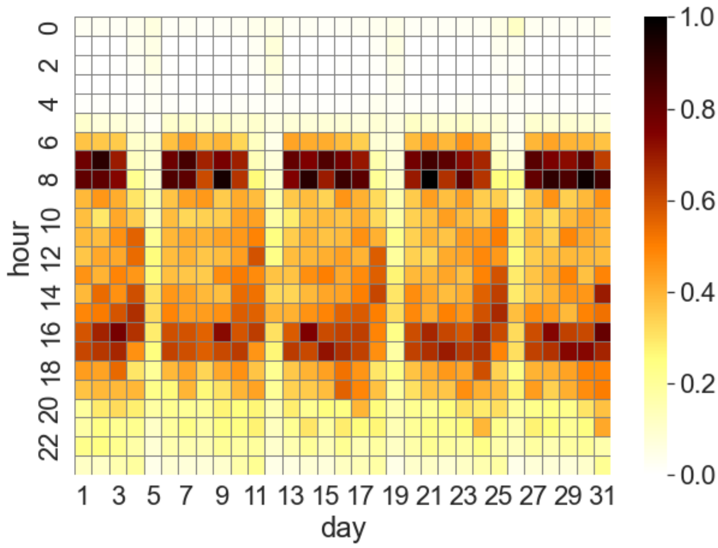


Figure 3.3: An example of production heatmap for one TAZ.

a day (comparing across rows) and between days (comparing across columns). These 1246 heatmaps of travel production are the basis for the following analyses.

3.2.2 Clustering temporal production patterns

The heatmaps of hourly travel production per TAZ are clustered based on (temporal) similarity using K-means clustering. Due to its easy application and effectiveness, the K-means clustering method is one of the most popular algorithms for clustering analysis (Poteraş et al., 2014; Szegedy et al., 2016; Cohn & Holm, 2021; Van Gansbeke et al., 2020). The method works by splitting the N-dimensional data set of M points (heatmaps) into K clusters such that the sum of the pairwise Euclidean distance between the points of each cluster is minimized (Hartigan &

Wong, 1979). In other words, this method aims to maximize the similarity between the points in the same cluster and maximize the dissimilarity of points from different clusters. The method's initialization is by randomly selecting K points as the cluster centroids. The clustering process has two significant steps:

- **Assignment:** assigning each point to its closest centroid. Mathematically this step refers to partitioning the points to the Voronoi diagram (Shamos & Hoey, 1975) generated by the centroids.
- **Update:** updating each cluster center to be the average of all points contained within them.

Applying the K-means method means that the number of clusters is exogenously chosen and thus needs to be justified. As we cannot determine how many patterns exist in advance, this is an unsupervised learning problem. That is, there are no "true labels" or ground truth available (i.e., there is no a-priori behavioral hypothesis on the existence of specific patterns), and therefore the appropriate number of clusters is best chosen by assessing the (dis)similarity within and between clusters, for different values of K. To calculate the goodness of clustering, we used the Silhouette index (Rousseeuw, 1987). There are several methods for clustering evaluation, such as distortion score (Camps-Valls, 2006), Rand index (Yeung et al., 2001), adjusted Rand index (Hubert & Arabie, 1985), and Silhouette index. The latter is particularly useful for an unsupervised evaluation of clustering.

The Silhouette ranges between -1 and 1, where high values show a well-matched point to its own cluster and poorly matched to the neighboring clusters. If many points have a negative value, the number of clusters needs to be modified. The Silhouette of point i is calculated as $s(i) = \frac{x(i)-y(i)}{max(x(i),y(i))}$ where, x is the average distance with points in other clusters (i.e., dissimilarity), and y is the average distance to points of the same cluster. uster is the average $s(i)$ of all the points inside the cluster, and the Silhouette index for all clusters (SI) is the average $s(i)$ of all points in all clusters. The optimal number of clusters happens when the SI is maximum. That is when on average, the difference between mean intra-cluster and nearest-cluster distance is the highest. For instance, if we calculate SI for the different number of clusters (i.e., K) in the range of $2,15$ and the maximum values of SI happen when $K = 3$, the optimal K equals 3.

To evaluate the separateness of clusters, the inter-cluster distance is commonly used (Liu et al., 2012). Visualizing the inter-cluster distances in two dimensions gives insight into the relative importance of clusters. To this end, multidimensional scaling (MDS) embed multi-dimensional cluster centers into 2-dimensional space (we refer the reader to Bengfort et al. (2018); Kruskal (1964)). This method preserves the distance to other centers. For instance, if cluster centers are close to each other in the original feature space, they are also close when embedded into 2-dimensional space. In the inter-cluster distance map, the clusters are sized according to the number of instances that belong to each cluster which in a sense reflects how important each cluster is.

Another issue that needs to be addressed is that the K-means method is inherently a linear algorithm (Ning & Hongyi, 2016). Therefore, it is unsuitable for complex nonlinear data distributions. To take the non-linearity of data into account, a deep convolutional neural network (DCNN) is used for feature extraction. The DCNN transforms input heatmaps to final representations, so-called feature vectors, which are more easily separable by a linear clustering algorithm (van Elteren, 2018) than the original heatmap. To this end, the DCNN identifies key (salient) features of the heatmaps for analysis and clustering, and after this transformation step, we have a set of feature vectors that are then used for K-means clustering (instead of applied directly on the heatmaps).

As DCNN, the state-of-the-art InceptionV3 based on transfer learning is used. The InceptionV3 architecture specifically improves adaptability to different scales, and overfitting is better prevented (Kaur & Gandhi, 2020). One way in which this is done is by using transfer learning. Transfer learning lets us transfer already trained model parameters to our new model and therefore accelerates its training (Wang et al., 2019). After training a DCNN on a large dataset, e.g., ImageNet, one can adopt transfer learning because the DCNN is able to learn generic features (think of edges or other geometric shapes) that are also applicable to other images (such as heatmaps) without the need for training from scratch. Furthermore, the weights of the DCNN, which is pre-trained on a large dataset, improve its accuracy for specific tasks such as pattern recognition tasks in which the amount of available training data is limited (Iglovikov & Shvets, 2018). This also holds for our data set. In this study, we thus extract feature vectors from the demand heatmaps using the InceptionV3 deep neural network, which is pre-trained on the ImageNet dataset consisting of millions of images used for object recognition and image classification. For details about this dataset, we refer the reader to Deng et al. (2009).

In summary, this step transforms the 1246 heatmaps of travel production into feature vectors that are then clustered based on (temporal) similarity, yielding K clusters with a distinct temporal production pattern, and in the process, determine how many (K) of such clusters/patterns exist.

The Supplementary data and selected hyperparameters associated with this chapter will be presented online.

3.3 RESULTS AND DISCUSSION

The first step in our research method was to apply the DCNN to the travel production heatmaps for feature extraction. Because we needed to use a pre-trained model (i.e., via transfer learning), we had to resize the input to the same format the network was originally trained on, that is, 224 by 224 pixels (numbers). Subsequently, the resulting feature vector had 224^2 features representing each heatmap. We clustered all these vectors of the entire 1246 TAZs in the Netherlands and calculated SI for various K values to determine the optimal number of clusters as shown in Figure 3.4. The scores on the y-axis are the average Silhouette index of all the features, SI, that each TAZ includes (i.e., travel production feature vector). Accordingly, the best cluster separation occurs when $K = 3$ because the maximum SI belongs to the case where $K = 3$. It is worth mentioning that SI for $K = 3$ is relatively close to that of $K = 2$. Still, this slight difference cannot be treated insignificantly. After the feature extraction step, the trip production features constitute large vectors containing many zeroes, and as the SI is the product of averaging the euclidean distances, any difference is meaningful — indicating non-zero elements of trip production features. Also, based on the inter-cluster distance map in Figure 3.5, the three clusters seem to be well-separated. Clustering of the temporal patterns of production was performed using inceptionV3 with the K-mean method.

The remainder of this section is divided into two parts: Firstly, we show and analyze the results of clustering the travel production heatmaps in Section 3.3.1; Secondly, in Section 5.3.1, we discuss the association between the extracted temporal clusters and spatial characteristics.

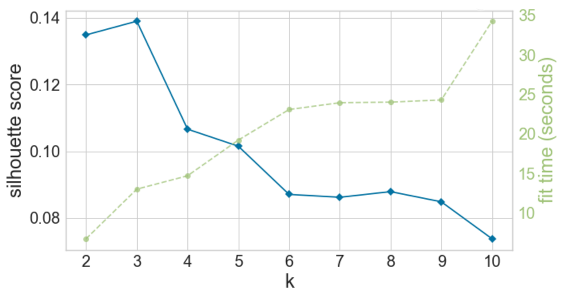


Figure 3.4: Silhouette index for finding the optimum number of clusters.

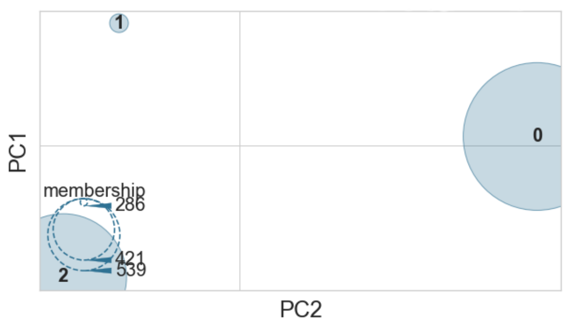


Figure 3.5: K-means inter-cluster distance map when K=3 and MDS is used for dimensionality reduction.

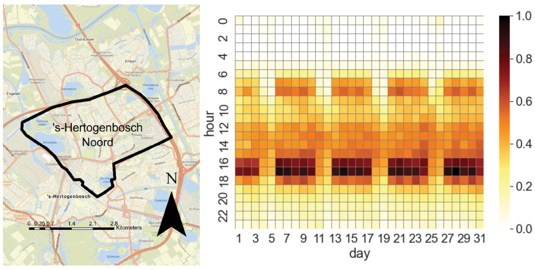
3.3.1 Travel production patterns

Examples of the three clusters of temporal heatmaps of travel production are given in Figure 3.6.

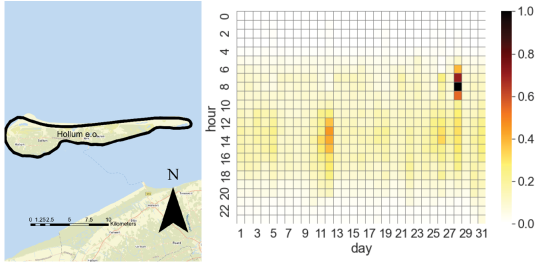
Of each cluster, two TAZ heatmaps are shown: the left images represent the closest heatmap to the cluster centroid, and the right pictures show the furthest heatmap from the cluster centroid.

For the first cluster, as shown in Figures 3.6(a) and 3.6(b), containing 286 TAZs, we generally observe the presence of afternoon peaks between 15:00 and 19:00 in travel demand. Therefore, this cluster is named *AP*. Unlike the working days, weekends do not have significant peaks. Thus the afternoon peak could be due to more work-related areas, i.e., people tend to leave work in the afternoon, which causes a rise in travel production. Figure 3.6(b) belongs to the first cluster, although both its location and temporal production pattern seem to be indicating an outlier. Figure 3.7 shows the Silhouette score distribution per cluster. As negative scores reflect poor heatmap-to-cluster matches, we can see that cluster *AP* contains more outliers than the other two.

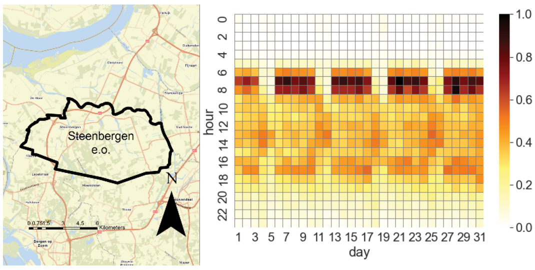
The second cluster shown by Figures 3.6(c) and 3.6(d), contains 421 TAZs, and displays more distinct morning peaks between 06:00 and 09:00 in travel demand. Hence, this cluster is named *MP*. As this morning peak is predominantly observed on working days, this could be due to residential areas, i.e., people leaving their houses in the morning. Compared to cluster *AP*, the smaller peak interval reflects more scheduled activities (e.g., starting time of work) in the mornings and less obligation to leave on time (e.g., from work) in the afternoons. Additionally, less regular activities like shopping and social events in the afternoon in cluster *AP* also trigger the longer peak range. In Figure 3.6(d), which falls further away from the cluster centroid, the afternoon peak starts to become more severe. In fact, a further increase in the afternoon values



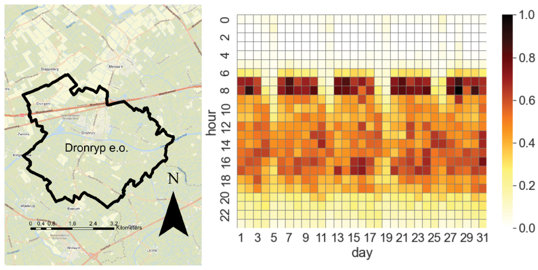
(a) Cluster AP



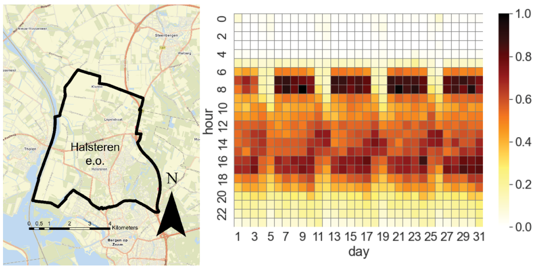
(b) Cluster AP



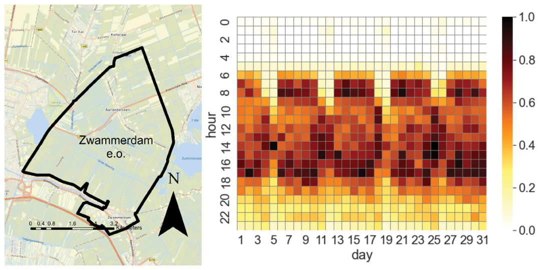
(c) Cluster MP



(d) Cluster MP



(e) Cluster *mix*



(f) Cluster *mix*

Figure 3.6: Travel production clusters.

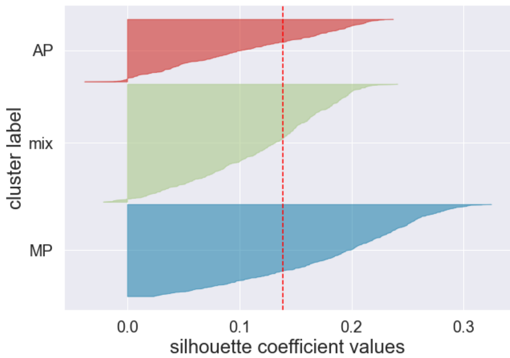


Figure 3.7: Silhouette score distribution per cluster.

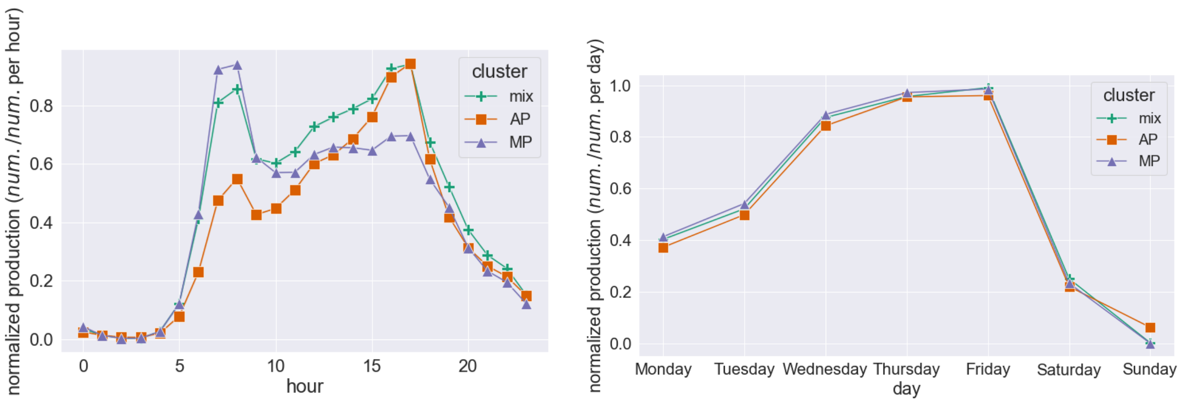
would probably cause this TAZ to be assigned to the third cluster. According to Figure 3.7, this cluster has higher Silhouette scores and lower outliers (i.e., negative Silhouette scores).

The third cluster, named *mix*, containing 539 TAZs, displays patterns other than those observed in cluster *AP* and *MP*. For instance, in Figure 3.6(e), morning and afternoon peak seems almost equally extreme with higher values and scheduled activities in the morning, which can be a presenter of suburban areas where a mix of residential and work-related activities is established. Figure 3.6(f), on the other hand, presents a different pattern: This TAZ seems to be a weekend trip producer as some peaks are observed during the weekends and Friday afternoons. Moreover, the afternoons seem to demonstrate higher values.

Visual inspection suggests only subtle differences between heat-maps 3.6(d) and 3.6(f). We did not apply a separate statistical test (e.g., Kolmogorov–Smirnov) because the two-stage workflow itself already guarantees separation: Inception-V3 creates feature vectors, and K-means then forms clusters by explicitly maximizing the distance between them. Running an extra test would simply confirm what the clustering algorithm has already enforced.

To investigate the dominant *within-day* patterns per cluster, Figure 3.8(a) shows the average normalized hourly travel production of the three clusters. To this end, for each TAZ, the hourly travel demand is averaged across all days of the month and then is min-max normalized across these 24-hour values. And then, for each cluster, these normalized 24-hour patterns are averaged across all TAZs inside that cluster. This leads to the results shown in Figure 3.8(a), which indeed confirm the earlier observations of morning peak, afternoon peak, and mixed temporal patterns.

To investigate the dominant *between-day* patterns per cluster, Figure 3.8(b) shows the average normalized daily travel production of the three clusters. To this end, for each TAZ, the hourly travel demand is summed per day, and then an average daily demand is computed for each day of the week, and then min-max normalized across these 7-day values. For each cluster, these normalized 7-day patterns are averaged across all TAZs inside each cluster. This leads to the results shown in Figure 3.8(b). The between-day patterns in all clusters demonstrate a similar trend, although cluster *AP* shows slightly higher values on Sundays, compensating for lower values during the rest of the week.



(a) Hourly production. (b) Daily production.

Figure 3.8: Average of normalized travel production of the three clusters.

Figure 3.9 shows the *absolute within-day* travel production in each cluster. The values are computed similarly as for Figure 3.8(a) but without normalizing. The shaded area shows the $90th$ and $10th$ percentile. As expected, the values are widely scattered, especially in cluster *AP*, with higher average production in almost all hours of the day. In contrast, cluster *MP* displays the lowest average production throughout the day. The morning and afternoon peaks of all clusters seem to be almost in the same time range; however, their difference in the production values are more significant during the afternoon peak. Overall, the means of clusters seem to be significantly different.

Figure 3.10 shows the *absolute between-day* travel production in each cluster. The values are computed similarly as in Figure 3.9 but without normalizing. The shaded area shows the $90th$ and $10th$ percentile.

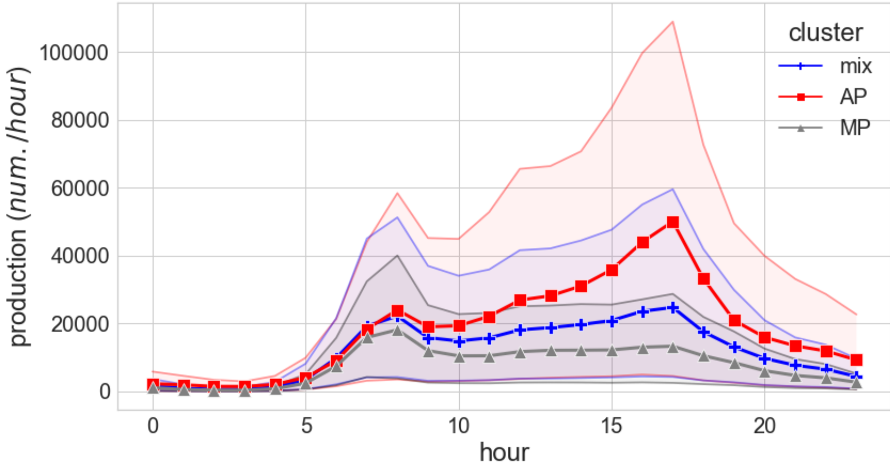


Figure 3.9: Average hourly production of the three clusters with 10th and 90th percentile as the shaded area.

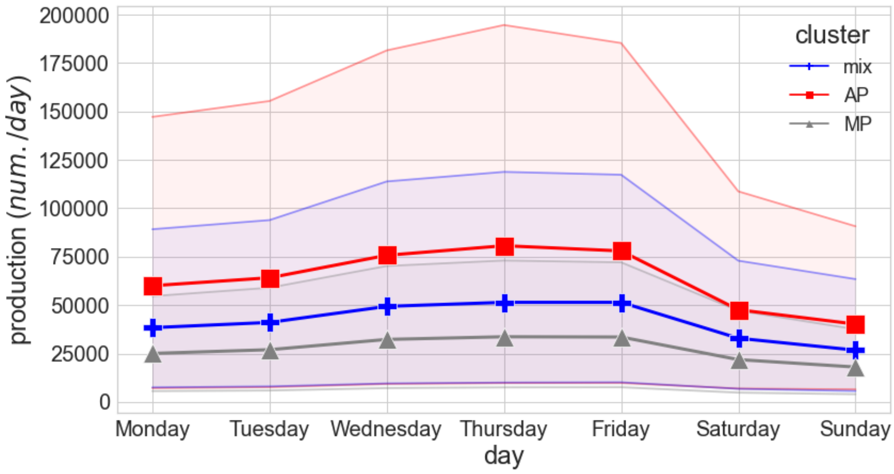
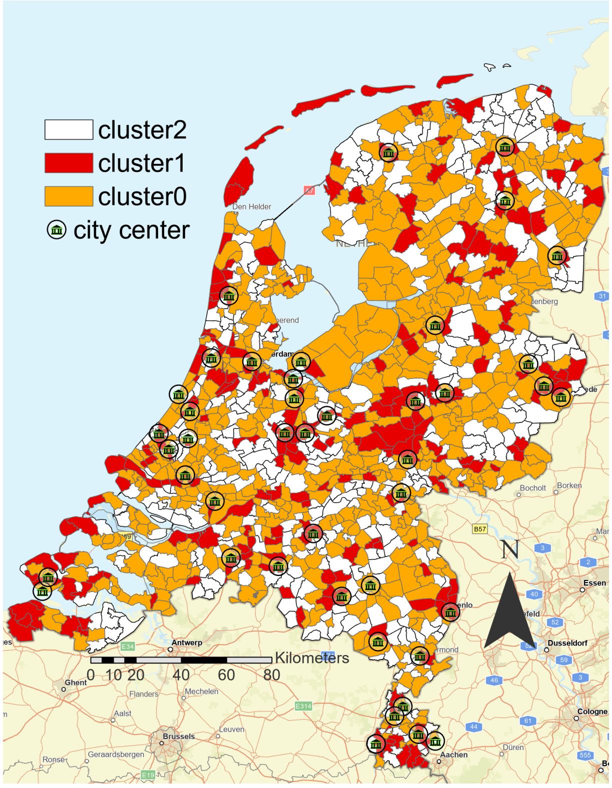
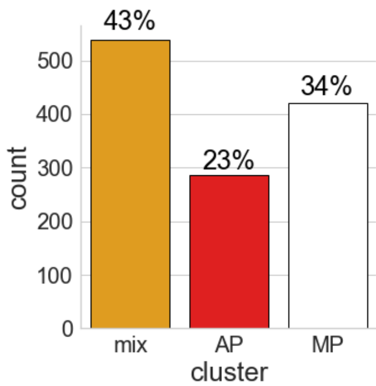


Figure 3.10: Average daily production of the three clusters with 10th and 90th percentile as the shaded area.

In the same way as hourly patterns, weekly average values seem to be well-separated. However, the Figure shows high variance meaning a widely scattered distribution of values. Moreover, the daily patterns seem very similar among all the clusters. A slight difference lies in the days with maximum averages. These are Thursday and Tuesday for cluster *AP*, which indicates working areas. The days with maximum production for cluster *mix* and cluster *MP* are Tuesday, Thursday, and Friday, which can hint that work is not as dominant as it is in cluster *AP*, i.e., as some part-time employees do not work on Fridays, having growth in the production values mainly indicates non-work-related trips. Furthermore, unlike Figure 3.8(b) in which cluster *mix* holds the maximum mean, Figure 3.10 display highest mean for cluster *AP*. Once again, it implies that productions in cluster *mix* have a smaller standard deviation (i.e., values are closer to their maximum) which produces higher values in the normalized production plot.



(a) Clusters and city centers.



(b) Cluster share count.

Figure 3.11: Travel production pattern clusters in the Netherlands.

3.3.2 Research limitations

This chapter introduces a data-driven framework to show day-to-day and within-day trip production patterns. Our method identifies the dominant temporal patterns without prior assumptions on the number of patterns. The last part of this research, as subsequently described in Chapter 5, investigates the association of the identified patterns with spatial characteristics. Clearly, several limitations should be borne in mind when interpreting the results:

- Raw CDR can offer ample information about millions of mobile phone users. Nevertheless, their limitation concerns the privacy issue. Consequently, the data available for conducting this study was initially processed, aggregated, and transformed into origin-destination tables by another party. Inevitably, the level at which we receive the data lacks details on the methodology used for deriving the resulting dataset. For instance, the algorithms for separating motor vehicles from other modes of transport, scaling the OD data to the entire dutch population, and mapping from the antenna Voronoi diagram to TAZs is unclear. Despite not fully understanding how the OD matrices were created, we used the data as given and focused on what we could learn from it. It's important to note that while we've done our best to explain the data, some of the finer details about how it was produced aren't clear to us. We accepted the matrices as they were, using them to guide our research and draw conclusions.

- Due to privacy reasons and ease of trip identification, the short-distance trips, mainly occurring inside each TAZ, were initially eliminated. Consequently, the present analysis only considers longer distance mobility with possibly more regular temporal patterns.

- In this research, we used origin-destination tables collected over only one month. This duration might not be sufficient to analyze the data's temporal stability. Furthermore, the resolution of the dataset is at the level of 4-digit postal code zones, which might be too coarse to evaluate the effect of special events on the trip production of TAZs.

3.3.3 Research implications

An implication of this study is the possibility of assessing the effect of developing/changing an urban area on trip production patterns. This can be particularly useful for decision and policy-makers. One possible direction for future research is using our framework to assess the effect of new urban areas on energy consumption and environmental footprint due to changes in temporal traffic patterns. Another benefit of the present study is that it introduces a framework for assessing the effect of special events on trip production patterns. To perform such an analysis, we need to select the spatiotemporal resolution of the data based on the extent of the event. Although this study focused on trip production patterns, one can apply a similar framework for trip attraction and OD demand.

3.4 CONCLUSIONS

The present study investigated the consistency and trustworthiness of the processed aggregated derivatives of GSM traces in the form of OD matrices. To this end, we analyzed the temporal patterns of travel production in TAZs of the Netherlands. The travel production of different areas reveals different temporal patterns within a day and between days. Clustering these patterns using a DCNN based on transfer learning with the K-means method introduced three distinct clusters in the Netherlands. One with a distinct afternoon peak, one with a distinct morning peak, and another with a mix of both. Further evaluation of the patterns shows robust temporal patterns in normalized production values. On the other hand, absolute values show a wider range of values for all the clusters reflecting both the regular and irregular patterns.

Notwithstanding the relatively limited sample (one month of travel production at a specific spatial-temporal scale), this work establishes a quantitative framework for detecting hourly and daily temporal patterns and how spatial urbanization level helps in demand modeling. Moreover, such analysis is required before using the processed demand data for policy-making and network development. Further research could also be conducted to determine the effectiveness of using the land-use characteristics (on top of other variables) in improving the demand prediction models.

Chapter 4

Prediction of Travel Demand at Multiple Spatial Scales

This chapter addresses the research gap in understanding the predictability of travel demand at different spatial scales. Despite a significant amount of research on travel demand prediction, there is a lack of direct comparison between studies due to differences in study design and the specific factors being considered. This chapter aims to fill this gap by directly comparing the predictability of trip production at various spatial scales, providing insights into how different spatial scales may impact the predictability of travel demand when using a gravity model.
This chapter is based on the following papers:

Eftekhar, Z., Pel, A., & van Lint, H. . (2023). A Cluster Analysis of Temporal Patterns of Travel Production in the Netherlands: Dominant within-day and day-to-day patterns and their association with Urbanization Levels. European Journal of Transport and Infrastructure Research, 23(3), 1–29. (published)

Eftekhar, Z., Behrouzi, S., Krishnakumari, P., Pel, A., & van Lint, H. . The Role of Spatial Features and Adjacency in Data-driven Short-term Prediction of Trip Production:An Exploratory Study in the Netherlands (under review)

Abstract

Large-scale prediction of trip production is one of the minimal ingredients for OD demand estimation and prediction. The main challenge in predicting trip production is its spatial-temporal dependence and heterogeneity. Most of the studies focus on the temporal correlations in predicting the demand but usually do not consider the spatial adjacency of mobility zones (MZ). Consequently, the effect of spatial uncertainty (associated with the spatial heterogeneity) on the trip production prediction across different spatial scales is not fully understood. Moreover, in large-scale trip production, one needs to train one separate model for each spatial area which requires extensive computation time and capacity. As the practicality of demand prediction model complexity depends on its required computational power, the computation time needs to be minimized. This research proposes a method that integrates graph convolutional neural network (GCN) into the long short-term memory network (LSTM). By introducing a nationwide graph reflecting the adjacency of MZs, spatial heterogeneity is well addressed in the prediction process. Also, instead of training one separate prediction model for each zone, one single prediction model is trained for the entire spatial extent. Therefore, the computational time decreases significantly. Furthermore, We quantitatively demonstrate the effect of spatial uncertainty on the prediction under various administrative spatial scales. The findings of this research have important implications for improving the OD demand prediction models.

4.1 Introduction

Short-term travel demand prediction is crucial for effective policy adjustments, management, and operations in various domains (Qian et al., 2020; Xiong et al., 2020). A critical aspect of demand prediction is forecasting the number of trips originating from a specific location, referred to as *trip production*. Although accurate trip production prediction is vital for estimating and predicting origin-destination (OD) demand, it remains a challenging task due to its complex spatial-temporal dependence and heterogeneity (Krishnakumari et al., 2019). In this chapter, “trip production” denotes the number of outgoing trips from a particular location or zone i.

Predicting trip production is challenging due to the heterogeneity caused by the diverse characteristics of traffic analysis zones (TAZs) (Shen et al., 2020). **Spatiotemporal heterogeneity** refers to the variations in the correlations or distributions of variables across different geographical regions and time intervals. Spatiotemporal heterogeneity in demand implies that the effects of land-use properties on travel demand patterns may not be consistent across different areas or periods. This heterogeneity within one TAZ and across TAZs originates from a range of factors, including diverse urbanization levels, population demographics and lifestyles, economic activities, transportation accessibility, and resource distribution across areas (Fotheringham et al., 1998). When a single TAZ encompasses a mixture of (diverse) characteristics, it can lead to spatial heterogeneity in trip production, introducing uncertainty in predicting trip production patterns. For example, trip production in certain parts of a TAZ may peak in the afternoon, while in others, it may peak in the morning, and in some cases, it may follow a completely different pattern (Atluri et al., 2018). Considering these challenges, this chapter aims to develop data-driven prediction models that explicitly account for spatial-temporal heterogeneity, investigate the effect of spatial scale on prediction accuracy, and examine the association between prediction errors, residual patterns, and socio-spatial features at different spatial scales. Understanding these relationships can significantly enhance the quality of trip production predictions and contribute to more effective urban planning and transportation policy decisions.

In the temporal dimension of demand data, various patterns, such as periodicity, linear and non-linear trends, and holiday effects, can be observed, significantly impacting the quality of predictions. This temporal variability is particularly substantial in urban areas, where morning and afternoon peaks contribute to nearly 50% of daily travel production (Lin & Shin, 2008). This variability not only exacerbates the spatial heterogeneity described earlier but also has its own effects. The coarser one discretizes the periods over which production is predicted, the larger the prediction errors may be, depending on both this temporal variability *and* the chosen period boundaries.

Numerous researchers highlight that neglecting this combined spatiotemporal heterogeneity results in higher prediction errors and that addressing spatial-temporal heterogeneity is essential for understanding and predicting travel production (e.g., (Anselin & Griffith, 1988; Deng et al., 2018; Cheng et al., 2019)).

Our literature review reveals a lack of data-driven models that consider these factors. Consequently, to inspire potential approaches, we first provide a brief overview of data-driven prediction methods applied to traffic data (speed and flow) to address these challenges.

4.1.1 Related Work

Various studies have investigated the temporal correlation of traffic data using diverse methods tailored to different application scenarios. Traffic forecasting approaches can be broadly divided into two categories: traditional statistical methods and deep-learning-based methods.

Traditional statistical methods include linear regression models (Sun et al., 2003), Kalman filtering (Yang, 2005), autoregressive integrated moving average (ARIMA) (Liu et al., 2021), K-nearest neighbor (KNN) (Xu et al., 2020), least squares support vector machines (LS-SVMs) (Zhang & Liu, 2009), particle filter (Wang et al., 2016), hidden Markov model (Qi & Ishak, 2014), and Gaussian process (Xie et al., 2010). These methods often rely on strict mathematical deductions and well-defined physical meanings, limiting their applicability to less complex traffic conditions and/or smaller traffic data sets. Furthermore, most traditional methods assume linearity (in their parameters) and consider stationary underlying processes generating the data, making them less suitable for non-linear dynamics and non-recurrent situations that are characteristic of traffic demand dynamics.

Deep-learning-based methods, the second category, include convolutional neural networks (CNNs) and recurrent neural networks (RNNs). CNNs, along with their deeper architecture (e.g. ResNet (Zhang, 2021)), are typically employed for spatial structure learning, while RNNs (e.g., LSTM (Hochreiter & Schmidhuber, 1997) and gated recurrent unit (GRU) (Chung et al., 2014)) are widely used for temporal and sequential learning. For instance, ST-ResNet (Zhang et al., 2017) employs a residual network for spatial correlation learning and LSTM for modeling time-series data. However, ST-ResNet and subsequent works (Zhang et al., 2017; Liang et al., 2018; Zhang et al., 2019) mainly focus on "coarse-level" citywide traffic flow estimation using taxi or bicycle data.

Deep-learning-based methods often require large datasets and high computational capacity due to the large number of parameters that need to be trained. Moreover, for large-scale trip production, separate models must be trained for each spatial area, necessitating extensive computation time and resources. When developing transportation models, it is essential to find a balance between the complexity of the model and the available computing resources (Raadsen et al., 2020). This is crucial because model acceptability often depends on computation time, which must be minimized without compromising accuracy. Additionally, most non-parametric methods mentioned above focus solely on temporal correlations without reflecting spatial heterogeneity.

Recent advancements in traffic prediction models have utilized deep learning techniques to capture spatial-temporal dependencies. For example, Wu et al. (2019) proposed a Graph WaveNet model for traffic forecasting, which integrates graph convolutional networks with temporal convolutional networks. Similarly, Guo et al. (2019) developed an attention-based spatial-temporal graph convolutional network (ASTGCN) for traffic flow prediction. These studies highlight the ongoing development of models that effectively capture complex spatial-temporal relationships.

Starting with travel demand prediction, the comparison study of Li & Axhausen (2019) on short-term traffic demand prediction methods reveals no singularly superior model across several time-series models, machine learning models, and three deep learning models across multiple datasets. On the whole, statistical approaches were less effective, in contrast, machine learning techniques and LSTM-based neural networks exhibited enhanced outcomes as measured by the SMAPE metric. Methods employing LSTM with one-hot encoding and LSTM with embedding techniques achieved commendable performance regarding RMSE. Generally, the LSTM-Neural Network model surpassed alternative models in comparative analyses over diverse geographical units.

In another demand prediction study, hierarchical reconciliation approaches have been explored to enhance deep-learning methods, as demonstrated by Khalesian et al. (2024), who emphasized the necessity of incorporating error analysis to maintain forecast accuracy within a feasible solution space, underscoring the flexibility and applicability of these methods across various scenarios of area-based traffic demand prediction. Similarly, deep spatio-temporal ConvLSTM models, like the one proposed by Wang et al. (2018), have shown great promise in forecasting travel demand by considering comprehensive time and space factors, marking a significant leap from traditional time-series models that often fall short in complex scenarios.

In the domain of traffic flow prediction, the spatial-temporal convolutional model introduced by Fu et al. (2021) stands out for its application on urban crowd density prediction using mobile-phone signaling data, demonstrating the potential of deep learning models to handle irregular-shaped divisions and capture the intricate spatial and temporal dependencies inherent in traffic prediction tasks. This approach aligns with the findings of Yang et al. (2018a), who utilized deep learning to predict daily usage of Bike Sharing Systems (BSS), indicating a shift towards data-driven models that leverage deep learning for enhanced predictive capabilities.

Furthermore, the research by Li et al. (2023) into urban rail transit (URT) underscores the impact of spatial features on prediction accuracy, presenting a deep learning-based passenger flow prediction method that incorporates spatial characteristics such as land use, regional location, and intermodal access. This methodological shift towards acknowledging the role of spatial features in prediction models offers a nuanced understanding of travel behavior and demand, facilitating more accurate forecasts.

DeepSTCL, a deep spatio-temporal ConvLSTM framework for travel demand prediction developed by Wang et al. (2018), represents another step forward, treating historical travel data like a video stream to predict future demand. This innovative approach underscores the potential of deep learning in tapping into spatio-temporal features for travel demand forecasting, achieving high accuracy and efficiency. In conclusion, the integration of deep learning into travel demand and traffic flow prediction models has facilitated a deeper understanding of complex spatial and temporal dynamics, significantly enhancing predictive accuracy.

Recently, researchers have explored graph-theoretic approaches to model traffic data collected from road sensors. The spatial correlations between traffic sensors are structured as a directed graph, with nodes representing sensors and edge weights indicating proximity between sensor pairs measured by road network distance. Given that sensor networks are naturally organized as graphs, recent advances in graph neural networks (Defferrard et al., 2016), particularly graph convolution networks (Kipf & Welling, 2016), have inspired several graph-based traffic prediction models (Zhang et al., 2019; Yu et al., 2017; Guo et al., 2019; Pan et al., 2019; Diao et al., 2019). For example, (Bruna et al., 2013; Welling & Kipf, 2016) propose a method integrating graph convolutional neural network (GCN) into the LSTM, addressing both spatiotemporal dependencies and heterogeneity.

The study by Rajabzadeh et al. (2017) presents an interesting short-term traffic flow prediction approach. This research acknowledges that most prediction models face limitations in predicting local variation patterns and handling dynamic traffic due to their reliance on training data. To address this issue, they propose a two-step approach that combines a baseline model constructed from historical data with a time-varying Vasicek (EV) model better to capture real-time variations in traffic flow during each day. What makes this study particularly interesting is its use of residuals, which are the differences between actual and predicted values, as a feature to improve traffic flow prediction. While the study does not explicitly identify and evaluate the patterns of residuals, its objective is to enhance prediction performance by considering daily uncertainty impacts by applying the EV model. This approach hints that evaluating prediction residuals can provide valuable insights into prediction errors and the factors that trigger them.

Inspired by the studies discussed, we propose using the LSTM approach as our benchmark model. To account for the spatial adjacency often overlooked in conventional forecasting methods, we then incorporate a GCN within the LSTM framework. Adjacency, in the context of TAZs, refers to the geographical proximity between zones. Two zones are considered adjacent if they share a boundary or are close enough to influence each other's travel patterns. This relationship is captured through an adjacency matrix, which encodes the connections among zones.

This deliberate integration aims to evaluate the incremental improvements in prediction accuracy afforded by considering the adjacency of zones. The LSTM model establishes a solid baseline, enabling us to highlight the specific advantages of integrating spatial characteristics through the GCN. Our approach does not purport to introduce a novel modeling technique per se; rather, it illustrates the potential for enhancing prediction accuracy by integrating spatial adjacency with established predictive models. By applying this LSTM+GCN framework to the context of trip production prediction, our study underscores the significant role of spatial adjacency. While numerous studies have explored OD matrix prediction utilizing either spatial or temporal prediction models separately, our study advances this by integrating both spatial and temporal dimensions using a GCN combined with LSTM. This integration allows the model to simultaneously consider the spatial adjacency of TAZs and their temporal trip production characteristics, enhancing the accuracy and reliability of predictions over traditional models that consider these aspects in isolation.

Several recent studies have proposed novel models that integrate spatial adjacency into traffic prediction. For example, Zhao et al. (2019) proposed the Temporal Graph Convolutional Network (T-GCN), which combines GCN and gated recurrent units (GRU) for traffic forecasting. Similarly, Yu et al. (2017) introduced a spatio-temporal graph convolutional network (STGCN) for traffic forecasting. These models demonstrate the effectiveness of incorporating spatial adjacency and temporal dynamics.

Beyond T-GCN (Zhao et al., 2019) and STGCN (Yu et al., 2017), other approaches have introduced more complex frameworks to improve GCN-based traffic prediction. These include models that leverage dynamic graphs to capture evolving network topologies, transformer-based GCN architectures that integrate attention mechanisms, and hierarchical GCNs (HGCNs) that exploit spectral clustering of regions. Such methods can potentially capture evolving spatial-temporal dependencies and multi-level spatial structures, leading to further enhancements in prediction performance (Wang et al., 2024; Yan & Ma, 2021; Zhang et al., 2021). Our work builds upon this foundation by integrating GCN with LSTM to capture spatial-temporal correlations in trip production prediction.

4.1.2 Research Objectives and Contributions

Drawing on the existing literature on traffic data, this research delves into the emerging role of graph knowledge in the context of travel demand prediction. We propose a method that integrates GCN into the LSTM and aims to achieve three main objectives. First, we seek to determine the extent to which computation time and prediction accuracy are impacted by incorporating the adjacency of traffic analysis zones into the prediction model. Second, we explore the influence of spatial scale on spatial uncertainty (associated with spatial heterogeneity), as this aspect has remained relatively unexplored in the demand prediction field. We compare the prediction accuracy changes across multiple administrative spatial scales when using the same model.

Our choice to focus on the LSTM model and its integration with GCN (LSTM+GCN) was deliberate and aimed at highlighting the specific benefits of considering the adjacency of zones in forecasting models. The LSTM model served as a robust benchmark to establish a baseline for predictive performance. By integrating GCN, we sought to underscore how spatial characteristics could enhance prediction accuracy beyond the baseline. This approach was not intended to claim broad novelty in modeling techniques but rather to demonstrate the incremental gains in prediction accuracy that can be achieved by incorporating spatial adjacency into well-established models. Therefore, our study contributes to the field by applying the LSTM+GCN framework to the specific context of trip production prediction, emphasizing the role of spatial adjacency.

The integration of GCN with LSTM in a unified framework allows for the simultaneous modeling of both spatial relationships and temporal dependencies. This dual capability is particularly advantageous in our context for several reasons. By capturing the spatial adjacency of TAZs through GCNs and the temporal trends through LSTMs, the model can make more informed, context-aware predictions than would be possible by considering either aspect in isolation. This approach allows the model to utilize the full spectrum of available data—spatial configurations and temporal sequences—thereby maximizing the insights gained from the data and improving prediction robustness. Incorporating both spatial and temporal data reflects real-world conditions more accurately, making the model's outputs more reliable and applicable for planning and operational purposes.

Our proposed method offers several key contributions to the field of travel demand prediction:

1. By employing a nationwide graph to integrate spatial adjacency into an LSTM-based model, our approach simultaneously enables large-scale, national-level trip production predictions and forgoes the need to train multiple separate models for each region or scale. This unified framework not only broadens the spatial scope of demand predictions but also potentially streamlines computational processes and resource allocation.

2. Our model's incorporation of spatial adjacency information directly addresses the spatial heterogeneity inherent in TAZs, thereby refining the accuracy of trip production forecasts.

3. Through a nuanced examination of how spatial scale influences the predictability and uncertainty of trip production, our study sheds light on crucial considerations for transport modelers and policymakers, particularly regarding the implications of spatial resolution on demand forecasts.

In this study, we use processed aggregated derivatives of GSM traces in the form of motor-vehicles OD matrices of the Netherlands in March 2017. This implies the scope of this work pertains to motor-vehicle trip production patterns.

The remainder of this chapter is structured as follows: Section 4.2 delineates the research data and the methodology employed. Section 4.3 presents and discusses the developed trip production prediction models, elucidates the impact of spatial scale on spatial uncertainty and prediction accuracy, investigates prediction errors and residual patterns, and identifies the most pertinent socio-spatial features contributing to residual patterns across various spatial scales. Finally, Section 4.4 offers a conclusion, summarizing the key findings and their implications.

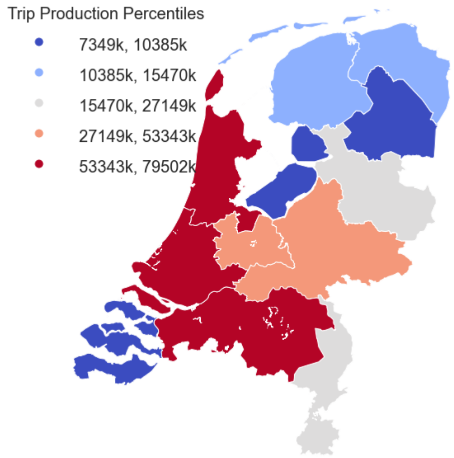
4.2 Methodology

4.2.1 Trip Production and Socio-Spatial Data

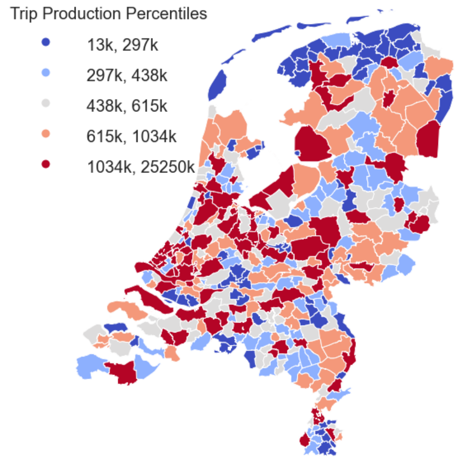
This study utilizes hourly trip production data for the entire Netherlands during March 2017. The data is aggregated over three spatial scales: *provinces*, *municipalities*, and *4-digit postal code zones*, resulting in 12, 390, and 1243 Traffic Analysis Zones (TAZs) throughout the Netherlands.

Trip production for TAZ i refers to the number of inter-zonal motor-vehicle trips originating from i. Trip production values are derived from the GSM traces of Dutch telecommunications company Vodafone, which holds a market share of approximately one-third of the Dutch population. Due to privacy concerns related to raw mobile phone data, another company processed the data. As a result, this study utilizes origin-destination (OD) matrices of motor vehicles based on TAZs in the Netherlands rather than mobile phone traces. These OD matrices have been scaled up to account for the entire Dutch population. For more details on the scaling procedure, refer to (Meppelink et al., 2020).

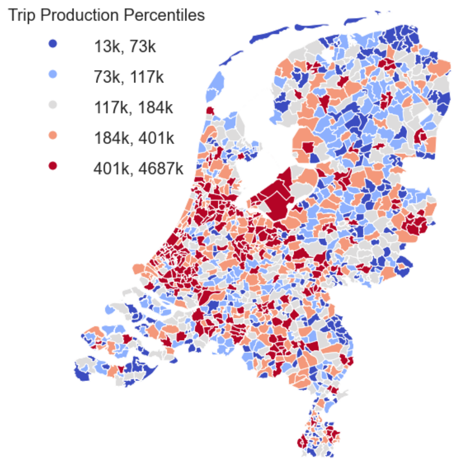
These OD matrices are pre-processed and reshaped to derive a vector of hourly trip production values per TAZ for each of the three spatial scales studied here in this research. The dataset contains over 365 million produced trips. Figure 4.1 displays the total monthly trip production per TAZ for the three spatial scales under study.



(a) Spatial scale of *provinces*.



(b) Spatial scale of *municipalities*.



(c) Spatial scale of *4-digit postal codes*.

Figure 4.1: Total monthly trip production per TAZ for the three studied spatial scales.

It is important to note that the trip production data is derived from GSM traces of Vodafone users, representing approximately one-third of the Dutch population. While this provides substantial coverage, there is a potential risk of user selection bias or over-representation of travel patterns from specific demographic groups. The data may not fully capture the travel behavior of non-Vodafone users or certain socioeconomic segments, such as elderly populations or individuals with limited access to mobile phones. This limitation is acknowledged, and the results should be interpreted with consideration of this potential bias.

The Central Bureau of Statistics (CBS) of the Netherlands provided socio-spatial data containing approximately 140 demographic and land-use variables for each TAZ (Centraal Bureau voor de Statistiek (CBS), 2017) at each of the three aforementioned spatial scales. CBS collects, edits, and publishes national statistics based on registers, surveys, and interviews. The variables used in our analysis correspond to data from 2017, aligning with the trip production data.

4.2.2 Prediction of Trip Production with LSTM

The trip production data exhibit strong seasonality effects and dynamic trends within recent time frames. Unlike a regular RNN, which struggles to establish long-term dependencies, the LSTM overcomes the vanishing gradient problem, enabling it to capture both short- and long-term temporal patterns in a time series. Consequently, we employ an LSTM to investigate the periodic (i.e., daily) dependencies and recent dynamic trends in trip production. We refer readers to (Hochreiter & Schmidhuber, 1997) for more details on the LSTM architecture.

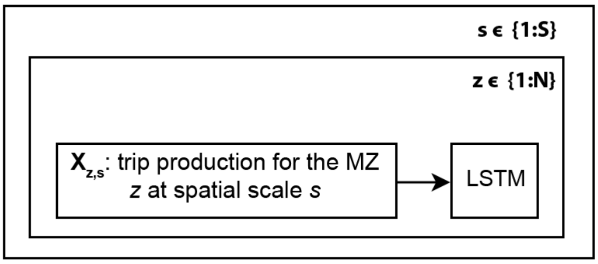


Figure 4.2: Experimental framework for the trip production using LSTM.

Figure 4.2 illustrates the experimental framework for predicting trip production using LSTM. The model takes as input the time series signals $X_{i,s}$ representing the trip production data for each TAZ i within the spatial scale s. The data are then fed into the LSTM network, which consists of multiple hidden layers. Each LSTM layer is composed of three gates: input, forget, and output gates. The gates function as follows:

- **Input Gate**: Determines which input information is allowed into the memory cell.
- **Forget Gate**: Controls which historical data should be retained in memory.
- **Output Gate**: Produces the final output of the memory cell and contributes a portion of the input information for the subsequent cell.

The LSTM network learns the temporal patterns in trip production data, capturing both short-term and long-term dependencies through these gate mechanisms. The output of the LSTM layer is passed to a fully connected layer, which predicts future trip production values based on the learned temporal features. In this framework, the trip production signal $X_{i,s}$ is normalized to ensure comparability across different spatial scales and TAZs, allowing the model to generalize better across regions.

4.2.3 Prediction of Trip Production with LSTM+GCN

Considering spatial correlation and heterogeneity between TAZs potentially improves trip production models. To address this, we integrate a Graph Convolutional Network (GCN) into the LSTM. We refer readers to (Bruna et al., 2013; Seo et al., 2018) for more details. A GCN is a spatial feature extraction model applicable to any topological structure graph. To compute the GCN at time t for an M-feature matrix $X_t \in \mathbb{R}^{N\times M}$, we first generate an undirected graph $G = (Z, E, A)$, where the nodes Z represent the TAZs, and $|Z|$ equals the number of TAZs, N, in each spatial scale.

Considering spatial correlation and heterogeneity between TAZs potentially improves trip production models. To address this, we integrate a Graph Convolutional Network (GCN) into the LSTM. We refer readers to (Bruna et al., 2013; Seo et al., 2018) for more details. A GCN is a spatial feature extraction model applicable to any topological structure graph. To compute the GCN at time t for an M-feature matrix $X_t \in R^{N\times M}$, we first generate an undirected graph $G = (Z, E, A)$, where the nodes Z represent the TAZs, and $|Z|$ equals the number of TAZs, N, in each spatial scale.

Nodes (Z): Each node represents a TAZ at a specific spatial scale. The total number of nodes $|Z| = N$ corresponds to the number of TAZs at that scale.

Edges (E): The edges represent the connections between neighboring TAZs. Two TAZs are considered neighbors if they share a common boundary. We use shapefiles of administrative boundaries and a spatial join operation to determine the neighbors.

The adjacency matrix $A \in \mathbb{R}^{N\times N}$ encodes the spatial relationships between TAZs, with A_{ij} defined as:

$$A_{ij} = \begin{cases} 1 & \text{if TAZs } i \text{ and } j \text{ are adjacent (i.e., share a common boundary),} \\ 0 & \text{otherwise.} \end{cases} \tag{4.1}$$

The GCN layer operates on the graph structure defined by the adjacency matrix A, aggregating information from neighboring TAZs to learn spatial features. The graph convolution operation is defined as:

$$H = \sigma(\tilde{A}XW), \tag{4.2}$$

where $\tilde{A} = D^{-1/2}(A+I)D^{-1/2}$ is the normalized adjacency matrix with added self-connections, D is the degree matrix of $A+I$, I is the identity matrix, X is the input feature matrix (trip production data), W is a learnable weight matrix, and σ is an activation function such as ReLU. This operation allows each TAZ to aggregate information from its neighbors, effectively capturing spatial dependencies.

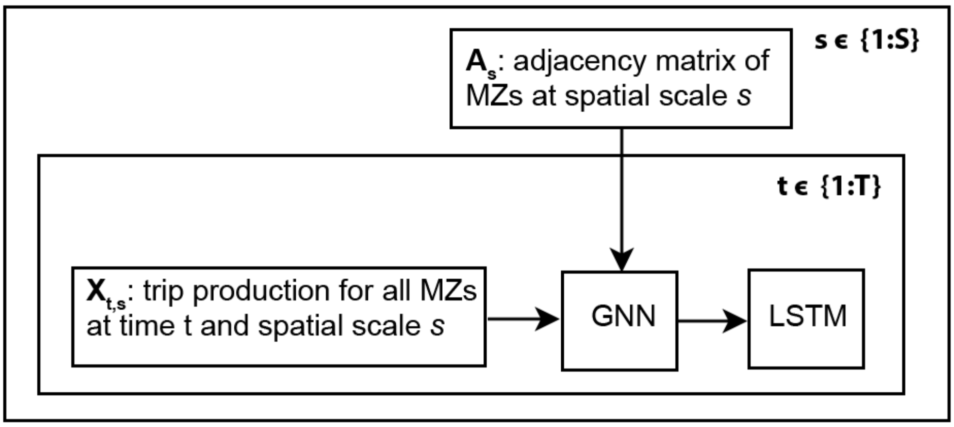


Figure 4.3: Experimental framework for the trip production using GCN+LSTM.

Figure 4.3 provides a detailed schematic diagram of the proposed model framework, illustrating the integration of GCN and LSTM layers for capturing spatial and temporal correlations.

We use the graph convolution operation to extract spatial features of the mobility network and the LSTM to extract temporal features of the signal. The LSTM input consists of the convolutional graph features concatenated with the original signals. In other words, for each spatial scale, the graph signals $X_{t,s}$ at time t, along with the adjacency matrix A_s, are used to compute the spatial features $H_{t,s}$.

As depicted in Figure 4.3, the spatial features $H_{t,s}$ obtained from the GCN layer are concatenated with the original graph signals $X_{t,s}$ to form the input $X_{t,s}^{GCN}$ for the LSTM layer. This concatenation ensures that both the spatial information from neighboring TAZs and the temporal information from the original signals are jointly utilized by the LSTM to make predictions. Therefore, the input to the LSTM layer is:

$$X_{t,s}^{GCN} = [X_{t,s}; H_{t,s}],$$

where $[\cdot;\cdot]$ denotes concatenation along the feature dimension.

Lastly, the LSTM output serves as the input for a fully connected layer that predicts the signal for the desired time. The process can be summarized as follows:

- **Spatial Feature Extraction:** Apply the GCN layer to learn spatial features ($H_{t,s}$) from the graph signals ($X_{t,s}$) and adjacency matrix (A_s).
- **Temporal Feature Extraction:** Concatenate the spatial features ($H_{t,s}$) with the original graph signals ($X_{t,s}$) to form the input ($X_{t,s}^{GCN}$) for the LSTM layer.
- **Trip Production Prediction:** Use the LSTM layer to predict future trip production values.

For further details on the architecture and implementation of our model, including the code and data used, we have published our code repository and dataset. The code repository can be found at Eftekhar et al. (2024a), and the dataset is available at Eftekhar et al. (2024b).

Thus, for each spatial scale, the graph signals ($X_{t,s}$) at time (t) are concatenated with the spatial features ($H_{t,s}$) to form the input for the LSTM ($X_{t,s}^{GCN}$). The LSTM output then serves as the input for a fully connected layer that predicts the signal for the desired time.

4.2.4 Experimental Design

We initially normalize the values for each TAZ to enable a fast and stable pattern comparison across various TAZs. For instance, Figure 4.4 shows the 2-D histogram of trip production for all the *provinces* in the Netherlands. As shown, the value ranges for different TAZs vary significantly. We can focus more on recognizing the patterns by normalizing the values in each TAZ. We later reverse the values back to their original range to evaluate prediction accuracy. We applied the Min-Max Scaling technique to normalize each production value x, i.e.,

$$x_{normalized} = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{4.3}$$

where $x_{normalized}$ is the normalized value, x_{min} and x_{max} are the minimum and maximum production values in the (month-long) time series for that particular TAZ. Consequently, the resulting normalized values range between 0 and 1. We then use the two proposed methods, LSTM and LSTM+GCN, to predict trip productions in each spatial scale. Table 4.1 provides a summary of the model parameters and hyperparameters used in this study.

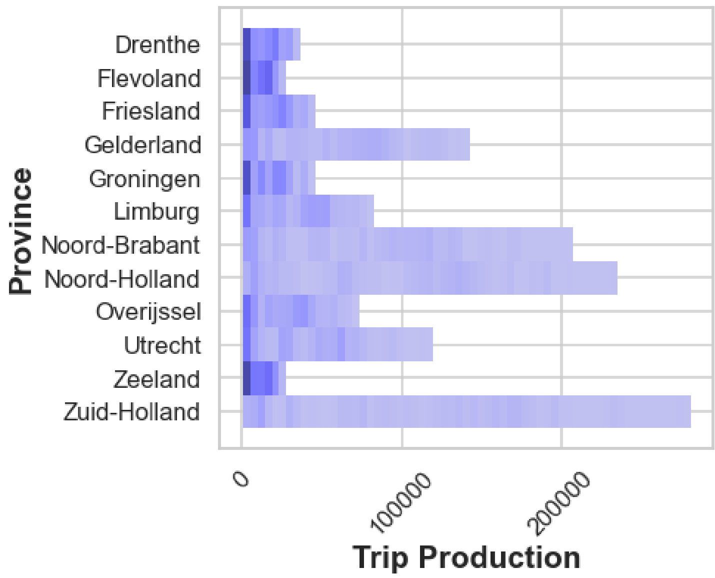


Figure 4.4: Trip production values among provinces during March 2017.

We employ two evaluation metrics at each prediction step to assess our model's performance: Mean Square Error (MSE) and Mean Absolute Percentage Error (MAPE). Moreover, we compare the accuracy of the two proposed models across the studied spatial scales using the same metrics. Furthermore, we report the computation time for each model to compare the required computation capacity.

In our analysis, the selection of Mean Squared Error (MSE) and Mean Absolute Percentage Error (MAPE) as pivotal metrics was guided by their respective abilities to elucidate distinct aspects of prediction. By harnessing both MSE and MAPE, our study captures a holistic view of the prediction models' performance—highlighting not just the magnitude of prediction errors but also their proportional significance relative to actual trip patterns.

MSE was chosen for its capacity to underscore and amplify the impact of larger errors in prediction. By squaring the differences between the predicted and the actual values, MSE penalizes more significant errors, thus sensitizing us to models that may be prone to occasional but substantial inaccuracies. This metric is especially critical in the prediction of trip production, where outsized errors can denote critical lapses in predicting peak travel demands, a vital consideration for transportation system planning and management.

MAPE, on the other hand, offers a normative perspective by expressing prediction errors as a percentage of the actual values. This normalization allows for a more intuitive comprehension of the model's accuracy, independent of the scale of the data. MAPE's percentage-based nature renders it particularly insightful for comparative analyses across regions or time periods with varying levels of trip production. It enables us to discern the relative predictive accuracy in a manner that is agnostic to the actual trip production magnitude, thereby facilitating equitable benchmarking across diverse scenarios. Please note that MSE is also used as the training loss function for both models. MAPE is not a training loss but an evaluation metric used to compare final prediction performance.

We conducted manual hyperparameter tuning to determine the optimal configurations for the GCN and LSTM models. This involved testing various combinations of the number of layers (1 to 3), hidden units (64, 128, 256), learning rates (0.001, 0.005, 0.01), and dropout rates (0.2, 0.5). While an exhaustive search was not possible due to computational limitations,

Table 4.1: LSTM and GCN+LSTM Hyperparameters

	LSTM	**GCN+LSTM**
GC layers	na	2
GC layers sizes	na	16, 10
GC activations	na	relu
GC dropout	na	0.5
GC optimizer	na	Adam
GC learning rate	na	0.01
GC weight decay	na	5e-4
LSTM layers	3	3
LSTM layers sizes	128, 256, 128	128, 256, 128
LSTM activations	relu	relu
LSTM dropout	0.2	0.2
LSTM optimizer	Adam	Adam
LSTM learning rate	0.001	0.001
Fully Connected Layer	Linear	Linear
Batch size	128	128
Number of epochs	100	100
Early stopping patience	10 epochs	10 epochs
Loss function	MSE	MSE

we selected the configurations that yielded the best validation performance, balancing model complexity and accuracy. Table 4.1 displays the settings we used for the implemented models in this study.

The experiments were conducted on a workstation with an Intel Core i7-9700K CPU @ 3.60GHz, 32 GB RAM, and an NVIDIA GeForce RTX 2080 Ti GPU. The models were implemented using Python 3.8 and PyTorch 1.7.1. All computations were performed locally, and the specifications are provided to ensure transparency and reproducibility.

4.3 Results and Discussion

In this section, we first present an initial exploration of trip production data and the predictions generated by the LSTM and LSTM+GCN models. In the second part of this section, we analyze the prediction results of the LSTM+GCN model in more detail. Here, we aim to identify residual patterns at each spatial scale and associate them with the socio-spatial features of each TAZ within the respective spatial scale.

4.3.1 Initial Exploration and Predictions: LSTM vs. LSTM+GCN Models

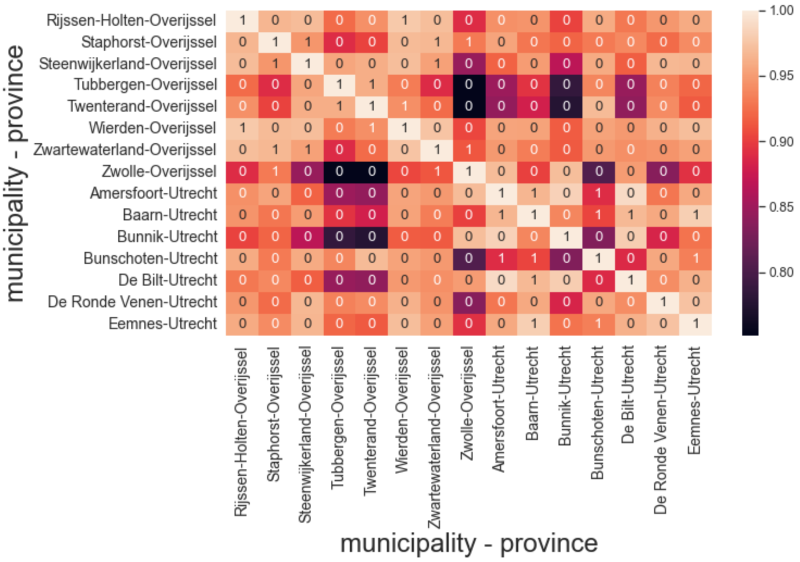
In order to investigate whether adjacent zones have similar trip production patterns, a correlation matrix is presented in Figure 4.5. Specifically, the matrix examines the correlation between several municipalities (Figure 4.5(a)) and several 4-digit postal code zones (Figure4.5(b)). Each cell in the matrix contains a value of either 1 or 0, indicating whether the two associated zones are adjacent or not. Despite the observation that lower correlated TAZs are primarily non-adjacent, the data indicates no significant linear correlation between adjacent zones, as evidenced by the accompanying figures. However, a conditional non-linear correlation may exist between adjacent TAZs, which the linear correlation matrix fails to capture. To overcome this limitation, non-linear models, such as deep neural networks, are employed in this study to model complex behaviors that may exist between adjacent zones.

While the linear correlation in a narrow 0.75–1.00 band (presented in Figure 4.5) indicates modest absolute differences, the figure still confirms that the lowest correlations predominantly occur between non-adjacent TAZ pairs. Importantly, the near-uniform linear range justifies our shift to non-linear models: weak linear separation implies that any meaningful adjacency effect must be captured through higher-order interactions, a task for which GCN layers are well-suited.

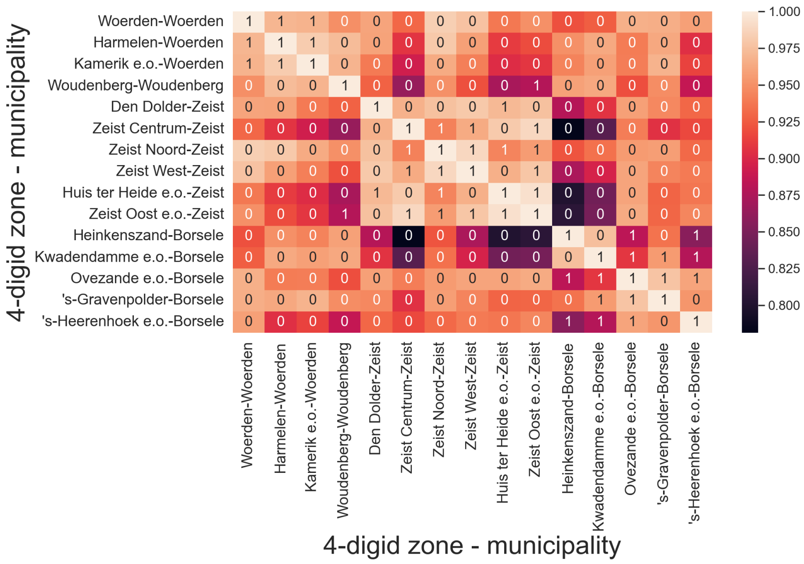
In our modeling process, the first step was to pre-process the data to conform to the required format for further analysis. This step acquired a normalized trip production vector for each TAZ within one of the three studied spatial scales. We then applied the LSTM and LSTM+GCN models and evaluated the LSTM model's predictions based on its MAPE evaluation metric. Figure 4.6 presents the worst TAZs in the three spatial scales under study, along with the LSTM+GCN model's predictions for those TAZs. Including adjacency data in the LSTM+GCN model seems to improve the accuracy of predictions, particularly for TAZs where the LSTM model struggled to capture peak trip production values. LSTM seems to produce higher errors in predicting regular peaks, especially at the *municipality* spatial scale, and might need larger training/validation dataset. LSTM+GCN, on the other hand, seems to perform more accurately in predicting the daily peaks with the same amount of training data.

Although we observe improvements in the LSTM+GCN model's accuracy, the latent factors contributing to these results are not immediately apparent. Further research into these models and the properties of each TAZ in the network is necessary to gain insights into these factors. To comprehensively evaluate the results of our two models, we examined the distribution of the two evaluation metrics across the under-study scales, as shown in Figures 4.8 and 4.9. These figures indicate that the LSTM+GCN model's average MAPE and MSE of predictions are lower than those of the LSTM model across all three scales under study.

Table 4.2 presents the summary of prediction metrics of MAPE and MSE across different spatial scales using LSTM and LSTM+GCN model. The empirical results presented in this table engender a thought-provoking dialogue about the expected trends in predictive performance across varying spatial resolutions. In an intuitive sense, the forecast accuracy is anticipated to deteriorate with the increase in spatial granularity due to the amplified noise and intrinsic variability within the travel patterns. However, the LSTM model's performance metrics indicate an



(a) Correlation matrix between municipalities.



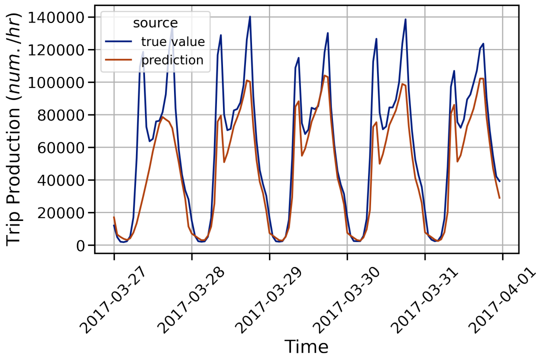
(b) Correlation matrix between 4-digit postal code zones.

Figure 4.5: Correlation of trip production between TAZs under different spatial scales.

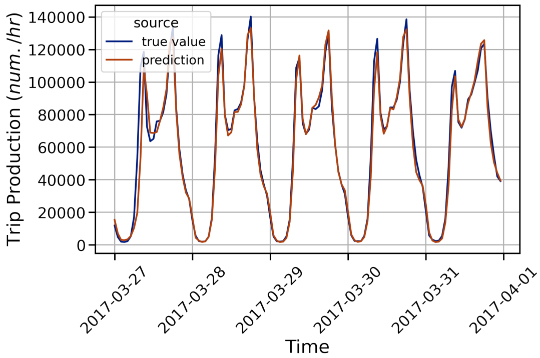
Model	Metric	Province (Avg ± Std)	Municipality (Avg ± Std)	4-digit Zones (Avg ± Std)
LSTM	MAPE	23.0% ± 3.4%	2.6e+14% ± 5.2e+15%	1.3e+14% ± 3.5e+15%
	MSE	2.4e+08 ± 3.6e+08	5.5e+05 ± 3.5e+06	5.4e+04 ± 3.7e+05
LSTM+GCN	MAPE	12.0% ± 1.9%	8.7e+11% ± 1.7e+13%	1.6e+13% ± 5.6e+14%
	MSE	5.0e+07 ± 6.7e+07	1.4e+05 ± 1.4e+06	1.5e+04 ± 9.2e+04

Table 4.2: Summary of prediction metrics across different spatial scales using LSTM and LSTM+GCN models, with MAPE expressed in percentage units.

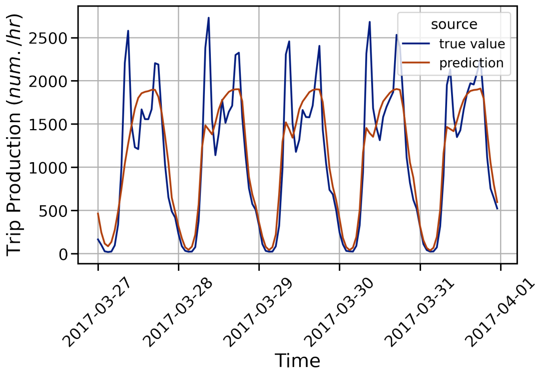
anomalously higher error for the municipality scale compared to the 4-digit postal code zones. This inversion of expected error magnitude necessitates a closer examination of underlying dy-



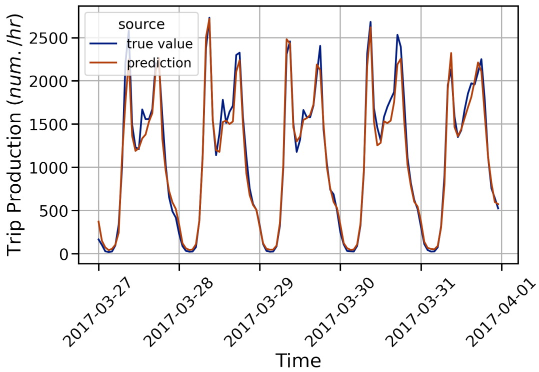
(a) ***Province*** prediction (*Gelderland*) with LSTM model.



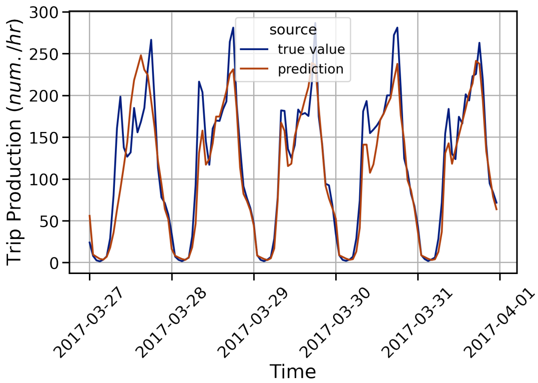
(b) ***Province*** prediction (*Gelderland*) with LSTM+GCN model.



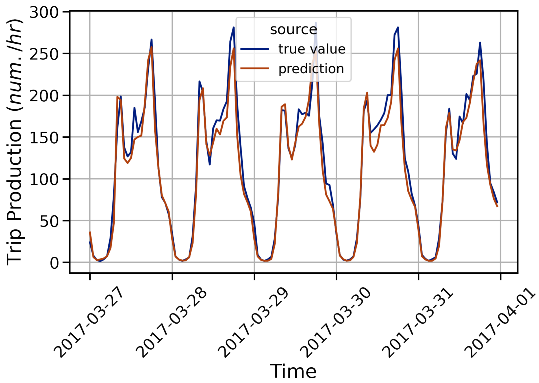
(c) ***Municipality*** prediction (*Rhenen*) with LSTM model.



(d) ***Municipality*** prediction (*Rhenen*) with LSTM+GCN model.



(e) ***4-digit postal code zone*** prediction (*De Steeg e.o.*) with LSTM model.



(f) ***4-digit postal code zone*** prediction (*De Steeg e.o.*) with LSTM+GCN model.

Figure 4.6: The worst predictions of trip production based on the lowest MAPEs in the network among the three spatial scales using LSTM (on the left side) and their associated prediction using LSTM+GCN (on the right side).

namics.

One plausible explanation for this counterintuitive finding could be the heterogeneity of socio-spatial characteristics within municipalities. The larger standard deviation in MAPE and MSE at the municipality level suggests a wider variability in prediction performance, which could stem from diverse commuting behaviors, land-use configurations, and transportation network complexities within these larger regions. Municipalities often encapsulate a mix of urban, suburban, and possibly rural settings, each with distinct travel demand patterns that could potentially confound the LSTM model, which does not explicitly account for spatial adjacency.

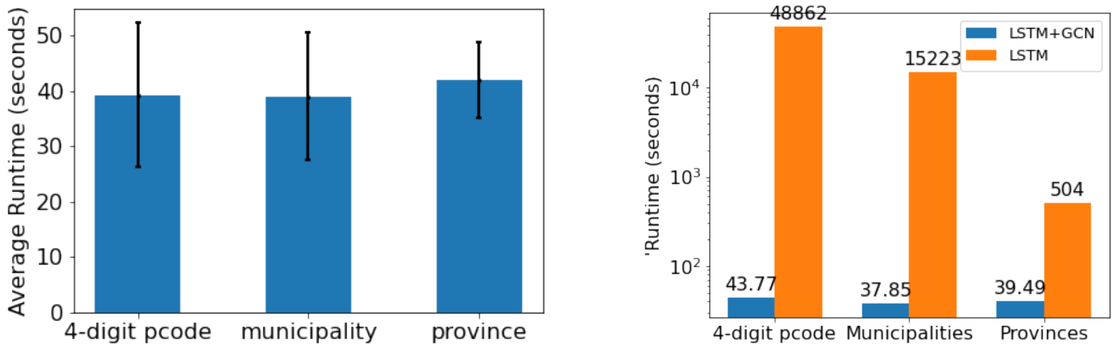
Conversely, the LSTM+GCN model, which integrates spatial adjacency into its predictive framework, exhibits a substantial improvement in MAPE for municipalities, although the MSE remains higher than that for the 4-digit zones. The improved MAPE implies that when spatial relationships are considered, the model becomes more adept at capturing the proportionate variances in trip production, especially in the context of municipalities where spatial adjacency plays a significant role. The persistence of higher MSE, despite a lower MAPE, might suggest that while the model is generally accurate, it is occasionally susceptible to larger errors—potentially from extreme values or outliers that are more pronounced in municipal data.

These results underscore the merit of integrating spatial adjacency to enhance predictive accuracy, especially where the spatial structure itself may be a pivotal determinant of travel patterns. The LSTM+GCN model's ability to leverage such spatial correlations ostensibly attenuates the prediction difficulty at higher spatial scales. Nonetheless, the nuanced nature of trip production across different spatial scales reaffirms the necessity for tailored-modeling approaches that can accommodate the unique features of each granularity level.

As we move from the *province* scale to lower levels of abstraction, the prediction accuracy declines, as indicated by increasing MAPE values. This trend suggests that prediction accuracy is lower at lower levels of abstraction. Nonetheless, it appears that accuracy improves at higher levels of abstraction, despite the increase in spatial uncertainty. This outcome may be because the aggregation of trip production patterns makes them more regular and hence more predictable. The worst predictions for both models are presented in Figure 4.6 and support our observation that the LSTM+GCN model predicts peak trip production values more accurately than the LSTM model. Furthermore, the worst predictions demonstrate that trip production at the *province* scale is more accurately predicted than at higher resolution scales. This outcome may be due to pattern aggregation and resolution lowering.

Our study's results indicate that incorporating adjacency data into the LSTM+GCN model enhances the accuracy of extreme value predictions while also significantly decreasing computation time. Figure 4.7(a) shows that the computation time distribution for the LSTM model has an average of 40 seconds per TAZ, with a more extensive interquartile range for 4-digit postal code zones, suggesting a more diverse range of zones. Please note that "per-TAZ" run time is well defined only for the disaggregate-LSTM set-up. As the LSTM+GCN model is fitted once on the whole graph, every zone shares the *same* training pass; presenting that as a box-plot would reduce to a single line.

The total computation times for the LSTM and LSTM+GCN models are illustrated in Figure 4.7(b). We noticed that the LSTM model's run time depends on the number of TAZs in the network, whereas the LSTM+GCN model's computation time does not fluctuate significantly with the number of TAZs. This is because the LSTM+GCN model is only trained once, rather than training one model for each TAZ in each spatial resolution. Therefore, changes in run time under different spatial scales are not substantial.



(a) Run time per TAZ of the LSTM model for the three spatial scales.

(b) Total run time of the LSTM and LSTM+GCN model for the three spatial scales.

Figure 4.7: Comparative run times of LSTM models.

Considering an alternative approach, one could estimate a single LSTM model for all TAZs without incorporating adjacency data or using the GCN model. This method would reduce computation time since it would no longer scale with the number of TAZs. However, the impact on prediction accuracy might differ between the disaggregate LSTM and aggregate LSTM+GCN models. Lacking the spatial relationships between TAZs, the single LSTM model might experience a decrease in prediction accuracy, as it would not capture the spatial heterogeneity and correlations among TAZs that influence trip production patterns. Regarding prediction accuracy, the single LSTM model without adjacency data would likely fall between the disaggregate LSTM models, which can capture the unique characteristics of each TAZ, and the aggregate LSTM+GCN model, which benefits from the inclusion of adjacency data to capture spatial relationships better. Thus, while using a single LSTM model for all TAZs without incorporating adjacency data offers the advantage of reduced computation time, it may involve a trade-off in prediction accuracy compared to the other approaches.

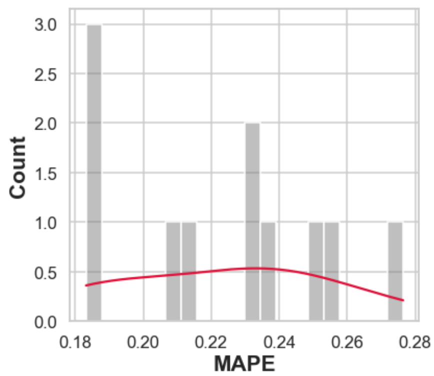
Altogether, our findings imply that incorporating adjacency data can improve prediction accuracy while decreasing computation time. Consequently, we analyzed and explored the predictions made using LSTM+GCN for the remainder of our study.

It is important to note that our empirical observations indicate that using a single model trained on the entire network may be more convenient than training separate models for each TAZ. However, we have not performed a formal computational complexity analysis or extensive runtime evaluations, and thus any conclusions about computation time savings remain preliminary.

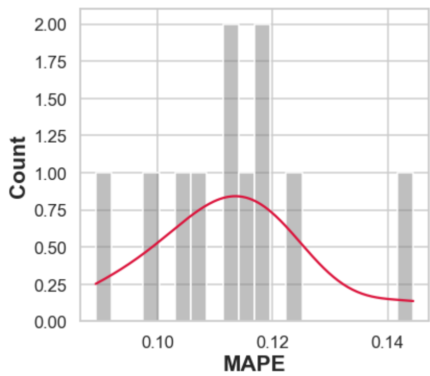
4.3.2 In-Depth Analysis of LSTM+GCN Prediction Results: Residual Patterns and Socio-Spatial Features

In Figure 4.10, we present the trip production and prediction results for Vlieland, a northern island that is both a municipality and a 4-digit zone. It is worth noting that Vlieland holds the highest MAPE among all 4-digit zones and municipalities, indicating a considerable challenge in accurately forecasting trip production in this region.

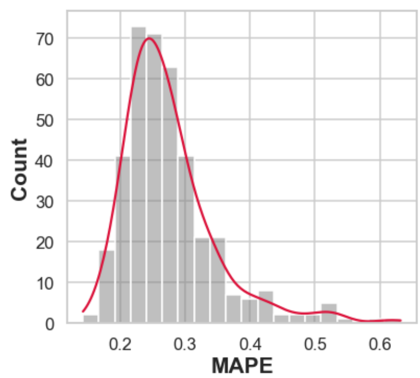
The observed trip production pattern of Vlieland in Figure 4.10(a) is characterized by an irregular profile with an extreme peak in the first half of the month, followed by scattered high



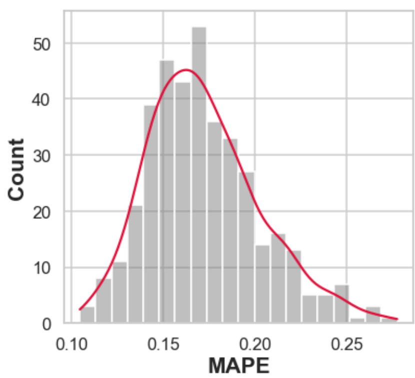
(a) LSTM, **provinces**.



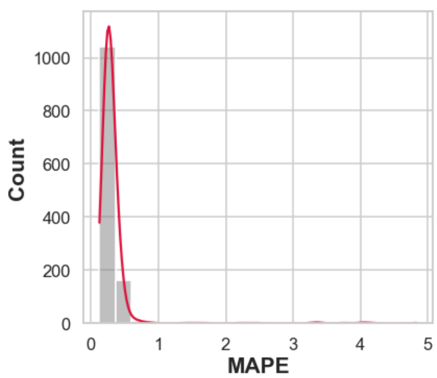
(b) LSTM+GCN, **provinces**.



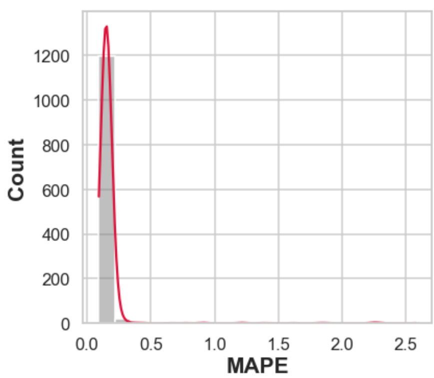
(c) LSTM, **municipalities**.



(d) LSTM+GCN, **municipalities**.

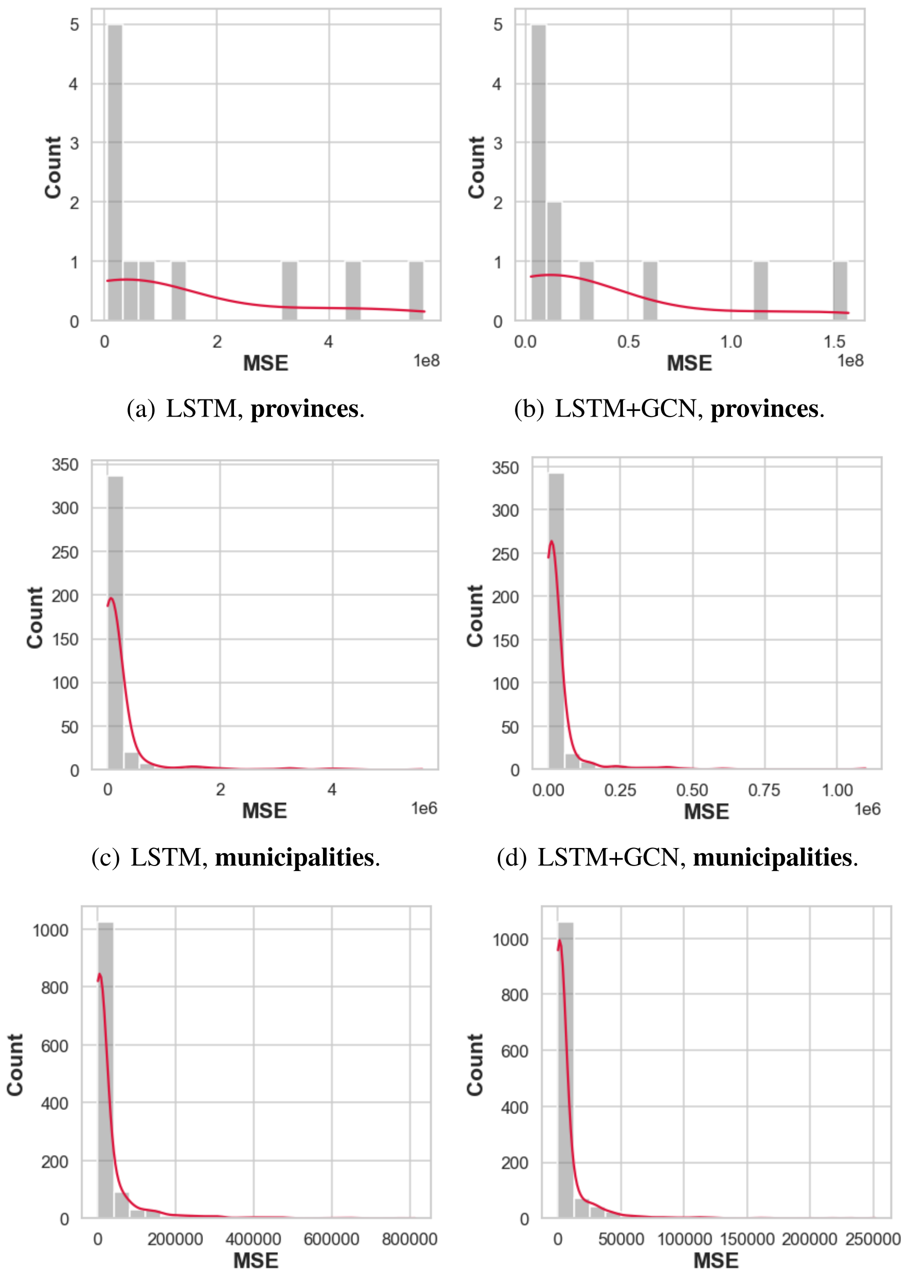


(e) LSTM, **4-digit postal codes**.



(f) LSTM+GCN, **4-digit postal codes**.

Figure 4.8: MAPE for actual versus predicted trip production values per TAZ for each of the three spatial scales, using LSTM (left) and LSTM+GCN (right) model.



(a) LSTM, **provinces**. (b) LSTM+GCN, **provinces**.

(c) LSTM, **municipalities**. (d) LSTM+GCN, **municipalities**.

(e) LSTM, **4-digit postal codes**. (f) LSTM+GCN, **4-digit postal codes**.

Figure 4.9: MSE for actual versus predicted trip production values per TAZ for each of the three spatial scales, using LSTM (left) and LSTM+GCN (right) model.

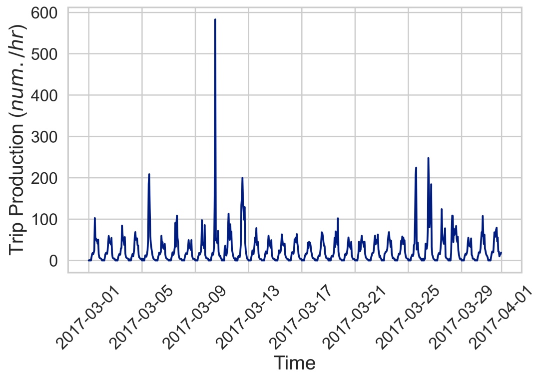
production in the last days of the month. This erratic pattern might be attributed to special events associated with this TAZ. Figure 4.10(b) and 4.10(c) present the trip production and prediction at the municipality and 4-digit zone levels, respectively. Accordingly, the prediction performance of the models varies across different spatial scales. Although the two models share similar hyper-parameters, they are trained on different training sets due to aggregation at a higher level of abstraction. The municipality-level model seems to capture the time series with more variation in the values, implying better predictions of production values influenced by seasonality. Conversely, the 4-digit model appears to be more biased towards predicting values closer to the average production line, hence less sensitive to extreme values in the production but better at predicting the off-peak values. These observations suggest the importance of considering the spatial scale in trip production and prediction analysis, as it can significantly affect the accuracy of the results.

Figure 4.11 presents the average hourly trip production and the prediction residual for various provinces in the Netherlands. This figure depicts these two variables, with Figure 4.11(a) showing the average hourly trip production and Figure 4.11(b) displaying the average hourly trip production prediction residual. The figure highlights that provinces such as Noord-Holland, Zuid-Holland, and Noord-Brabant, characterized by high population density and heavy industrial and commercial areas, have higher trip production values with high variability. This variability is shown in the shaded areas between the 10th and 90th percentile of the values. Figure 4.11(b) implies that residual peaks occur around the peak hours of trip production, although with somewhat different patterns. For instance, Zuid-Holland has the highest average hourly trip production during the afternoon rush hour, while the residual peak happens during the morning peak, which has the highest variation in trip production. These observations suggest that the prediction error is correlated with the variation in trip production, which holds implications for improving the accuracy of transportation demand models. However, further research is needed to explore these correlations and their underlying causes in more detail.

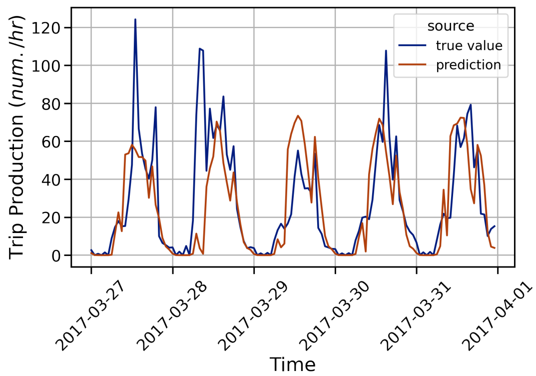
To investigate the correlation between trip production prediction error and variation in trip production among the 4-digit zones, Figure 4.12(a) plots the Mean Squared Error (MSE) against variance. The figure reveals a reasonably linear correlation between the two variables, indicating that the prediction error is positively correlated with trip production variation. However, a few outliers with high MSE and relatively low variance suggest high prediction errors despite low variation in trip production. Figure 4.12 highlights these TAZs in red, pointing to the zones with a high prediction error despite low variation in trip production. Interestingly, all the islands are among the outliers.

The robustness of the proposed LSTM+GCN model is crucial given the diversity of urban and rural dynamics in different regions. Our model was trained and evaluated on regions with varied geographical and traffic conditions, specifically at multiple spatial scales including provinces, municipalities, and 4-digit zones. The model consistently demonstrated its ability to capture spatial-temporal dynamics effectively, achieving high predictive accuracy across urbanized and rural regions.

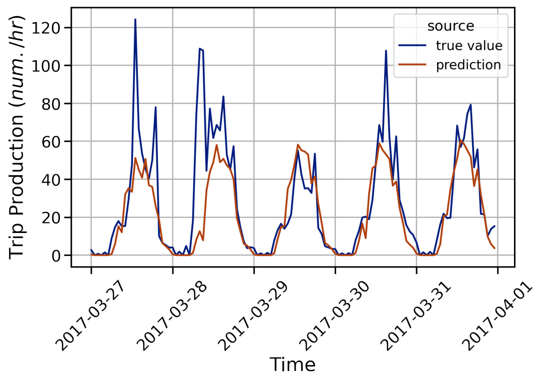
Despite the expected variations in traffic patterns between densely populated urban regions and sparsely populated rural regions, the model achieved consistent predictive accuracy in all scenarios. This was particularly notable during rush hour periods and off-peak periods, as well as across weekday and weekend traffic. Leveraging the complementary strengths of both GCN and LSTM, the model could effectively capture spatial and temporal correlations even in regions with varied urban dynamics.



(a) Actual trip production of Vlieland throughout March 2017.

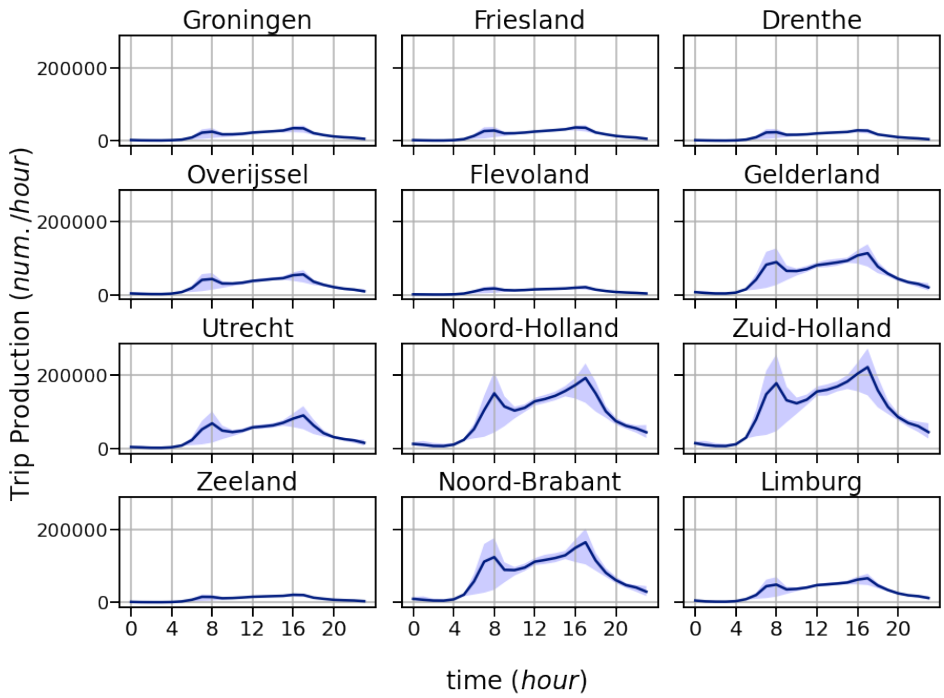


(b) Prediction of trip production for Vlieland at the **municipality** level.

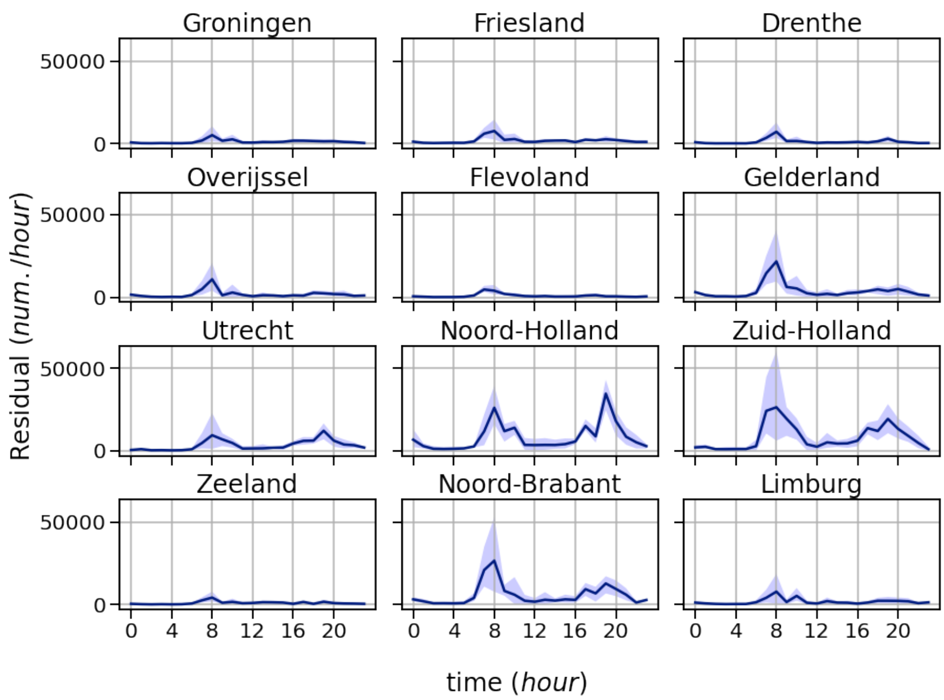


(c) Prediction of trip production for Vlieland at the **4-digit postal code** level.

Figure 4.10: Vlieland trip production prediction under two spatial scales throughout the last working week of March 2017.



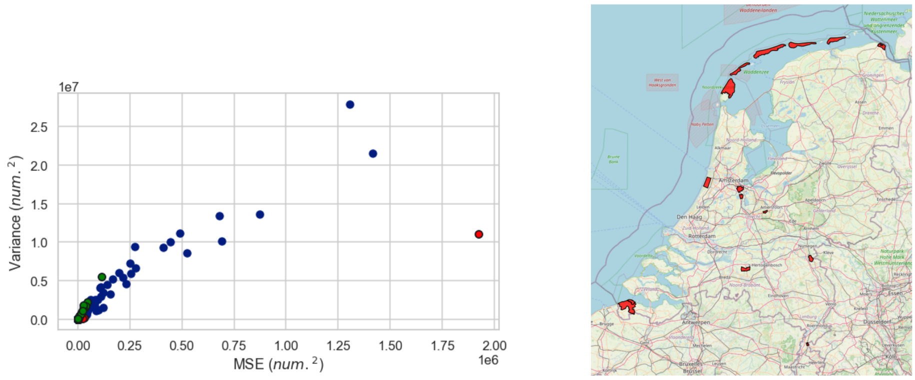
(a) Trip production.



(b) Prediction residual.

Figure 4.11: Average hourly trip production and prediction residual time series among the provinces in The Netherlands.

These results highlight the robustness and generalizability of the LSTM+GCN model, making it suitable for trip production prediction across diverse geographical areas and traffic conditions.

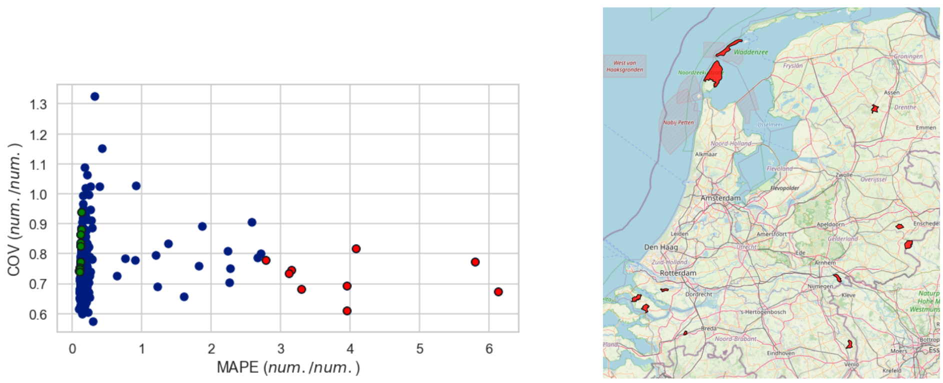


(a) MSE of trip production prediction vs. the variance of trip production.

(b) Zones with the highest MSE/variance ratio.

Figure 4.12: MSE of trip production prediction vs. the variance of trip production among 4-digit zones in The Netherlands.

To investigate the relative prediction error, Figure 4.13(a) explores the relationship between Mean Absolute Percentage Error (MAPE) and Coefficient of Variation (COV) in trip production among the 4-digit zones. The figure indicates a positive correlation between MAPE and COV, suggesting that the prediction error increases with an increase in trip production variation. However, some outliers with high MAPE and low COV suggest high prediction errors despite low variation in trip production. Figure 4.13 highlights these TAZs in red. Overall, the results suggest that the prediction of trip production becomes more challenging in certain zones due to underlying characteristics.



(a) MAPE of trip production prediction vs. the COV of trip production.

(b) Zones with the highest MAPE/COV ratio.

Figure 4.13: MAPE of trip production prediction vs. the COV of trip production among 4-digit zones in The Netherlands.

4.4 Conclusion

This study presented a GCN+LSTM framework that integrates spatial adjacency into trip production prediction across multiple spatial scales. We demonstrated incremental gains in predictive accuracy by considering spatial heterogeneity and identified socio-spatial features critical to understanding residual patterns. These findings support the development of more tailored, accurate, and practical demand prediction models.

In this chapter, we addressed two main objectives. The first objective was to develop data-driven short-term prediction models for trip production patterns that explicitly consider spatial-temporal heterogeneity. To achieve this, we used GCN and added the adjacency information of the TAZs in the network. By doing this, we reduced the computation time while increasing the prediction accuracy. Additionally, using one model for the entire network instead of one for each TAZ decreased the number of models required for prediction.

The second objective of this study was to investigate the effect of spatial scale on spatial uncertainty and, consequently, on trip production prediction accuracy. We found that introducing the adjacency of the TAZs in the network decreased the prediction computation load using GCN. Furthermore, reducing spatial resolution, i.e., going to higher levels of abstraction, showed increased prediction accuracy despite adding uncertainty to the spatial correlation of TAZs. We attribute this increase to the aggregation of trip production patterns, which may become more regular and, thus, more predictable.

The findings of this research have several implications for understanding how spatial heterogeneity affects demand prediction and for transportation planning and policy-making. The results of this study are helpful for transport modelers to consider the spatial scale in selecting the relevant types of variables used in their models for travel demand prediction. For instance, recognizing that at lower levels of spatial abstraction, a combination of land-use and demographic information contributes significantly to residual patterns, while at higher spatial scales, land use plays a more critical role. This awareness allows for more tailored and accurate modeling approaches, leading to better-informed decisions regarding resource allocation, infrastructure development, and service planning.

Incorporating spatial adjacency into trip production prediction models using GCN has demonstrated improvements in prediction accuracy and computational efficiency. This enhancement enables transportation planners to efficiently handle large-scale networks without compromising accuracy, facilitating the development of more effective transportation strategies. Understanding the influence of neighboring TAZs on trip production can help in designing more efficient transit networks and optimizing traffic management strategies that account for spatial interactions between regions.

Although we focused on the prediction of trip production, the insights gained from this study assist in OD matrix estimation and prediction. Furthermore, while our study focused on motor-vehicle OD matrices, addressing trip production patterns specific to motor vehicles, it is essential to note that the methodology employed in this research is mode-agnostic. This versatility implies that our approach could potentially be adapted and applied to various other modes of transportation, broadening the scope and impact of the findings presented in this paper.

Overall, this research lays the groundwork for developing more complex demand prediction models with higher accuracy without requiring high computational capacity. By refining models to consider nuanced socio-spatial dynamics, we enhance their predictive accuracy and applicability, supporting the creation of more responsive and equitable transportation systems.

The study also enhances our understanding of spatial uncertainty across multiple spatial scales and how it affects the prediction of trip production. However, a major limitation of this study is that the adjacency of the TAZs is the only feature used for generating the adjacency matrix in our GCN. Other socio-spatial features, as we showed in this research, also affect the heterogeneity of trip production in areas. For instance, the spatial urbanization level, defined based on the combined land-use characteristics and demographics of an area, has been shown to affect trip production patterns and cause heterogeneity. Therefore, there is a definite need to consider other contributing spatial features for defining the adjacency matrix and comprehensively defining a dynamic (showing that features can change over time) adjacency matrix to address trip production heterogeneity.

Compared to previous studies that utilized traditional time series models or standalone deep learning models, our GCN+LSTM model demonstrates improved prediction accuracy by effectively capturing spatial-temporal dependencies. For instance, studies like Zhao et al. (2019) and Yu et al. (2017) also highlighted the benefits of incorporating spatial information. Our findings align with these results and further extend them by analyzing multiple spatial scales and focusing on trip production prediction at a national level.

As a future direction, we are committed to exploring cutting-edge methodologies in the domain of demand prediction, with a keen focus on benchmarking an array of forecasting paradigms. This includes delving into models that intricately weave together more sophisticated spatial data representations alongside advanced machine learning algorithms. Our objective is not merely to augment the predictive precision of these models but to catalyze a transformative shift in their capability to anticipate travel demand dynamics accurately. By charting this course, we aspire to contribute to the perpetual enhancement of forecasting models, ensuring they remain both relevant and robust in the face of evolving transportation landscapes and the complex interplay of spatial-temporal factors they encompass.

Chapter 5

Sensitivity of Travel Demand Patterns to Socio-Spatial Characteristics

---

This chapter addresses the potential research gap in understanding the relationship between dominant demographic and land-use information (i.e., socio-spatial characteristics) and the predictability of demand patterns at different spatial scales in the context of trip production prediction. The chapter proposes a data-driven approach that analyzes several data sources on factors such as population density, collective land-use characteristics, average age, income, household size, and employment status. The analysis aims to understand the complex relationships between these factors and the predictability of origin-destination (OD) matrix predictions. While these socio-spatial characteristics are commonly used in trip production models, there is a lack of understanding about how different demographic factors impact the accuracy of these models and how they vary at different spatial scales. Therefore, this chapter aims to provide further insights into how these characteristics affect the predictability of trip production and to identify the most relevant factors for predicting trip production at different spatial scales.
This chapter is based on the following papers:

Eftekhar, Z., Pel, A., & van Lint, H. . (2023). A Cluster Analysis of Temporal Patterns of Travel Production in the Netherlands: Dominant within-day and day-to-day patterns and their association with Urbanization Levels. European Journal of Transport and Infrastructure Research, 23(3), 1–29. (published)

Eftekhar, Z., Behrouzi, S., Krishnakumari, P., Pel, A., & van Lint, H. . The Role of Spatial Features and Adjacency in Data-driven Short-term Prediction of Trip Production:An Exploratory Study in the Netherlands (under review)

Abstract

Expanding upon the findings in Chapter 3, further analysis of the mixed-level areas shows a more complex relationship between temporal heterogeneity and spatial characteristics. Population density seems to impose additional uncertainty on the temporal patterns. All in all, feature selection and spatial and temporal discretization play essential roles in identifying the dominant trip production patterns. **In continuation with the trip production predictions elaborated in Chapter 4,** we examine prediction errors by identifying and evaluating dominant patterns of trip production prediction residuals. Then, we pinpoint the most critical demographic and land-use features of TAZs contributing to their associated residual clusters and compare them across different spatial scales. Our analysis extends beyond mere prediction errors to systematically dissect and interpret the prevailing patterns in trip production prediction residuals. By identifying the demographic and land-use characteristics most instrumental in shaping these patterns across various TAZs, our research deepens the collective understanding of trip production variability. Such insights pave the way for the formulation of more precise and robust demand prediction models tailored to different spatial scales.

Beyond the methodological innovation, this paper significantly contributes to the analysis of prediction results at multiple spatial scales, examining the association between prediction accuracy (residual patterns) and the types of built environment variables. This research advances travel demand models by systematically integrating socio-spatial characteristics, such as land use, points of interest, and demographics, identified as key influencers of travel demand across various scales. This allows for the creation of more effective models, tailored to the actual needs of society. By refining models to consider these nuanced socio-spatial dynamics, we enhance their predictive accuracy and applicability, supporting the development of responsive and equitable transportation systems. This approach not only improves model accuracy but also ensures transport planning and policies are based on a comprehensive understanding of the factors driving travel demand, leading to more targeted and effective transport planning and policy-making.

In this study, we use processed aggregated derivatives of GSM traces in the form of motor-vehicles OD matrices of the Netherlands in March 2017. This implies the scope of this work pertains to motor-vehicle trip production patterns.

5.1 Introduction

Travel demand patterns are characterized by spatial-temporal heterogeneity, as the amount of travel differs strongly across areas as well as time-of-day and day-of-week (Shen et al., 2020). Temporal variability is especially significant when modeling motor vehicle demand in urbanized areas where morning and afternoon peaks account for almost 50% of daily travel demand (Lin & Shin, 2008). Spatial heterogeneity is derived from diverse urbanization levels, economic activities, lifestyles, transportation accessibility, and resource distribution between areas, e.g., Fotheringham et al. (1998). This spatial-temporal heterogeneity, including the identification of patterns therein, is an important part of understanding travel demand.

Many studies analyze the relationship between travel demand and land-use properties using methods based on Ordinary least squares (OLS) (Yang et al., 2018b; Maat & Timmermans, 2006). OLS generally assumes homogeneous regression relationships in the data. However, spatiotemporal data has the basic features of both spatiotemporal dependence and heterogeneity (Ermagun & Levinson, 2018; Atluri et al., 2018). For instance, in the temporal dimension of traffic data, the traffic demand of areas is typically similar to that of recent times. Moreover, they might show periodicity, trends, and holiday effects. In the spatial dimension, the traffic state of each road segment is often (but not always) similar to upstream and downstream traffic conditions. This is an example of spatiotemporal dependence (Cheng et al., 2013). Due to complex nonlinear variations, the traffic demand also has different distributions in different geographical regions and time intervals. For instance, for some areas, the peak hours of travel production happen in the afternoon, while it is in the morning for other areas. This phenomenon represents spatiotemporal heterogeneity (Atluri et al., 2018). Overlooking spatiotemporal heterogeneity gives rise to some errors, for instance, misinterpretation of coefficients and inaccuracy in estimations (Anselin & Griffith, 1988; Deng et al., 2018; Cheng et al., 2019).

To account for spatial heterogeneity, many extended OLS-like models have been developed, amongst which the geographically weighted regression (GWR) model (Brunsdon et al., 1996) is widely used in transportation studies. For example, Cardozo et al. study the relationship between transit travel demand and land use mix, bus accessibility, and road density using a GWR model. Whereas the GWR model can sufficiently describe spatial heterogeneity, it does not address temporal heterogeneity. Typically, days are divided into multiple periods, in which average (proportional) values are considered for modeling.

To incorporate temporal heterogeneity, a geographically and temporally weighted regression (GTWR) method to predict transit travel demand was first applied by Ma et al. (2018). However, little is still known about how different areas have various spatiotemporal patterns in travel production associated with their urban development. Thus, this chapter will delve into these uncharted territories, seeking to understand the factors that mediate the interactions between urbanization and travel production in time and space. Through this study, we aim to clarify how different urban factors such as land use and demographic information interact, influencing travel patterns. Our findings will provide valuable insights to improve travel prediction models. Additionally, these insights have the potential to guide urban planning decisions and transportation policies. The sequence of analyses builds upon the previous chapters' findings to contribute a new layer of understanding to our study.

5.2 Methodology

This chapter outlines the methods used to explore how trip production is influenced by various levels of urbanization, building directly on the patterns identified in Chapter 3. The methodology employed here is two-fold. First, it examines the link between urbanization levels and trip production patterns. This foundational analysis sets the stage for the sections that follow.

The subsequent part of this chapter takes a different angle, focusing on the examination of trip production prediction residuals. This segment, along with the next, is grounded in the results and discussions from Chapter 4. It involves a detailed analysis of the residuals, emphasizing their significance and the insights they can provide.

In the third segment, the discussion transitions into an exploration of the connections between various socio-spatial characteristics and the patterns present in the residuals. By doing so, it integrates diverse demographic and land-use factors, assessing how they correlate with the predictability of trip production patterns.

Collectively, these methodological steps create a comprehensive approach to understanding trip production in the context of urban settings.

5.2.1 Association with urbanization levels

In this step, we analyze the degree to which the K clusters, **found in Chapter 3**, are associated with specific urbanization levels. That is the degree to which the urbanization level of a TAZ can be used as a predictor for to which cluster of temporal production patterns it belongs.

Comparing the overall area of different land-use types (derived from Open Street Map (OSM) (OpenStreetMap contributors, 2017) data) in the clusters helps relate the resulting clusters (i.e., temporal production patterns) to land-use types. Assuming a non-linear relationship, we propose a tree-based ensemble machine learning method, eXtreme Gradient Boosting (XGBoost) Ensemble, to model the relationship between the land-use feature and clusters. Due to the regularization term in the loss function, one can achieve the lowest complexity with the highest accuracy. For more details on the XGBoost algorithm we refer the readers to appendix A.1 and Chen & Guestrin (2016). The framework of our method, called the OVR-SMOTEXGBoost ensemble model, for our multi-class imbalanced data is shown in Figure 5.1 and the three main steps in this algorithm are described as follows.

1. Decompose the multi-class classification problem into multiple independent binary classification problems. Traditionally, classification methods are designed for two-class (binary) problems. Furthermore, as the generality of multi-class classification problems naturally makes learning more complex, an intuitive approach is to solve such problems by decomposing them into several binary classification problems. In One-vs.-Rest (OVR) strategy, one develops multiple classifiers (i.e., one classifier per class indicating being or not being in a specific class) (Hong & Cho, 2008). Based on the OVR decomposition method, the initial training set is decomposed into three two-class training sub-samples: $Train_1$ for class1, $Train_2$ for class1, and $Train_3$ for class3. Then using a binary logistic loss function, each classifier calculates the probability of each class. Then each instance is classified into the class of the highest probability.

2. Balance the training sets. An issue regarding the classification is the imbalanced learning problem. Underrepresented data and class distribution skew affect the learning algorithm performance and cause this problem (He & Garcia, 2009). The synthetic minority oversampling technique (SMOTE) is one of the most recognized data augmentation methods to address the imbalanced data problem (Zhai et al., 2022). It attempts to balance class distribution by oversampling minority class instances by randomly replicating them.

3. Use the training sets to train the SMOTE-XGBoost ensemble model for a binary class of TAZs prediction (i.e., each of three OVR classifiers).

To avoid overfitting, we use a K-fold strategy with $k = 10$ for the training and testing (i.e., cross-validation) (Note that K here is completely unrelated to K in the earlier K-means clustering method in that these are two independent parameters, but that simply both these methods happen to use the same letter). K-fold divides the data elements into K groups of samples, i.e., k folds, and the training of the model is achieved using K-1 folds, and testing is done on the left-out fold. For instance, given a dataset of 100 samples, 90 (i.e., nine folds) are used for training and validation. However, the predictive accuracy is tested on the remaining ten (i.e., one

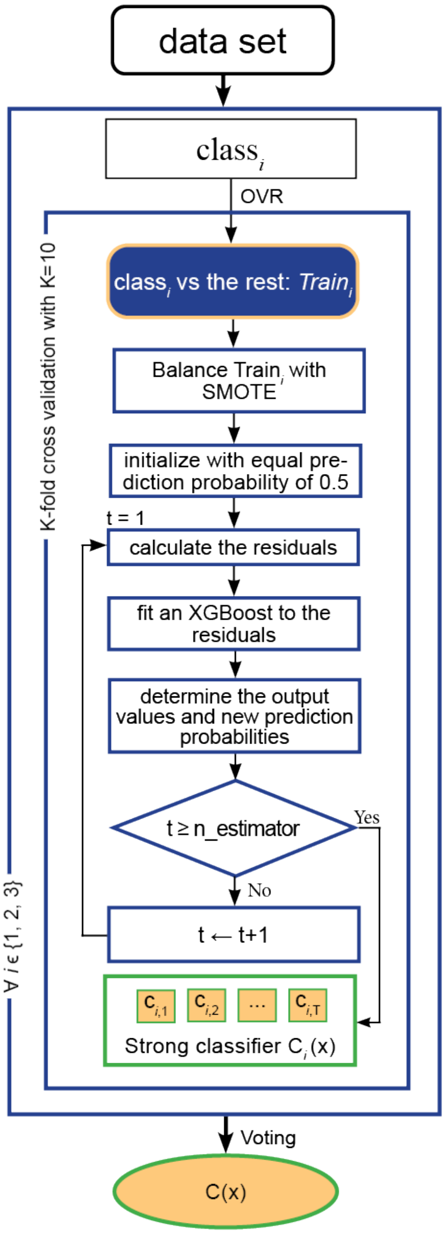


Figure 5.1: The framework of the OVR-SMOTE-XGBoost ensemble model.

fold) samples. This process is iteratively repeated for all ten folds (with no overlap) to achieve a robust accuracy in prediction.

To further investigate the main spatial patterns inside a cluster, we apply a hierarchical clustering analysis (HCA). This step leads to several spatial sub-clusters explaining the temporal heterogeneity observed in the temporal clusters. HCA is generally a clustering analysis method that tries to build a hierarchy of clusters. In this study, we used an Agglomerative approach where each TAZ starts in its own cluster (i.e., bottom of the hierarchy). At each iteration, the similar clusters, based on their spatial characteristics, merge with others until one cluster or N clusters are formed (i.e., top of the hierarchy). In this research, we calculate the similarity between two clusters using Ward's method, which is the sum of the square Euclidean distances between the two associated clusters.

5.2.2 Residual Analysis

To investigate prediction errors, **based on the prediction of trip production in Chapter 4**, we analyze the residual heatmap of hourly trip production prediction per Traffic Analysis Zone (TAZ). First, we predict trip production and compute the residual as the absolute difference between the predicted and actual values. To concentrate on discerning patterns, we normalize the residuals per TAZ using Min-Max Scaling. Consequently, each zone has a heatmap of normalized residual values, with each cell ranging between 0 and 1. Each heatmap's horizontal and vertical axes represent the days of the prediction horizon and the hours of the day, respectively. As an example, Figure 5.2 shows the heatmap of trip production residual for a randomly selected TAZ. This representation enables us to observe the temporal patterns of residuals within a day

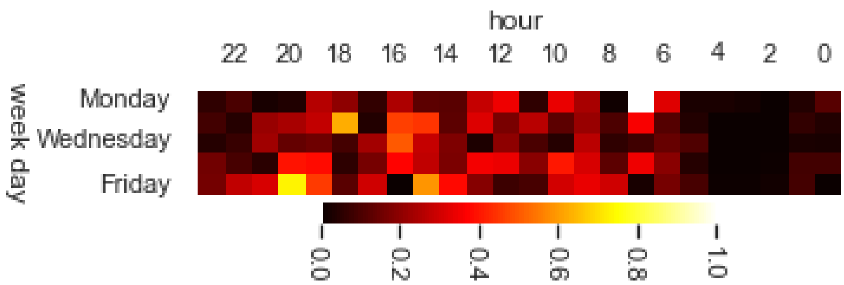


Figure 5.2: An example heatmap of trip production residual of a random TAZ.

(by comparing across columns) and between days (by comparing across rows). The heatmaps for each under-study spatial scale serve as the foundation for subsequent analyses.

To identify patterns of residuals, we cluster heatmaps of TAZs per spatial scale based on temporal similarity using K-means clustering. Owing to its straightforward application and interpretability, which is crucial in exploratory analyses for generating clear, understandable results, the K-means clustering method is a widely employed algorithm for clustering analysis (Poteraş et al., 2014; Szegedy et al., 2016; Cohn & Holm, 2021; Van Gansbeke et al., 2020; Sun et al., 2021). The method aims to partition the N-dimensional dataset of M points (heatmaps) into K clusters, minimizing the sum of the pairwise Euclidean distance between the points in each cluster (Hartigan & Wong, 1979).

To identify patterns of residuals, we employed K-means clustering due to its simplicity, efficiency, and interpretability, which are crucial for exploratory analyses. K-means is a widely used algorithm that partitions the data into K clusters by minimizing the within-cluster sum of squares (Lisboa et al., 2013). While other clustering methods such as hierarchical clustering or DBSCAN could be considered, K-means provides a straightforward approach that aligns well with the objectives of our study. Exploring other clustering methods is an avenue for future research.

The clustering process comprises two primary steps:

- **Assignment:** Assigning each point to its closest centroid, mathematically referring to the partitioning of the points to the Voronoi diagram (Shamos & Hoey, 1975) generated by the centroids.

- **Update:** Updating each cluster center to be the average of all points contained within them.

As the K-means method requires exogenous determination of the number of clusters, we must justify our choice. Since we cannot a priori determine the number of existing patterns, this is an unsupervised learning problem without "true labels" or ground truth. Therefore, the optimal number of clusters is determined by assessing the (dis)similarity within and between clusters for various values of K. We use the Silhouette index (Rousseeuw, 1987) to measure the goodness of clustering.

The Silhouette index ranges between -1 and 1, where high values indicate a well-matched point to its own cluster and poorly matched to neighboring clusters. If many points have negative values, the number of clusters should be adjusted. The optimal number of clusters is determined using the Silhouette Elbow method, which identifies the point at which the rate of improvement in the Silhouette index slows down.

However, the K-means method is inherently linear (Ning & Hongyi, 2016) and unsuitable for complex non-linear data distributions. To account for data non-linearity, we use a deep convolutional neural network (DCNN) for feature extraction. The DCNN transforms input heatmaps into feature vectors, which are more easily separable by a linear clustering algorithm (van Elteren, 2018) than the original heatmap.

The state-of-the-art InceptionV3 based on transfer learning is employed as the DCNN in this research. The InceptionV3 architecture is specifically designed to improve adaptability to different scales and prevent overfitting (Kaur & Gandhi, 2020). Transfer learning allows us to transfer pre-trained model parameters to our new model, thereby accelerating its training (Wang et al., 2019). Utilizing a DCNN trained on a large dataset, such as ImageNet, enables the extraction of generic features applicable to other images, such as heatmaps, without the need for training from scratch. Furthermore, the pre-trained DCNN weights enhance the accuracy of specific tasks, such as pattern recognition, when the available training data is limited (Iglovikov & Shvets, 2018). In this study, we extract feature vectors from the demand heatmaps using the InceptionV3 deep neural network pre-trained on the ImageNet dataset, which contains millions of images for object recognition and image classification (Deng et al., 2009). The compatibility of K-means with the linearly separable data produced by this feature extraction method, further justifies its use, as it balances the analytical approach by complementing the complexity of DCNN.

We use the Inception V3 architecture, pre-trained on ImageNet, to extract latent feature vectors from these residual heatmaps. By passing the residual matrix through the Inception V3 network, we obtain a latent representation—a vector—that preserves the essential information contained in the original matrix form. While this latent vector is not fully interpretable in terms of direct physical meaning, it provides a lower-dimensional feature space in which the residual patterns are more easily separable. This transformation allows us to apply K-means clustering to identify dominant temporal residual patterns more effectively than if we clustered the raw matrices directly.

Ultimately, this step transforms residual heat maps into feature vectors clustered based on temporal similarity. This process yields K clusters with distinct temporal residual patterns and determines the number of such clusters or patterns.

5.2.3 Association of Residual Patterns with Socio-Spatial Variables

In the final step, we analyze the degree to which the K clusters per spatial scale are associated with demographic and land-use variables. We assess the extent to which the socio-spatial characteristics of a TAZ can predict the cluster of temporal residual patterns it belongs to. We compare the distribution of various points of interest (POI) and demographics (derived from CBS) within the clusters to relate the resulting clusters (i.e., temporal residual patterns) to land use and socio-economic characteristics. Assuming a non-linear relationship, we propose a tree-based ensemble machine learning method, eXtreme Gradient Boosting (XGBoost) Ensemble, to model the relationship between land-use features and clusters. The regularization term in the loss function allows for achieving the lowest complexity with the highest accuracy.

To identify the most important socio-spatial features contributing to residual patterns (i.e., clusters), we use the "gain" metric. In XGBoost, the gain metric represents the improvement in accuracy brought by a feature to the branches it is on. Essentially, it measures the importance of a feature in the model by calculating how much the feature contributes to the overall performance of the model. For more details on the XGBoost algorithm used in this study, we refer readers to 5.2.1 and A.1 and (Chen & Guestrin, 2016).

5.3 Results and discussion

This section is structured to present the findings of our research, aligned with the methodology outlined earlier. We begin by discussing the results that explore the relationship between travel production patterns and spatial characteristics. This part provides a foundation for understanding trip patterns within different urban contexts. Following this, our attention shifts to the patterns identified in the prediction residuals of trip production. To conclude, we focus on the results from our assessment of the connection between these residual patterns and socio-spatial variables. This final part synthesizes our findings, underscoring the critical socio-demographic and land-use factors that are intertwined with variations in trip production predictability.

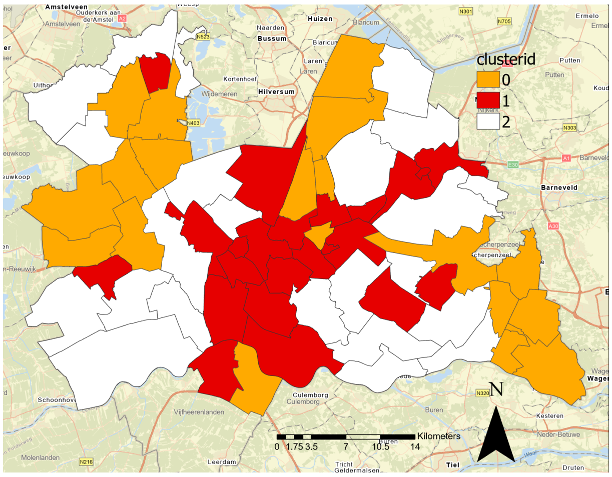
5.3.1 Association between travel production patterns and spatial characteristics

Observing the three clusters in space suggests that zones in cluster *AP* which constitute a smaller proportion of zones (as shown in Figure 3.11(b)), happen to be in more urbanized areas. In fact, as displayed in Figure 3.11(a), out of 43, 39 city centers fall into this cluster. Moreover, the non-metropolitan (i.e., the least urbanized) areas happen to be in cluster *MP* and *mix*. Therefore we hypothesize that these (temporal) clusters are (spatially) associated with urbanization levels. For instance, usually, the majority of farmlands belong to the non-metropolitan (i.e., the least urbanized) areas, thus possibly also to clusters *MP* and *mix*.

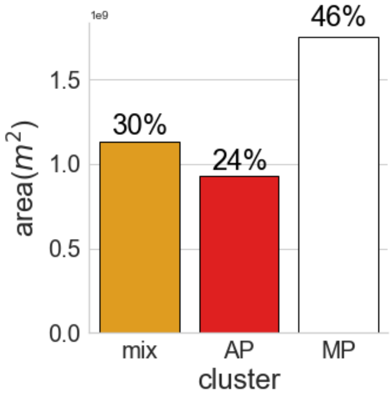
To test this, the association of each cluster with the following land-use characteristics is analyzed:

- population
- commercial and industrial buildings,
- rail and service roads,
- cycleway and footway,
- car and bicycle parking.

This is initially done for the province of Utrecht (as shown in Figure 5.3(a)), but the subsequent model will be based on all 1246 TAZs in the Netherlands. Figure 5.3(b) shows the distribution of TAZs for each cluster in the province of Utrecht.

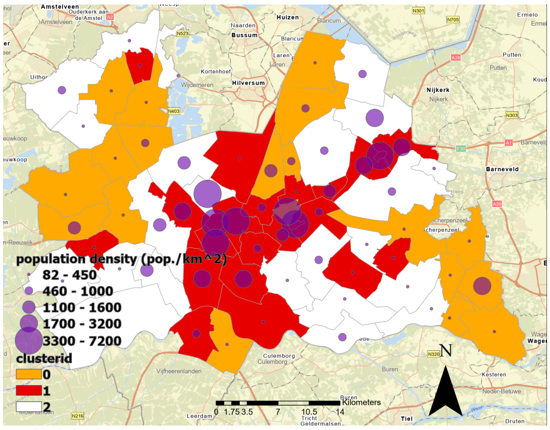


(a) The three clusters.

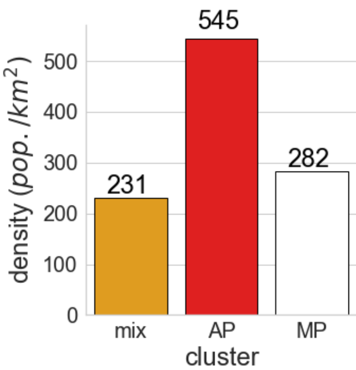


(b) Distribution of total area over the clusters.

Figure 5.3: Share of clusters from the total area of Utrecht province.



(a) Population density of the zones and the three clusters.

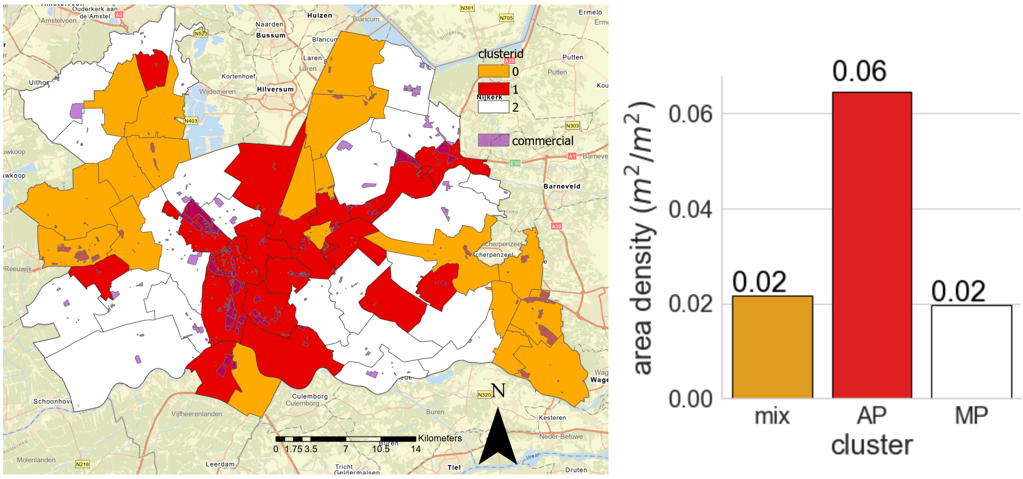


(b) Population density over the clusters.

Figure 5.4: Share of clusters from the total population of Utrecht province.

Looking specifically at the province of Utrecht, Figure 5.4(b) shows indeed a higher average population density in cluster *AP*. Figure 5.5 also shows that the majority of commercial/industrial areas, as a representation of urban areas, in the province of Utrecht belong to cluster *AP*. Figures 5.6, 5.7(b), and 5.8(b) indicate that transportation infrastructure are densely distributed in cluster *AP*. However, Figure 5.9(b) shows a high density of farm and meadow lands in cluster *MP* and *mix*. Also, Figure 5.9 shows that 48% of farmlands which are usually interpreted as rural areas, fall into cluster *MP*. Therefore, density in the mentioned characteristics of each TAZ may be associated with their cluster.

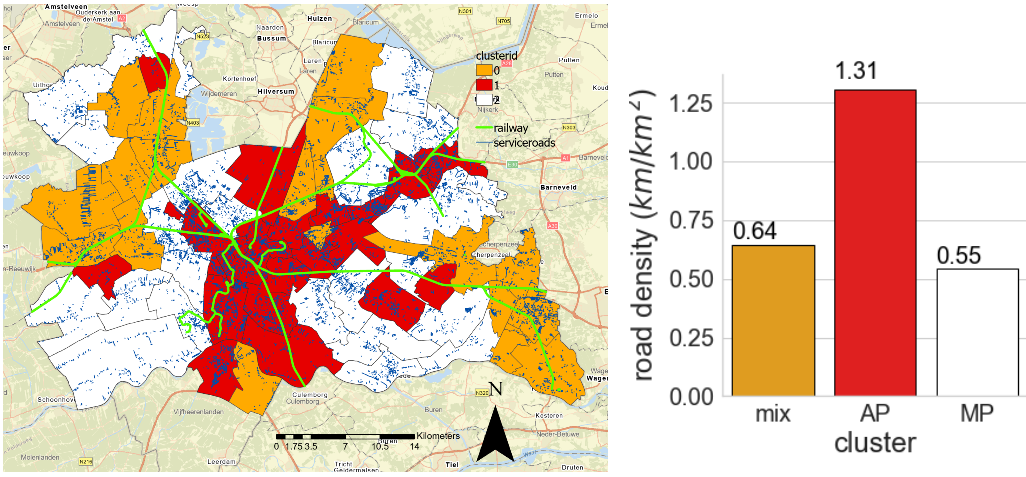
To test this hypothesis at the level of all 1246 TAZs in the Netherlands, an OVR-SMOTE-XGBoost ensemble classification model is trained in which the inputs are the land-use characteristics (i.e., densities) and the output is the cluster, i.e., *AP*, *MP*, or *mix*. The model tries to



(a) Commercial/industrial land-use.

(b) Density of commercial/industrial area over the clusters.

Figure 5.5: Share of clusters from commercial/industrial areas.



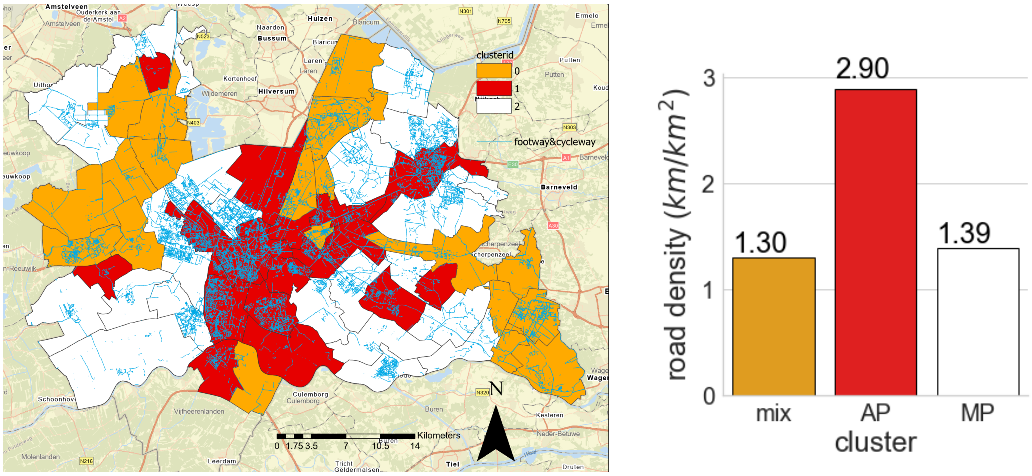
(a) Service roads and railways.

(b) Distribution of service roads and railways total length over the clusters.

Figure 5.6: Road density of clusters from the total length of service roads and railways.

reconstruct the clusters using the land-use characteristics and population density (as presented in Figures 5.4 to 5.9) associated with each TAZ. We used a K-fold cross-validation strategy with $k = 10$ to achieve robust accuracy in prediction. Accordingly, Figure 5.10 shows the confusion matrix of the associated classification.

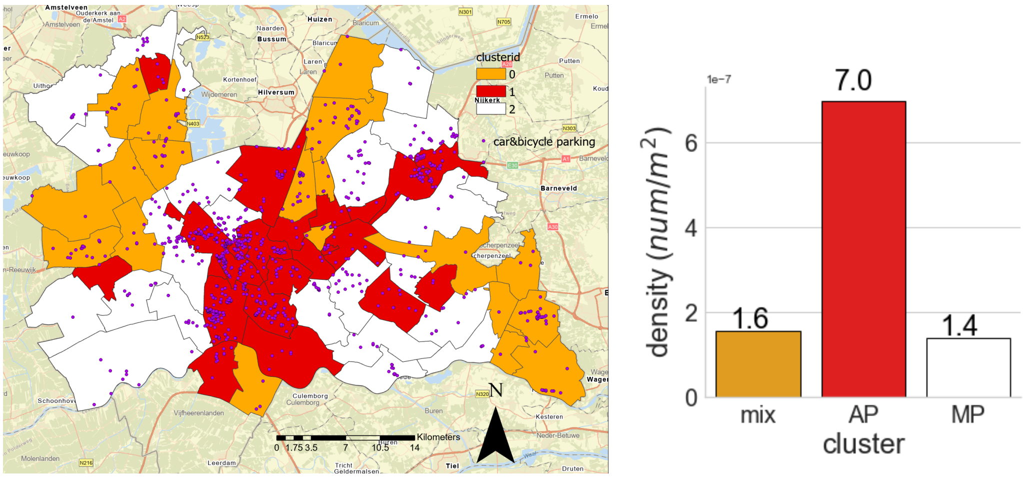
Note that due to our model structure, the possible bias caused by imbalanced training data is avoided. One can imply this by considering that classification accuracy in clusters *AP* is the highest (i.e., 64%), although the number of such TAZs is the lowest (i.e., 286 out of 1246). If the model was biased toward the higher populated cluster, *mix*, the model accuracy would have been around 43%, which is derived from the proportion of TAZs belonging to *mix*. However, our accuracy is higher (53%). Therefore, it seems that the model is not biased due to imbalanced data. In fact, the accuracy of identifying cluster *mix* seems to be the lowest. This may be due to the similarity of patterns in cluster *mix* to both other clusters, i.e., cluster *mix* constitute mix of clusters *AP* and *MP*. Therefore, misclassification is more probable in *mix*. Especially, making



(a) Foot and cycle ways.

(b) Road density of foot and cycleway total length over the clusters.

Figure 5.7: Share of clusters from the total length of foot and cycle ways.



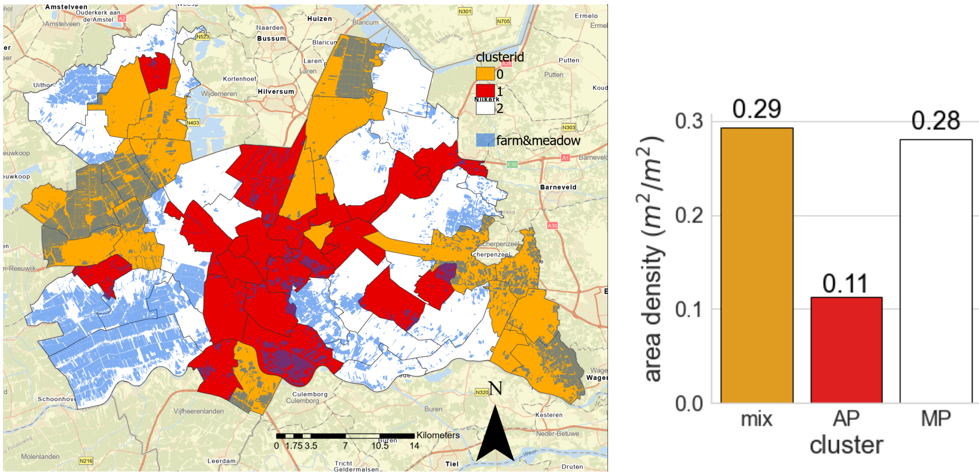
(a) Car and bicycle parking garages.

(b) Density of car/bicycle parking total area over the clusters.

Figure 5.8: Share of clusters from the total area of car and bicycle parking garages.

distinction between *MP* and *mix* seems to be more difficult. This might be due to less extreme peaks in *MP* than in *AP*; thus more difference between *AP* and *mix*. Overall, the land-use feature we used seems to describe *AP* to a greater extent. To look closer, Figure 5.11 presents the confusion matrices of all 12 provinces in the Netherlands. In almost all provinces, the accuracy in *mix* is lower than in the other two. Only one province (Zeeland) shows a different result; In fact, it has the highest accuracy for *mix*. The cause of this behavior needs further analysis, but it can be related to the exceptional location of this province as a river delta situated at the mouth of several major rivers at the country's border(as shown in Figure 5.12).

To take a closer look at *mix*, Figure 5.13 shows the most extreme TAZs associated with false negatives and positives of *mix*. That is, the TAZs with the highest probability to be false positive and false negative of the *mix* cluster. $C_{i,j}$ is the most extreme TAZ whose true label is i, and the predicted label is j, for instance. $C_{mix,MP}$ shows the TAZ in the *mix* cluster, which is incorrectly classified as a TAZ in *MP*. The normalized trip production heatmap shows the



(a) Farmland land-use.

(b) Density of farmland/meadow area over the clusters.

Figure 5.9: Share of clusters from farmland/meadow areas.

true labels \ predicted labels	*mix*	*AP*	*MP*	Σ
mix	219(41%)	115(21%)	205(38%)	539
AP	64(22%)	181(64%)	41(14%)	286
MP	126(30%)	35(8%)	260(62%)	421
Σ	409	331	506	1246

Figure 5.10: Confusion matrix for recognizing TAZs in the clusters, i.e., either AP, MP, or mix.

dominant pattern of *MP*. However, the irregular peak in the afternoon of Friday, March 24th, might have triggered a miss-cluster for Sleeuwijk. What stands out is that all the other three extreme zones in Figure 5.13 fall in Zuid-Holland, the highest-populated province in the entire Netherlands. An explanation might be that irregularities or inconsistencies between the spatial characteristics and normalized trip production patterns are most likely a consequence of the high population density and the associated uncertainties.

The spatial characteristics of the extreme cases in Figure 5.13 are in Table 5.1. Comparing these values with the mean and standard deviation of the clusters (i.e., true labels) in Table 5.2 gives insight into the miss-classification stimuli in extreme cases. For instance, the trip production heat map of $C_{(AP,mix)}$, Leiden West, in Figure 5.13 implies the dominant behavior of cluster *AP*. However, despite high densities in most spatial features and thus being similar to cluster *AP*, high density in Farm & Meadow triggered the miss-classification to *mix*.

Model structure and lack of post-analysis play a vital role in analyzing and alleviating the errors in *mix*. In our probabilistic OVR classification model, we decomposed the three-class problem into three independent binary class problems. Then, using a binary logistic loss function, we calculated the probability of each class. Each instance is then classified into the class of the highest probability. It is worth mentioning that the correct class might be the second highest

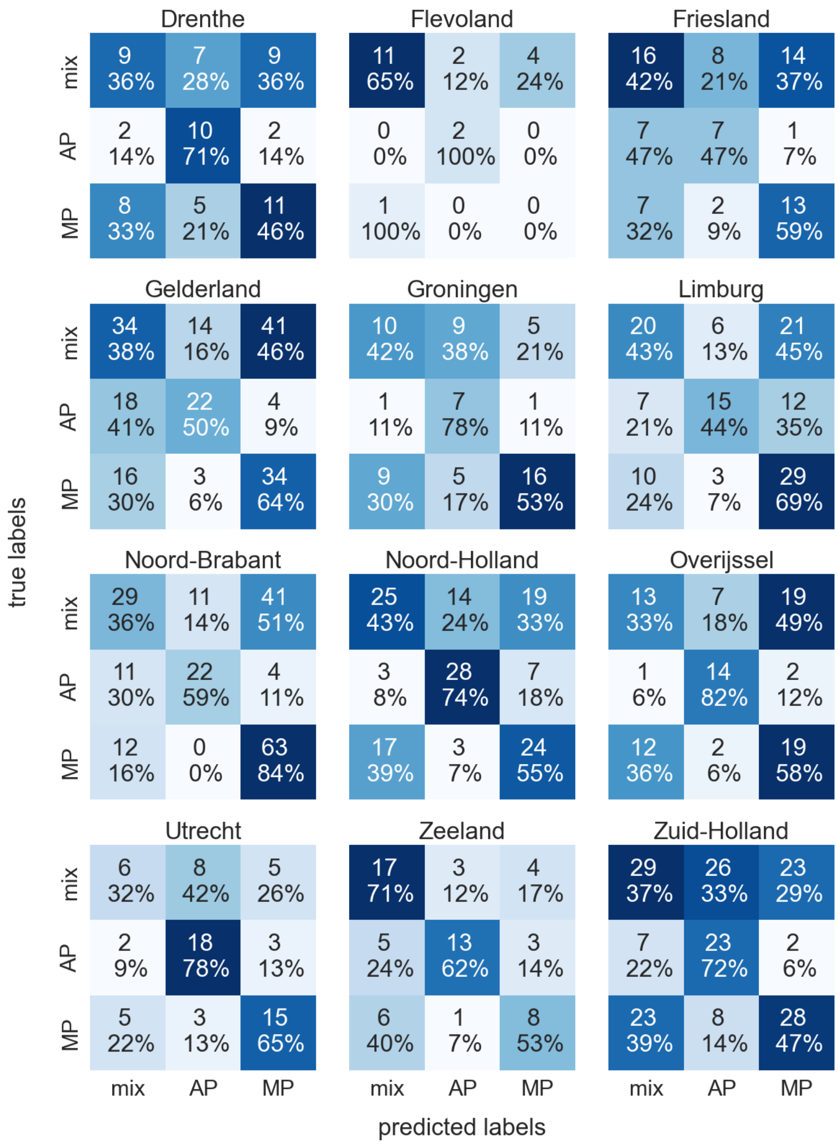


Figure 5.11: Confusion matrix of provinces for recognizing TAZs in the clusters, i.e., either AP, MP, or mix.

probability, with a slight difference from the highest class. If we aim to analyze the more severe errors, we need to flag close probabilities and evaluate the confusion matrix afterward. Then the analysis can focus on more significant errors. In this regard, Figure 5.14 shows that in about 43% of the missclassified TAZs, associated with *mix* (i.e., either actual or predicted cluster is *mix*), less than 20% probability difference is observed between the first and second most probable cluster. Therefore, these edge cases need a post-analysis step to exclude them from genuine prediction errors.

Having considered the above post-analysis, two main components seem to cause the identification errors in *mix*:

1. **Feature selection** plays a crucial task in improving the accuracy and preventing overfitting of pattern recognition algorithms. To obtain the most informative feature subset, we select the important features and delete the unimportant features from the original subset. In this research, we only considered land-use features for our classification al-

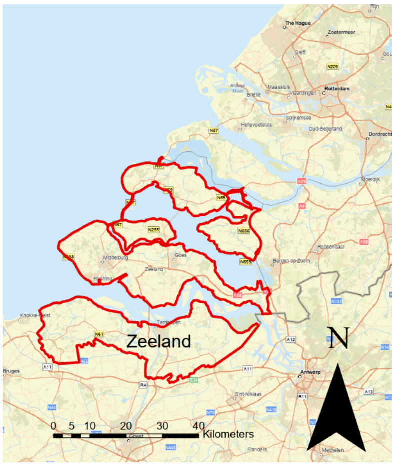


Figure 5.12: Province of Zeeland in the southwest border of the Netherlands.

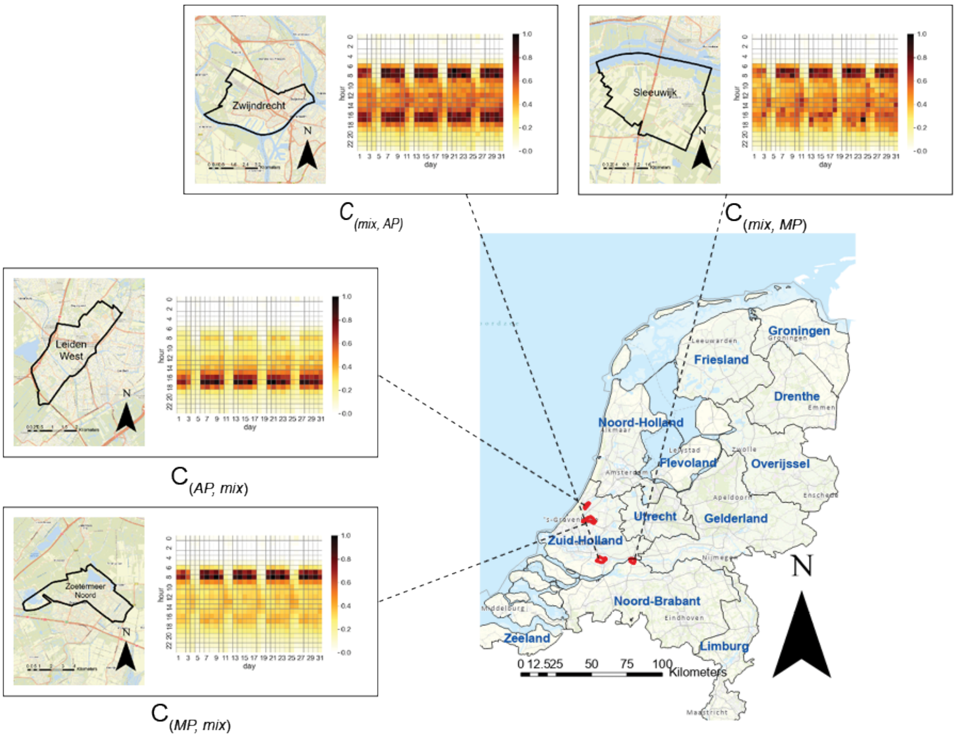


Figure 5.13: TAZs with the highest probability to be false positive and false negative of the ***mix*** *cluster.*

gorithm. Demographic information such as age, occupation, and income level is another feature group that can partially explain travel behaviors. For instance, a student city reveals a different dominant departure time pattern than a city with a higher population ratio of labor workers. As a result, the trip production pattern is also different based on the demographic profile of the TAZ. Another challenge regarding the feature selection is the dynamic relationships between the target classes(i.e., patterns) and candidate features. Therefore it is also challenging to measure the relationship between candidate features, the selected features, and target classes(i.e., patterns) in the feature selection process.

2. **Spatial and temporal discretization problem** is an inefficiency in the land-use data

Table 5.1: Spatial characteristics of the extreme TAZ in Figure 5.13.

TAZ	Zwijndrecht	Sleeuwijk	Leiden West	Zoetermeer Noord
True label	*mix*	*mix*	*AP*	*MP*
Predicted label	*AP*	*MP*	*mix*	*mix*
commercial & Industrial land-use density (m^2/m^2)	0.074	0.005	0.062	0.014
Farm & meadow land-use density (m^2/m^2)	0.023	0.089	0.011	0.064
Footway & cycleway density (km/km^2)	2.54	0.52	6.56	5.99
Railway & service road density (km/km^2)	3.57	0.20	2.87	1.28
Parking garage density (Num/km^2)	0.37	0.12	0.23	0.29
Population density (Num/km^2)	952	214	1521	1220

Table 5.2: Mean and standard deviation of Spatial characteristics of the true labels (i.e., clusters).

Cluster		*AP*	*MP*	*mix*
Commercial & industrial land-use density (m^2/m^2)	mean	0.029	0.007	0.013
	std	0.047	0.010	0.019
Farm & meadow land-use density (m^2/m^2)	mean	0.064	0.107	0.100
	std	0.068	0.075	0.077
Footway & cycleway density (km/km^2)	mean	1.57	0.80	0.98
	std	1.63	0.96	1.17
Railway & service road density (km/km^2)	mean	1.18	0.35	0.5
	std	1.32	0.40	0.52
Parking garage density (Num/km^2)	mean	0.76	0.20	0.21
	std	1.67	0.49	0.68
Population density (Num/km^2)	mean	558	261	290
	std	834	497	439

causing the inability to separate the different travel behavior patterns. The discretization step specifying the data resolution is essential for macroscopic modeling, deriving the necessary high-level insights for traffic planning and management, and reducing the computation time by decreasing the level of details. In this regard, two main factors must be addressed: Firstly, how much spatial variation is observable in each TAZ. Secondly, how often the land-use data is updated (i.e., the temporal variation)? For instance, investigating $C_{MP,mix}$ in Figure 5.13 reveals that in west side of *Zoetermeer Noord* a large area, called "De Nieuwe Driemanspolder", has turned into recreational meadow between 2017 and 2020. This overhaul adds up to the area density of *Farm&Meadow*, causing the trip production pattern to become *MP*. As the land-use data in this study does not seem to be updated frequently, this major change is not reflected, leading to errors in identifying trip

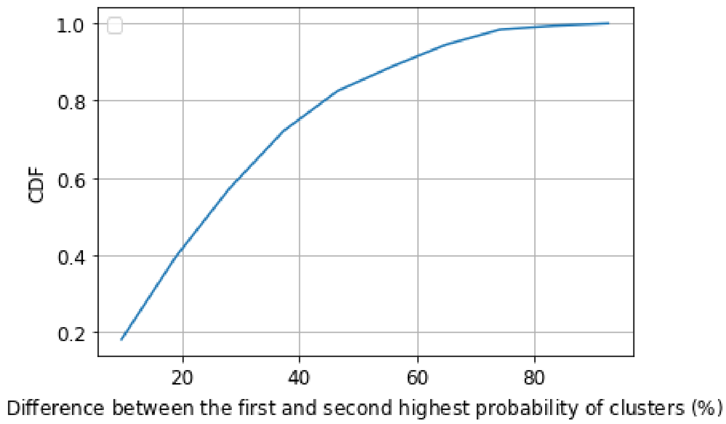


Figure 5.14: Cumulative density function (CDF) of the probability difference between the first and second most probable cluster. This visualization is on TAZs of the false positive and false negative of the mix cluster.

production patterns. Information on spatial and temporal variations of land-use data helps one adjust the abstraction level for drawing relevant insights and design a traffic zoning system with a more crisp separation between different travel behaviors.

To investigate the spatial patterns of the **mix** cluster more closely, we applied hierarchical *clustering on the land-use densities* of all TAZs. Accordingly, we identified three main clusters with different spatial properties as shown in Figure 5.15. Sub-clusters $\{0,1,2\}$ containing 68, 171, and 300 TAZs of *mix*, respectively. Sub-cluster 2 is more similar to *AP*, owing to higher densities of all the mentioned spatial characteristics except for *Farm&Meadow*. This similarity can also be observed in Figure 5.16 and 5.17, showing the average hourly and daily patterns in the three sub-clusters, respectively. On the other hand, sub-cluster 1 contains TAZs with a slightly higher median density of *Farm&Meadow* areas. Unlike cluster *MP*, which also had the highest density of *Farm&Meadow*, Sub-cluster 1 does not have the lowest densities in the other five variables. These areas might refer to the sub-urban residential TAZs connecting the metropolitan to rural areas. These areas show a morning peak indicating commuting travels and an afternoon peak due to non-commuting activities such as shopping, recreational, and social activities. Sub-cluster *0* shows fairly regular morning and afternoon peak with lowest population density and the median *Farm&Meadow* density lower than sub-cluster *1*.

We recapped the overall findings of this research in Figure 5.18. The first clustering step resulted in three dominant temporal patterns in the trip production of all TAZs. Then we performed a correlation analysis to understand the association between spatial characteristics and the dominant patterns. More complex patterns inside *mix* induced us to perform the second clustering on the spatial characteristics of TAZs inside *mix*.

5.3.2 Association of Residual Patterns with Socio-Spatial Characteristics

In this section, we investigate the patterns and characteristics of trip production residuals. To accomplish this, we generated a heatmap of trip production residuals for each TAZ. The heatmap displays the normalized prediction residual of the associated TAZ for each hour of the day (y-axis) over the working days of the week (x-axis) as displayed in Figure 5.2 in the previous section. Each TAZ has its specific day-to-day and within-day residual patterns. To better

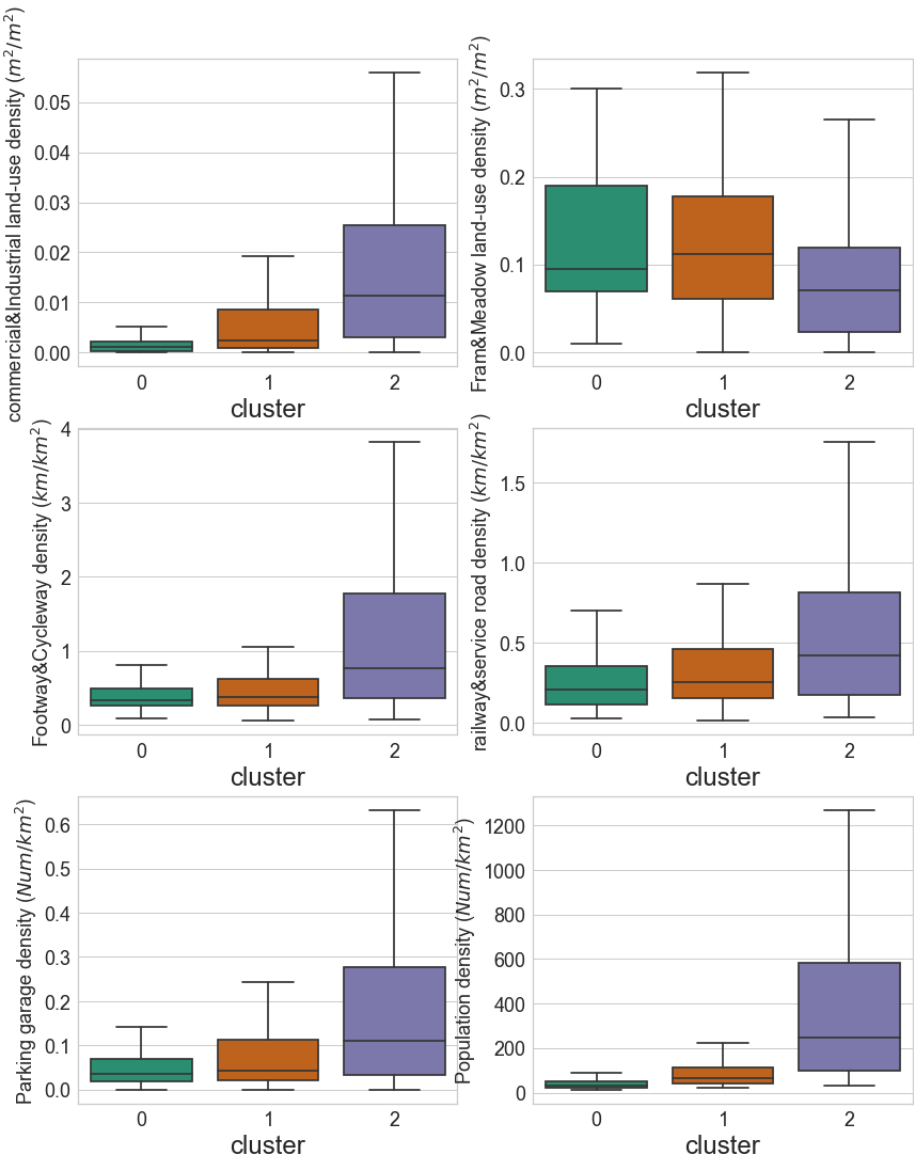


Figure 5.15: Comparing the spatial characteristics of mix cluster's sub-clusters.

understand the dominant patterns of residuals, we analyzed the heatmaps of each spatial level separately, i.e., municipality and 4-digit zones. We employed a DCNN for feature extraction, followed by the K-means clustering algorithm to identify the dominant patterns in each spatial scale. We used the Silhouette score elbow method to determine the optimal number of clusters. Figures 5.19(a) and 5.19(b) depict the Silhouette plots at the municipality and 4-digit zone levels, respectively. Based on these plots, we found four and five clusters suitable for the municipality and 4-digit zone levels, respectively.

Figure 5.20 shows the spatial distribution of identified clusters for TAZs in the Netherlands at different spatial scales: Municipality (5.20(a)) and 4-digit zone level (5.20(b)). At the municipality scale, clusters 0 and 1 appear to be concentrated on the west side of the Netherlands, while clusters 2 and 3 are more prevalent on the eastern side of the country. However, at the 4-digit level, a distinct pattern is not observable by solely examining the spatial distribution. These findings suggest that additional underlying factors might contribute to the formation of production residual clusters and thus warrant further analysis.

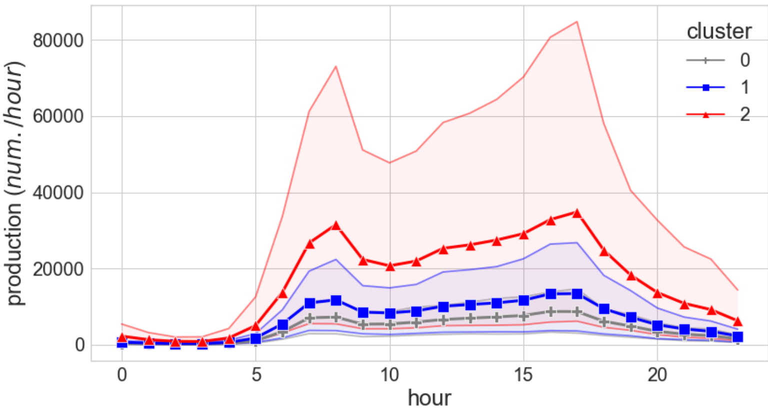


Figure 5.16: Average hourly production of the three sub-clusters of the mix cluster with 10th and 90th percentile as the shaded area.

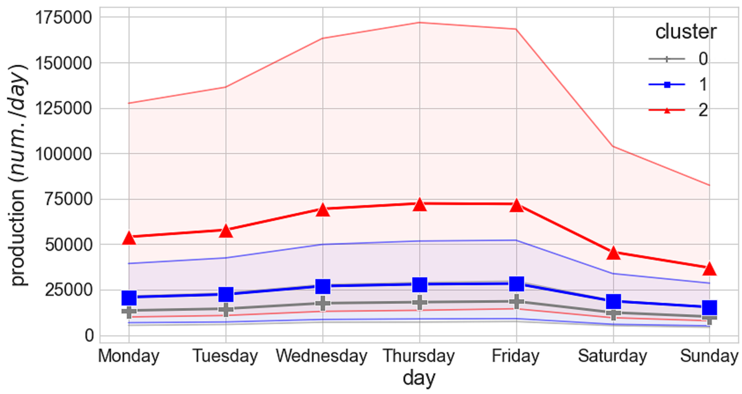


Figure 5.17: Average daily production of the three sub-clusters of the mix cluster with 10th and 90th percentile as the shaded area.

Figures 5.21 and 5.22 compare the residual patterns with the trip production patterns across identified clusters under the municipality and 4-digit spatial scales, respectively. This analysis aims to gain insight into the relationship between trip production and its associated prediction residual. The results shown in Figure 5.21 indicate that the most severe within-day trip production peak occurs commonly between all four clusters at around 6 p.m. or with a slight difference at around 9 a.m. However, the peak of residual is different among clusters. Specifically, cluster 0 displays the highest residual among all clusters, with the residual peak occurring at 4 p.m. and with a slight difference at 7 p.m., while the morning peak occurs at 8 a.m. In cluster 1, residual peak is at 8 a.m., and the afternoon peak occurs at 4 p.m. For cluster 2, the morning and afternoon peaks of the residual are equally severe, and they happen around 8 a.m. and 4 p.m., respectively. In cluster 3, the prediction residual peak is at 7 p.m., and the morning peak occurs at 8 a.m.

Overall, the identified clusters seem to relate to the districts where their prediction residual occurs and the severity of the morning and afternoon peaks. It is interesting to note that, unlike the trip production pattern, whose only afternoon peak occurs at 6 p.m., the residual patterns display a double afternoon peak at 4 and 7 p.m., i.e., 2 hours before and one hour after the

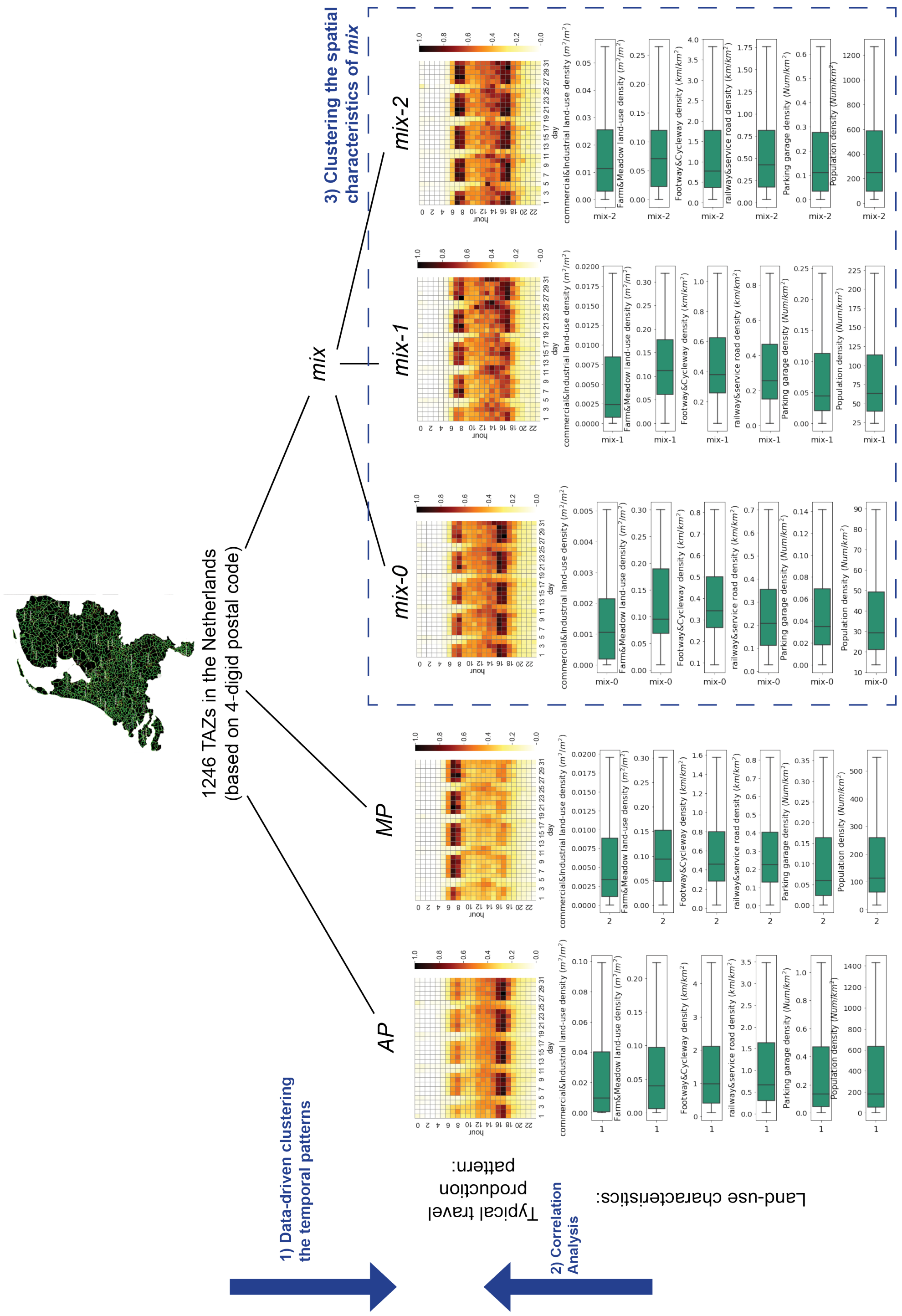
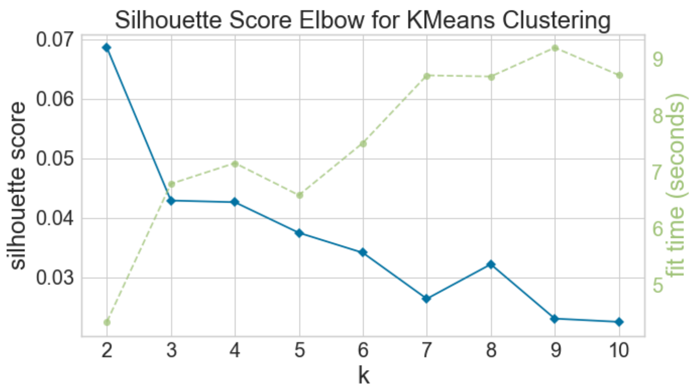
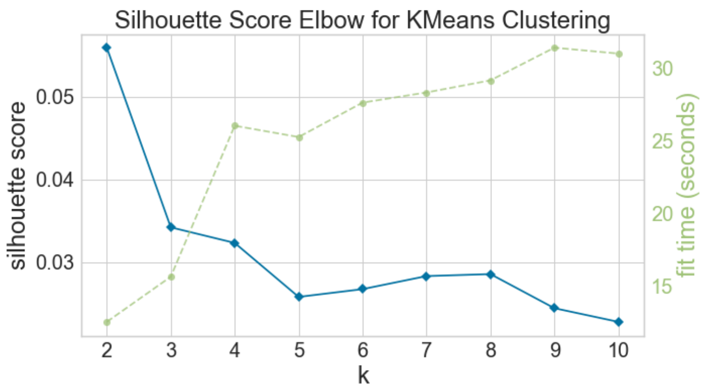


Figure 5.18: Overall findings of the study.

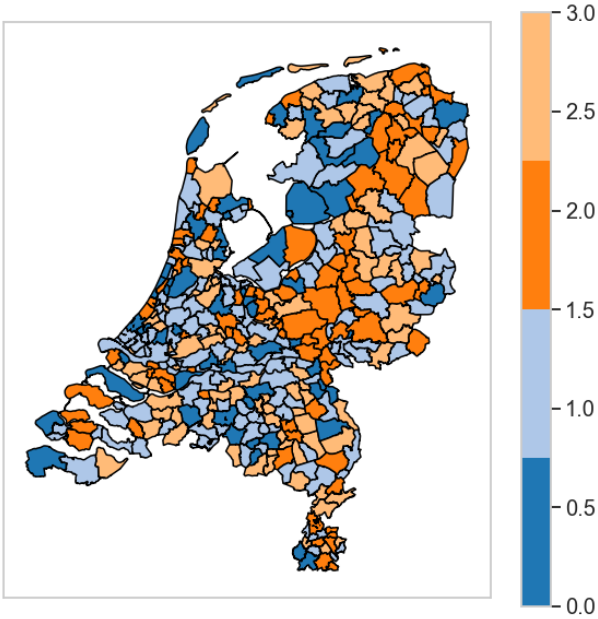


(a) Silhouette plot for municipalities.

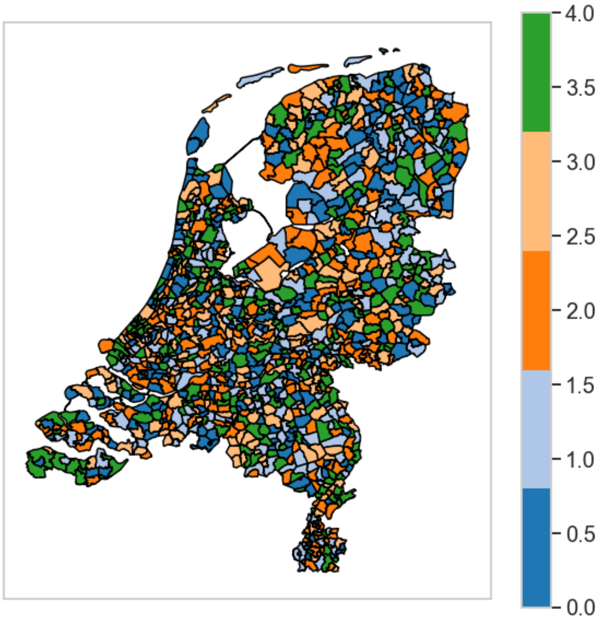


(b) Silhouette plot for 4-digit zones.

Figure 5.19: Silhouette score elbow plots.

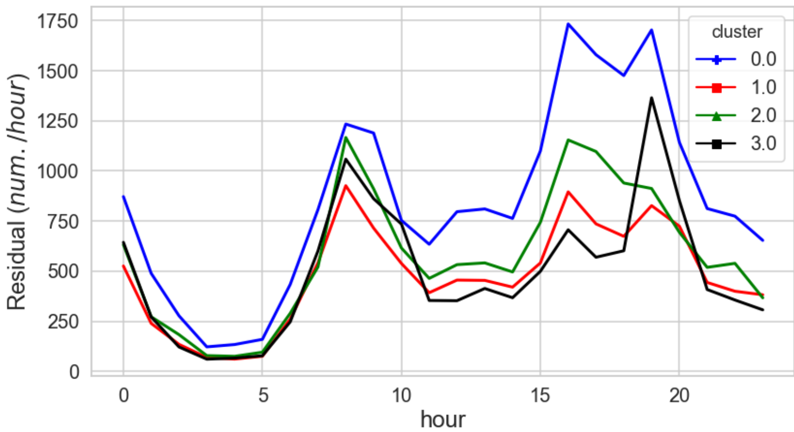


(a) Municipalities.

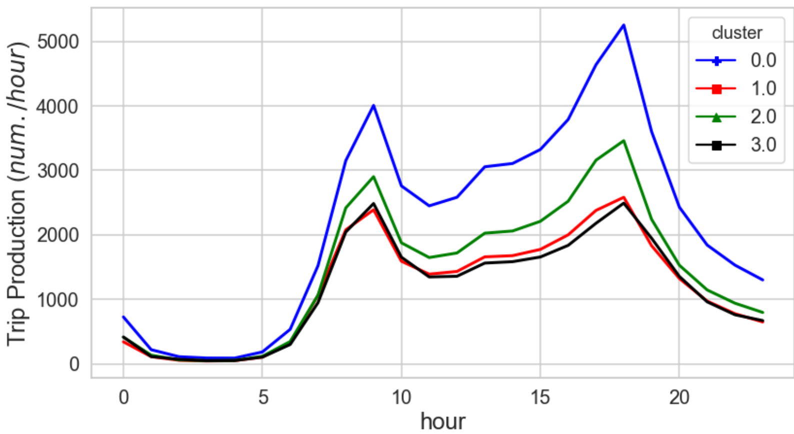


(b) 4-digit postal code zones.

Figure 5.20: Spatial distribution of clusters for TAZs in the Netherlands.



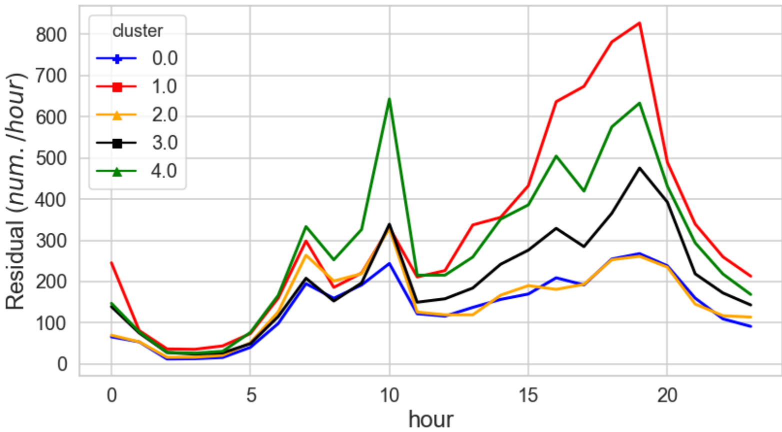
(a) Average hourly trip production residuals.



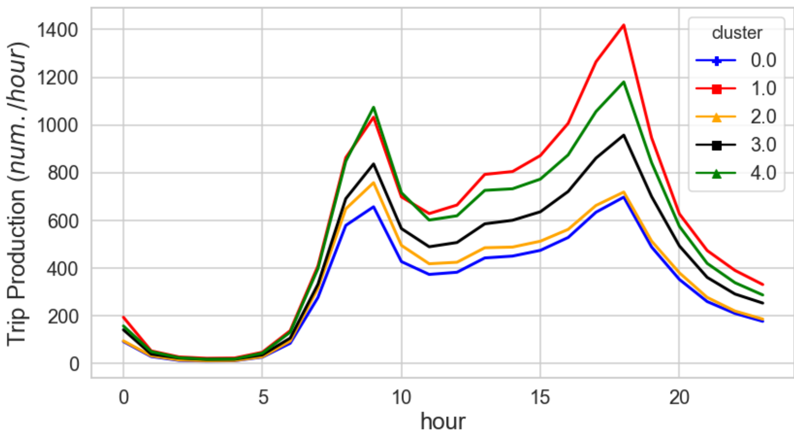
(b) Average hourly trip production.

Figure 5.21: Comparing average hourly trip production and residuals across identified clusters at the ***Municipality*** *Scale.*

production peak in the afternoon. However, the morning peak of residual happens at the same time, around 8 a.m., meaning one hour before the production morning peak. Cluster 0, with the highest residual average at all hours of a day, has a double severe afternoon peak, suggesting that high production results in more irregularities scattered before (i.e., from two hours before) and after (i.e., to one hour after) the actual afternoon peak of production. Moreover, morning activities seem to follow a more strict schedule, as the residual elevation is scattered in a narrow range (i.e., from one hour before to the peak hour of trip production) among all four clusters at this scale. Except for the peak values of residual, the residual patterns follow the discerned patterns in the trip production.



(a) Average hourly trip production residuals.



(b) Average hourly trip production.

*Figure 5.22: Comparing average hourly trip production and residuals across identified clusters at the **4-digit postal code** scale.*

Figure 5.22 illustrates the average hourly residual clusters at the 4-digit level and compares it with the corresponding trip production clusters. The analysis reveals that the residual patterns differ from those identified at the municipality scale, indicating the importance of analyzing them to gain insight into the relationship between trip production and its associated prediction residual at different spatial scales. Specifically, the study finds that the patterns of trip production across these clusters are similar to those identified at the municipality scale, with the morning and afternoon peak of trip production occurring at around 9 a.m. and 6 p.m., respectively, across all clusters. However, unlike the municipality scale, the residual patterns at the 4-digit level exhibit double morning peaks occurring at 7 a.m. and 10 a.m., two hours before and one hour after the morning peak of trip production. Additionally, the afternoon residual peak is reduced to a single peak at around 7 p.m., one hour after the afternoon trip production peak. Further analysis of the five clusters at the 4-digit TAZs level reveals that the severity of the first and second morning and afternoon peaks differs between the clusters, with cluster 1 exhibiting the highest afternoon peak and cluster 4 having the most severe morning peak.

In our analysis, we employed the XGboost algorithm and used the “gain” metric to assess the importance of demographic and land-use features associated with the identified prediction residual clusters. “Gain” is a measure of the relative contribution of each feature to the model calculated as the average gain of the feature when it is used in all possible splits of a tree in the ensemble. The “gain” metric provides an intuitive measure of feature importance and identifies the most relevant features associated with the prediction residual clusters. The higher the gain value of a feature, the more important it is in predicting the outcome variable and the more significant its contribution to the model identifying its residual pattern.

This section presents the results of an importance analysis conducted using the “gain” metric in the XGboost algorithm to identify the most relevant demographic and land-use features that distinguish residual patterns in two spatial scales, municipality, and 4-digit. Box plots of the feature importance scores are displayed in Figures 5.23 and 5.24, where the y-axis represents the feature value, and the x-axis shows the cluster names. It should be noted that the plot displays the distribution of feature values among the TAZs within each cluster rather than representing the gain value itself.

Figure 5.23 shows the results for the municipality scale, where the identified features are mainly related to points of interest (POIs), including the average number of nearby primary schools, cafes, restaurants, and general practices, which tend to increase the residual and make trip production less predictable. In addition, the concentration of older or newer-built houses in the TAZ affects the prediction error, while the proportion of younger residents appears to make trip production more regular, as observed in cluster 0. Notably, cluster 3 exhibits a severe late afternoon peak in residual, an hour later than the trip production peak, with a higher average electricity consumption.

Figure 5.24 displays the importance analysis results for the 4-digit scale, where socio-economic features become more critical in distinguishing the residual cluster of TAZs. The first 12 most important features include residents with a non-EU immigrant parent, the percentage of high-income residents, residents with unemployment, social or disability benefits, the percentage of newborn babies, and population density. Land use and urbanity features also play an important role in distinguishing the residual cluster. For instance, the number of close hospitals, cinemas, and performing art locations, the density of commercial/industrial areas, the urbanity score, and the concentration of middle-aged houses are all relevant. Cluster 1, with the highest afternoon peak in residual, seems to have more close hospitals, a higher proportion of residents with a non-EU immigrant parent, and entertainment locations such as cinemas and performing art venues, as well as a higher percentage of residents with social benefits.

Overall, the results suggest that the prediction of trip production relies more on demographic information when analyzing at lower levels of abstraction, whereas, at higher levels of abstraction, spatial features such as land use and built environment variables play a more critical role in causing irregular demand patterns.

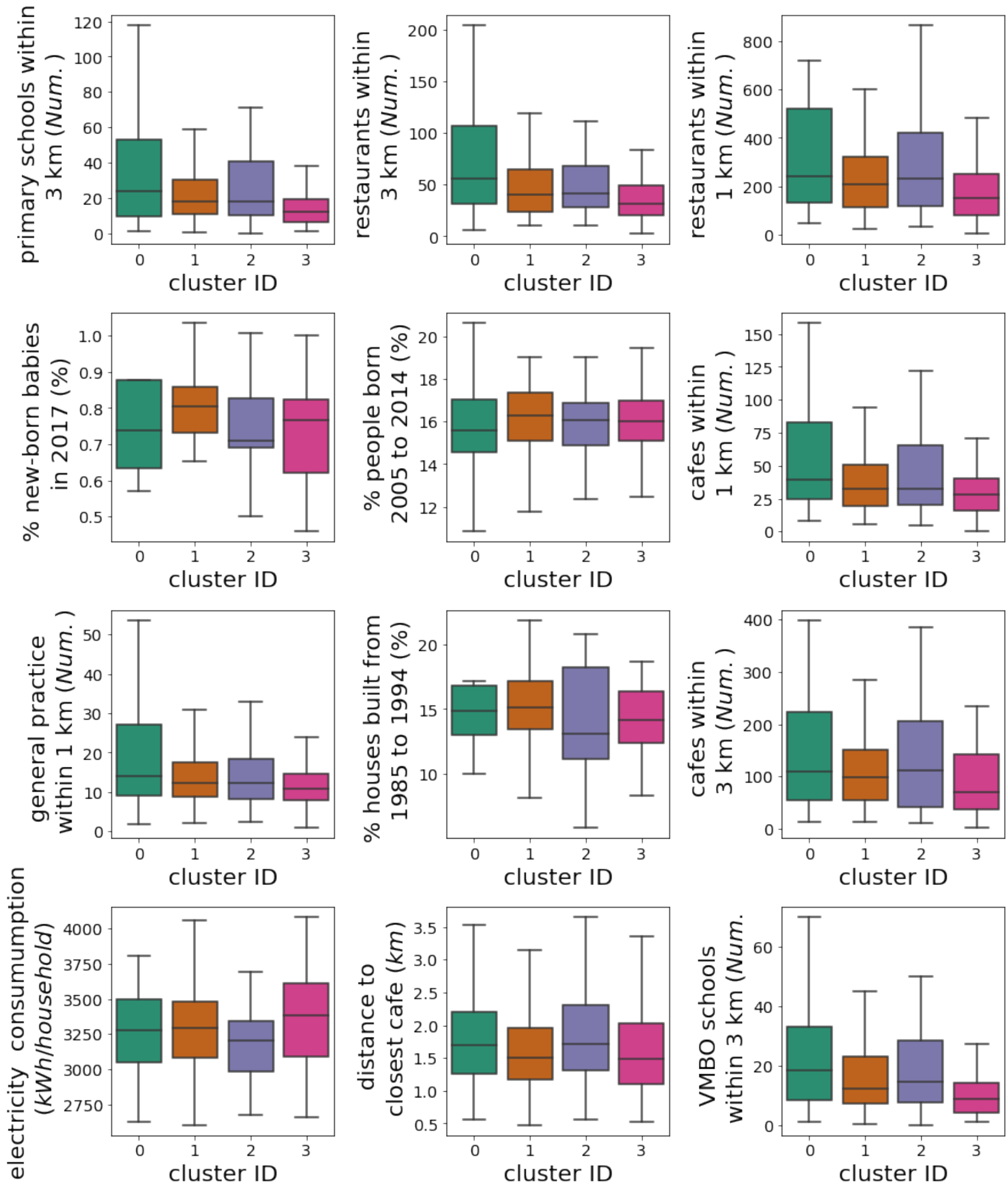


Figure 5.23: Box plots of the most relevant features to the identified residual clusters at the municipality spatial scale.

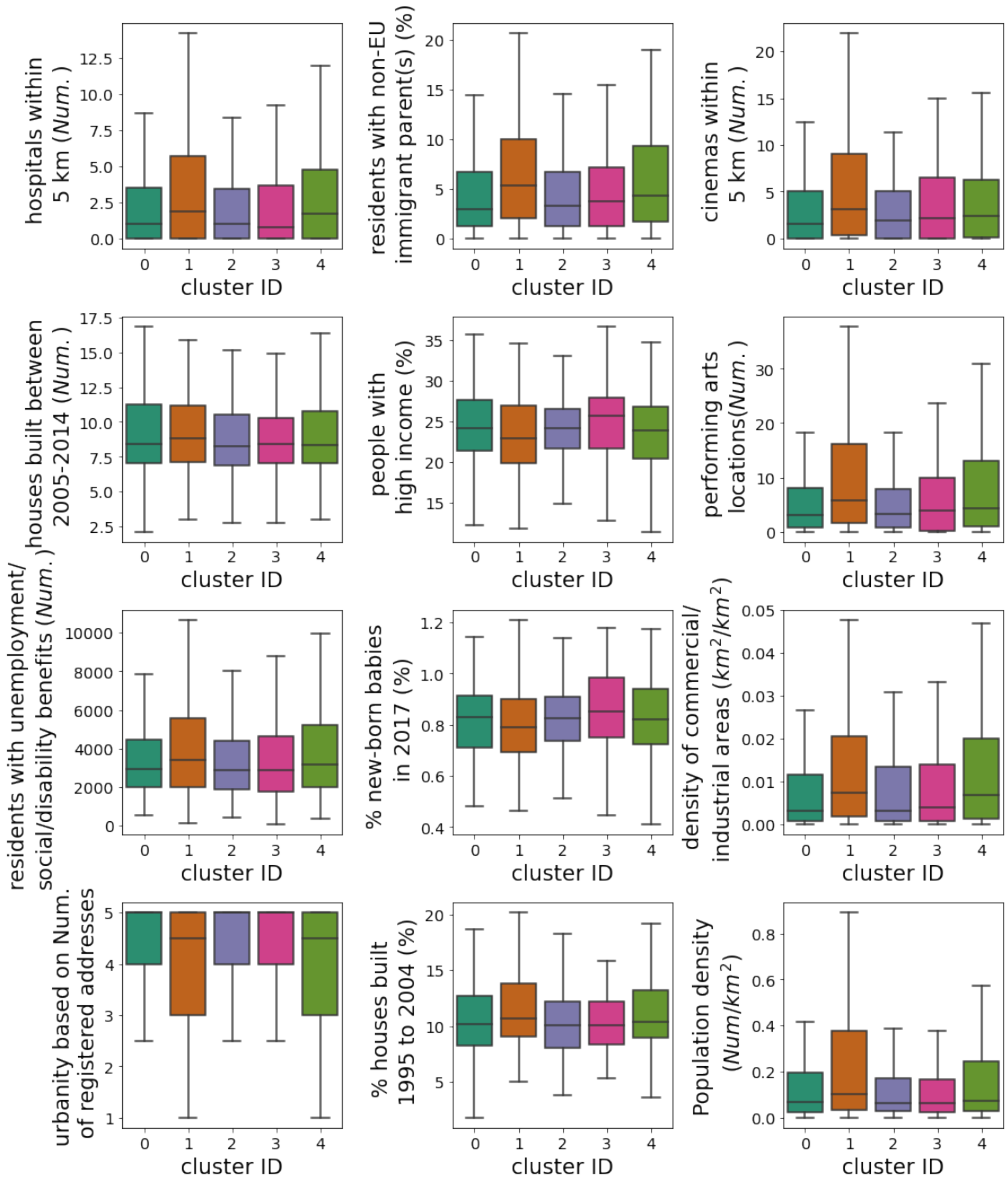


Figure 5.24: Box plots of the most relevant features to the identified residual clusters at the 4-digit spatial scale.

5.3.3 Research Limitations

This research, while contributing significant insights into the field of short-term trip production prediction, acknowledges certain limitations that should be considered when interpreting the results and applying the methodologies.

1. The primary data source for this study is derived from GSM traces provided by a major Dutch telecommunications company, Vodafone. While these traces offer valuable insights into travel patterns, they inherently come with significant limitations that might affect the study's outcomes. The data covers approximately one-third of the Dutch population, primarily Vodafone users. This coverage limitation potentially introduces biases, as the user demographic may not accurately represent the wider population's travel behavior, especially across different socioeconomic groups. Also, the method used to infer trip productions from GSM data relies on assumptions about travel modes, which may not always hold true. The accuracy of these inferences can be affected by the granularity of the data and the algorithms used to interpret signal patterns, possibly leading to misclassifications (e.g., distinguishing between travel by public transport and private vehicles). Furthermore, the accuracy of GSM-based studies is highly dependent on the spatial distribution and density of network cells. Areas with sparse cellular coverage or where cell towers are unevenly distributed can lead to significant gaps in data, affecting the reliability of trip estimations. Additionally, the data's temporal coverage (e.g., time of day, day of the week) can also influence accuracy. For instance, trip productions during peak traffic hours might be overrepresented due to higher mobile phone usage, skewing the understanding of typical travel patterns. Given these constraints, particularly the black-box nature of the processed GSM data, this study acknowledges the limitations in the accuracy and reliability of the trip production values derived.

2. The scope of this study is geographically limited to the Netherlands. While this provides detailed insights into Dutch urban and transportation network, the findings might not directly translate to other regions with different geographic, demographic, and infrastructural characteristics.

3. The study adopts LSTM integrated with GCN as the primary predictive model. This choice, while based on the model's ability to handle spatial-temporal correlations and variations, does not imply it is the optimum method for all trip production prediction scenarios. The field offers a variety of other data-driven models that might provide different levels of effectiveness depending on specific use cases.

4. Another limitation of our study is the lack of comparison with a broader range of state-of-the-art models. While our focus was on demonstrating the benefits of integrating spatial adjacency into LSTM models, future research should include comparisons not only with models like T-GCN (Zhao et al., 2019) and STGCN (Yu et al., 2017) but also with advanced graph-based architectures that incorporate dynamic graphs (Wang et al., 2024), transformer-based GCNs (Yan & Ma, 2021), and hierarchical GCNs (HGCNs) (Zhang et al., 2021). Such evaluations would provide a more complete perspective on how effectively our approach captures spatial dependencies relative to more complex models.

5. The decision to use K-means clustering for residual analysis was guided by its simplicity and effectiveness. However, this approach might not capture the intricacies of more

complex, non-linear data distributions compared to other advanced clustering algorithms.

6. The LSTM+GCN model, though efficient in handling large-scale data, involves a certain level of computational complexity. This aspect might limit its applicability in environments with constrained computational resources.

7. The findings and conclusions drawn from this study, particularly those related to the residual analysis and socio-spatial feature associations, are based on the specific context of the Netherlands. The transferability of these insights to other contexts requires a separate study.

8. A limitation of our study is the absence of more granular socioeconomic data, such as income levels, education levels, and job types, in our analysis. Incorporating these detailed socioeconomic variables could potentially enhance the model's predictive capabilities and provide deeper insights into trip generation patterns. Future research could explore the integration of such data to analyze their impact on trip generation forecasts.

9. Another area for future research is the interpretability of the GCN+LSTM model. Employing techniques such as attention mechanisms, feature importance analysis, or explainable AI methods could provide deeper insights into the model's decision-making process and identify key influencing factors, enhancing transparency and trust in the model's predictions.

10. While our model shows promise, practical deployment for real-time prediction adjustments involves challenges not addressed in this study. Future work could focus on optimizing the model for real-time applications, exploring techniques to improve computational efficiency, and evaluating its performance in operational environments. Addressing these considerations would enhance the model's practicality for policymakers and transportation planners.

11. While we discuss certain computational aspects of our approach, such as training a single model for all TAZs instead of one model per TAZ, our primary contribution lies in demonstrating the incremental predictive accuracy gained by integrating spatial adjacency. We did not conduct a formal theoretical computational complexity analysis or extensive runtime evaluations. Future research could focus on detailed computational profiling, optimizing the method for real-time deployment, and comparing its computational performance against more computationally efficient models.

12. Our residual analysis utilized K-means clustering to identify patterns; however, more sophisticated methods such as hierarchical clustering, DBSCAN, or spatial econometric approaches might provide deeper insights into residual patterns. Future research could explore these methods to better understand the factors contributing to prediction inaccuracies and potentially improve upon the results of our existing analysis.

To sum it up, this study's primary contributions lie in the multi-scale analysis of trip production prediction and the exploration of residual patterns and socio-spatial features across different spatial scales. While acknowledging the aforementioned limitations, the research provides a foundational framework that can be built upon and adapted for further studies in this domain.

Future research could address these limitations by incorporating a broader data scope, exploring alternative predictive models, and adapting the methodologies to diverse geographical and urban contexts.

5.4 Conclusion

The first purpose of this research was to assess the association between temporal production patterns and spatial urbanization. Taken together, the results of this research support the idea that using land-use characteristics can improve the task of dynamic travel production prediction. In fact, observing trip production clusters in space and comparing them with the distribution (i.e., density) of land-use characteristics suggested different urbanization levels for each cluster: urban, rural, and mixed-level. An urban area mainly presents the city center with a high density of urban facilities, reveals a sharp afternoon production peak indicating more work-related activities. While a rural area shows more distinct production values in the morning, suggesting more residential land use. Mixed levels display other patterns in time and space. Further analysis of the mixed-level areas shows a more complex relationship between temporal heterogeneity and spatial characteristics. Population density seems to impose additional uncertainty on the temporal patterns. All in all, feature selection and Spatial and temporal discretization play essential roles in identifying the dominant trip production patterns.

Another objective of this study was to investigate the prediction errors, residual patterns, and their association with socio-spatial features at different spatial scales. The study's results showed a correlation between production variability and prediction error. The study also found that residuals peaked on the shoulders of production peaks rather than during peak hours. Additionally, the residual pattern of two TAZs might differ even if they have the same production patterns, which suggests that latent features are causing such differences. The study explored a range of features to identify the most important socio-spatial features contributing to identifying residual patterns at each spatial scale. The results indicated that under the lowest level of abstraction, i.e., 4-digit postal code zones, a combination of land-use and demographic information contributes to their residual patterns. However, as the spatial scale increases towards the municipality level, land use plays a more critical role in identifying the residual pattern, i.e., prediction errors. These findings provide valuable insights into identifying residual patterns and the contribution of socio-spatial features to such patterns, which can inform the development of more accurate and effective travel demand prediction models.

Our analysis of prediction errors, residual patterns, and their association with socio-spatial features helps identify areas where certain demographic or land-use characteristics contribute to demand variability. Policymakers can leverage this information to implement targeted interventions, such as adjusting land-use policies or enhancing transportation services in areas where prediction errors are linked to specific socio-spatial factors, ultimately better meeting the travel demands of different regions.

Chapter 6

Conclusions, implications and recommendations

In this final chapter, we reflect on the significance of the studies conducted in this thesis, addressing the main research question: ‘What are the spatial and temporal patterns of travel demand considering the input data quality, spatio-temporal context, the objective spatial scale, and the socio-spatial characteristics?’ This question, stemming from the identified knowledge gaps in the introduction, has guided our investigation and shaped the methodologies applied. We start with a summary of our key findings. Next, we discuss the implications of our findings for practitioners, policymakers, and society and identify potential opportunities for extending this research. Finally, we offer recommendations for future work in this area.

6.1 Summary of Key Findings

This thesis undertook a comprehensive exploration of spatial and temporal patterns of travel demand. By examining the interaction between data quality, spatial scale, and socio-spatial characteristics, we have identified insights crucial for estimating and predicting travel demand. Key findings include:

Chapter 2 demonstrates that GSM data provides a basis for understanding user locations over time, but additional information is needed to accurately discern stay locations. A Kernel-based approach proves effective for this, even when utilizing data from as little as one percent of the population. The temporal resolution of GSM data was highlighted as critical, with larger polling intervals adversely affecting the accuracy of OD matrices. This established a connection between temporal data characteristics and the threshold for identifying stay and pass-by locations, with findings indicating that larger polling intervals (PI) should not exceed the duration threshold for identifying stays to maintain OD matrix integrity. An optimal PI exists, enhancing structural similarity to ground-truth data, calculated to be just over half the minimum stay duration. However, short-duration activities remain challenging to detect, and unusual travel behaviors are prone to misclassification.

Chapter 3 describes an empirical analysis of trip production patterns using aggregated GSM data from the Netherlands, clustered into distinct temporal patterns using deep learning methods. The study identified three primary clusters with distinctly different patterns throughout the day and week, which correlate with urbanization levels (Urban, Rural, and Mixed). Utilizing an OVR-SMOTE-XGBoost ensemble, the research validates how accurate trip production patterns can be predicted based on land-use characteristics. These findings are key to improving data-driven demand estimation and transportation policies. Although input data are processed and aggregated derivatives of GSM traces due to privacy reasons, the study confirms its utility to represent both regular and irregular travel patterns, providing a quantitative approach to understanding travel behavior and aiding policy making.

Chapter 4 concludes that the integration of a graph convolutional neural network (GCN) with a long-short-term memory network (LSTM) presents an effective method for predicting large-scale trip production, taking into account spatial-temporal heterogeneity and adjacency of mobility zones. This integration reduces computational demands as it uses a single model for the entire network instead of separate models for each zone. The study found that accounting for spatial adjacency improves prediction accuracy in all three studied spatial scales (namely *Province*, *Municipality*, and *Traffic Analysis Zone*) while reducing computational time. Furthermore, it observed that larger spatial scales, which aggregate data and increase abstraction, can lead to more accurate predictions despite introducing some spatial uncertainty. This finding suggests that aggregation may lead to more regular, and hence predictable patterns. Although the research primarily addresses motor-vehicle trip production, the method is mode-agnostic, suggesting wider applicability.

Building on Chapter 3, **Chapter 5** presents a nuanced examination of trip production, revealing that mixed-level cluster areas show intricate interplays between temporal heterogeneity and spatial characteristics, with population density introducing additional uncertainties in temporal patterns. The analysis highlights the importance of careful feature selection and the discretization of spatial and temporal data in describing the prevailing trip production patterns. The study progresses by assessing prediction errors and residual patterns in trip production forecasts, linking these to key demographic and land-use characteristics of Traffic Analysis Zones across various spatial scales. This comprehensive analysis identified specific socio-spatial factors, such as land-use diversity, employment density, and population demographics, that significantly influence trip production variance. By quantifying the impact of these factors on different spatial scales, the study provides insights that can be directly applied to improve the accuracy of demand prediction models tailored to various levels of spatial aggregation. Utilizing GSM-derived motor vehicle Origin-Destination (OD) matrices from the Netherlands in March 2017, the research focuses on motor-vehicle trip production behaviors.

The research begins with discerning the association between temporal production patterns and levels of urbanization. The trip production clusters observed in space, compared with the distribution of land use types, revealed clear distinctions in urbanization between the urban, rural and mixed clusters. Metropolitan areas showed pronounced afternoon peaks in trip production, reflecting work-centric activity, while rural areas displayed morning peaks, suggesting residential tendencies. Mixed-level areas, however, demonstrated varied patterns both temporally and spatially, with population density complicating the temporal trends. The critically important roles of feature selection and the discretization of data in space and time are highlighted in isolating key trip production patterns.

The study also delved into the analysis of prediction errors and the patterns of residuals, exploring their relationship with socio-spatial features at different spatial scales. It revealed that the variability of trip production is correlated with prediction errors and that the residuals tend to spike around, but not during, the production peaks. It was found that identical production patterns across traffic analysis zones could yield different residual patterns, hinting at the influence of underlying latent features. At the finer spatial resolution of 4-digit postal codes, both land-use and demographic factors were significant contributors to residual patterns. Conversely, at broader scales such as municipalities, land-use became increasingly dominant in explaining the residuals. These insights are instrumental in recognizing residual patterns and the influence of socio-spatial attributes on them, which could significantly enhance the relevance and applicability of travel demand forecasting models.

6.2 Conclusion

This thesis set out to explore the spatial and temporal patterns of travel demand by considering the input data quality, spatio-temporal context, spatial scale, and socio-spatial characteristics. By addressing the identified knowledge gaps, our research has advanced the understanding of how these factors influence travel demand patterns and their estimation and prediction.

Firstly, we tackled the lack of standardized approaches for processing GSM data for origin-destination (OD) estimation. Our findings demonstrate that the quality of travel demand estimation depends not merely on data resolution or accuracy but on data that is fit for the specific purpose. By developing a kernel-based approach for interpreting GSM data with minimal assumptions (Chapter 2), we showed that appropriate preprocessing can significantly enhance the reliability of OD matrices. While the GSM data itself was limited, the training set, based on detailed travel diaries from a small sample of users, proved sufficient to refine and validate the model. This insight shifts the focus from acquiring high-resolution data to ensuring that data collection and processing methods align with the modeling objectives.

Secondly, we addressed the gap in understanding spatial and temporal patterns in trip production. Through empirical analysis using deep learning methods (Chapter 3), we identified distinct temporal patterns of trip production that correlate with levels of urbanization. This revealed that collective travel demand exhibits regular patterns that can be systematically classified, providing insights into aggregated travel behaviors crucial for policy-making and transport planning. This finding enhances our understanding of how urban form influences travel behavior at a collective level, offering practical implications for targeted interventions in urban and transportation systems.

Thirdly, we investigated the relationship between spatial scale and the prediction of trip production. Our research indicates that larger spatial scales, which involve greater data aggregation, can lead to more accurate trip production predictions due to the smoothing of variability and amplification of regular patterns (Chapter 4). However, this comes at the expense of losing finer spatial detail. This highlights the importance of carefully selecting the spatial scale in travel demand modeling to strike a balance between prediction accuracy and the granularity required for specific policy or planning objectives.

Lastly, we explored the correlation between socio-spatial characteristics and travel demand patterns at different spatial scales. We found that land-use diversity and demographic factors significantly influence trip production variance and prediction errors, with their impact varying across spatial scales (Chapter 5). This finding emphasizes the necessity of integrating socio-spatial features into travel demand models to improve their predictive capabilities and ensure that the models remain relevant for practical applications in diverse geographic and urban contexts.

Reflecting on these findings, it becomes evident that while our initial research questions were pertinent, they might benefit from reformulation in light of the complexities uncovered. For instance, the relationship between spatial scale and socio-spatial characteristics warrants further investigation to better capture the nuances of collective travel behaviors across different contexts. Moreover, our work underscores that improving travel demand estimation requires not only better data but also a deeper understanding of how factors such as urbanization, land use, and demographics interplay to shape aggregated travel behaviors at varying spatial and temporal scales.

In conclusion, this thesis advances the field of travel demand estimation by offering empirical evidence and novel methodologies that provide a richer understanding of spatial and temporal travel demand patterns. By demonstrating the significance of fit-for-purpose data, uncovering the influence of urbanization on temporal patterns, highlighting the trade-offs inherent in spatial scale selection, and quantifying the role of socio-spatial characteristics, we contribute a framework for enhancing the accuracy and applicability of travel demand models. These insights have direct implications for transport planners and policymakers, enabling more effective, data-driven strategies for managing and predicting travel demand in diverse urban contexts.

6.3 Limitations and Future Research Directions

Building on the insights from this thesis, we propose two impactful future research directions that extend the application of our developed frameworks to address areas of urban sustainability and adaptive planning for special events. These directions not only showcase the potential utility of our research but also suggest pathways for practical implementation in urban planning and policy formulation.

- **Sustainable Urban Development Framework:** Our study, by analyzing the relationship between trip production patterns and socio-spatial characteristics, provides foundational insights into how urban development affects travel behaviors. Specifically, we have shown how land-use diversity and population density influence temporal patterns of trip production. Building on these insights, future research can develop a comprehensive framework that integrates our predictive models with energy consumption and emission factors associated with different travel patterns. By linking changes in temporal traffic patterns to energy use and environmental impacts, this framework can help quantify the ecological consequences of urban expansion. Policymakers can then utilize this information to design urban development strategies that promote sustainable transportation modes, reduce reliance on private vehicles, and minimize environmental footprints.

- **Adaptive Planning for Special Events:** Our research introduces a methodological framework that can be used to examine how special events alter typical travel behaviors and trip production patterns. This framework arms transport planners with the ability to develop adaptive transport strategies. These strategies can mitigate congestion and enhance mobility during events, improving urban livability.

While the findings of this research represent a significant advancement in transport planning and demand estimation, several inherent limitations must be acknowledged. These limitations not only serve as caveats for the present work but also as potential avenues for future research.

6.3.1 Data Representativeness and Privacy Concerns

- **Synthetic Data Utilization:** The reliance on synthetic data, despite its utility for privacy preservation and validation, raises questions regarding the direct transferability of the models to real-world scenarios. Future studies should aim to apply the developed frameworks to empirical datasets while navigating the associated privacy concerns and ethical implications.

- **Processing and Transformation Limitations:** The initial processing and aggregation of mobile phone datasets were conducted externally, which may lead to information loss.

Detailed methodologies used by third parties for data transformation and derivation of OD matrices were not fully transparent, potentially obscuring some underlying data characteristics. Ensuring full traceability of data transformation processes will enhance the reliability of future research outcomes.

6.3.2 Geospatial and Temporal Considerations

- **TAZ and Base Station Coverage Discrepancies:** Assumptions regarding the correspondence between TAZs and base station coverage areas might not hold true universally. This simplification could impact the fidelity of OD matrices, particularly in non-urban areas. Future studies should incorporate precise base station coverage data to refine the spatial granularity of demand estimation.

- **Short-Trip Exclusion:** The exclusion of short-distance trips potentially overlooks significant urban mobility patterns. Incorporating these trips would present a more comprehensive view of urban dynamics and enhance model sensitivity to all forms of mobility, including active transportation modes.

- **Temporal Data Range:** The study's confinement to a single-month data collection period limits the temporal generalizability of the findings. Longitudinal data spanning various seasonal and event-specific contexts would be instrumental in verifying the robustness of demand estimation over time.

6.3.3 Methodological Refinements

- **Bayesian Model Simplifications:** The naive approach of the Bayesian model adopted in activity category detection could be refined. Expanding the dataset size and complexity could help in transitioning towards a fully Bayesian model, capturing the intricate correlations between variables for more nuanced inference.

- **Positioning and OD Accuracy:** The study focused predominantly on the impact of positioning intervals, leaving positioning accuracy and disturbances, such as ping-pong handovers, less explored. Future research should aim to integrate these factors, potentially through empirical studies, to provide a holistic understanding of their effects on OD accuracy.

- **OD Matrix Scaling and Transport Mode Consideration:** The transition from sample OD matrices to population-wide estimations and the mode of transport consideration were outside the scope of this research. Further methodological advancements are needed to incorporate comprehensive scaling techniques and mode differentiation, enhancing the model's practical applicability.

6.3.4 Technical and Computational Challenges

- **GCN Feature Limitations:** The study's Graph Convolutional Networks (GCNs) primarily utilized TAZ adjacency. Future iterations should integrate additional socio-spatial features to account for area heterogeneity, possibly leading to more precise and contextually informed predictions.

- **Urbanization and Heterogeneity:** The impact of spatial urbanization levels on trip production was identified as significant but not fully integrated into the predictive models. Research expanding on these socio-spatial dynamics could yield a richer understanding and prediction of travel demand patterns.

These limitations underscore the complex and multifaceted nature of transport planning and demand estimation. As such, they pave the way for future research to refine these initial models, adapt them to real-world datasets, and explore the nuances of urban mobility with greater precision and granularity.

In sum, this thesis lays a path forward for enhancing travel demand prediction through a meticulous investigation into the interaction between data integrity, spatial-temporal patterns, and socio-spatial factors. The insights garnered from this study provide a strategic blueprint for academia and industry professionals in urban and transportation planning, underscoring the potential for data-driven methodologies to reshape the landscape of mobility analytics and policy development. By addressing the outlined limitations and capitalizing on the identified research avenues, future work has the opportunity to refine these predictive models further, ultimately fostering more sustainable and efficient urban transportation systems.

Chapter A

A.1 XGBoost ensemble model

This part discusses more details on the XGBoost algorithm (we refer the readers to Chen & Guestrin (2016) for more details).

XGBoost is an ensemble of decision trees; it consists of sequentially developed decision trees where each tree works to improve the performance of the prior tree (Srivastava et al., 2022). Ensemble methods generally try to reduce the bias or variance of several weak learners by combining them into a strong learner (i.e., a learner with low bias and variance). Boosting is an ensemble method in which weak learners are fitted sequentially and aggregated to the ensemble model. In each step, the training set is updated to focus more on the weakness of the current ensemble. In other words, each model in the sequence does the fitting by giving higher weight to misclassified data points. If the weak learner of each step depends on the gradient direction of the loss function at each step, this method is also called Gradient Boosting Machines (GBM) (Friedman, 2002). The advantage of XGBoost over non-extreme gradient boosting methods is the regularization term in the loss function, which helps prevent over-fitting.

Assume our data set is $\chi = \{(x_i, y_i) : i = 1, ..., n,\ x_i \in \mathbb{R}^m,\ y_i \in \mathbb{R}\}$; hence we have n observations each having m features corresponding to their associated label y. Then $\hat{y}_i$ can be defined as a result of an ensemble, with T additive functions, represented by the generalized model as follows:

$$\hat{y}_i = \Phi(x_i) = \sum_{t=1}^{T} f_t(x_i) \;, \tag{A.1}$$

where f_t is a decision tree, and $f_t(x_i)$ is the score given by the $t-th$ decision tree to the $i-th$ data point. The objective function that needs to be minimized to select the function f_t consists of two terms of *training loss*, $L(y_i, \hat{y}_i)$ and *regularization*, $\Omega(f_t)$:

$$obj(\Phi) = \sum_i L(y_i, \hat{y}_i) + \sum_t \Omega(f_t) \tag{A.2}$$

The *training loss*, L, estimates the model's goodness of fit based on the training data. A common form of L for classification, which is used in this research, is the logistic loss (i.e., binary logistic) for $y \in \{0, 1\}$ (Bishop & Nasrabadi, 2006):

$$L_{Logistic} = -\frac{1}{N} \sum_{i=1}^{N} (y_i log(p_i) + (1 - y_i) log(1 - p_i)) \;, \tag{A.3}$$

where y_i is the true value, $p_i \in [0, 1]$ denotes the probability prediction, and N is the number of samples. An ideal classifier has a logistic loss close to zero.

In order to prevent the model from becoming too complex, Ω applies the penalty as follows:

$$\Omega(f_t) = \gamma T + \frac{1}{2}\lambda||\omega||^2 \ , \tag{A.4}$$

Where γ controls the penalty for the number of leaves, T and λ is the parameter for controlling the magnitude of leaf weights ω in the decision tree. The purpose of having the regularization term in the objective function is to simplify the model and prevent over-fitting.

Learning the tree structure is more difficult than the traditional optimization problem, where you can simply take the gradient. In other words, training all the trees simultaneously is not a straightforward task; therefore, XGBoost uses an additive method that optimizes the learned tree and adds a tree at each step. In the $t-th$ iteration, we need to add the following f_t, which minimizes the objective function:

$$obj^t = \sum_{i=1}^{n} L(y_i, \hat{y}_i^{t-1} + f_t(x_i)) + \Omega(f_t). \tag{A.5}$$

This function can be simplified and approximated by the Taylor expansion:

$$obj^t \simeq \sum_{i=1}^{n} [L(y_i, \hat{y}_i^{t-1}) + g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i)] + \Omega(f_t) \ , \tag{A.6}$$

Where the functions g_i and h_i, the first and second order gradient of the loss function, are defined as follows:

$$g_i = \partial_{\hat{y}_i^{(t-1)}} L(y_i, \hat{y}_i^{(t-1)}) \tag{A.7}$$

$$h_i = \partial^2_{\hat{y}_i^{(t-1)}} L(y_i, \hat{y}_i^{(t-1)}). \tag{A.8}$$

We can rewrite Equation A.6 by expanding Ω and find the optimal output value (i.e., weight) $\tilde{\omega}_j$ for leaf j as follows:

$$\tilde{\omega}_j = -\frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda} \ , \tag{A.9}$$

Where I_j is the instance set of leaf j. Replacing Equation A.9 and A.4 in A.6 gives us the following optimal value of the loss function which is used as a similarity score for measuring the quality of each tree structure:

$$\tilde{obj}^t = -\frac{1}{2} \sum_{j=1}^{T} \frac{(\sum_{i \in I_j} g_i)^2}{\sum_{i \in I_j} h_i + \lambda} + \gamma T. \tag{A.10}$$

For binary logistic loss function we use for the classification, $g_i = -(y_i - p_i)$ (i.e., the residual), and $h_i = p_i(1 - p_i)$ which can be replaced in Equation A.9 and A.10.

To simplify evaluating tree structure when adding new branches to the tree (i.e., evaluating the split candidates), a greedy algorithm is used. This algorithm starts from one single leaf and adds new branches to the tree iteratively. Therefore, after the tree splits from a given node, the formula for loss reduction (i.e., gain) is as follows:

$$obj_{split} = \frac{1}{2}\left[\frac{(\sum_{i \in I_L} g_i)^2}{\sum_{i \in I_L} h_i + \lambda} + \frac{(\sum_{i \in I_R} g_i)^2}{\sum_{i \in I_R} h_i + \lambda} - \frac{(\sum_{i \in I} g_i)^2}{\sum_{i \in I} h_i + \lambda}\right] - \gamma \ , \tag{A.11}$$

Where I_L and I_R are subsets of the available observations in the left and right nodes after the split. I is the subset of the available observations in the current node so that $I = I_L \cup I_R$. Moreover, the tree structure will continue to split if obj_{split} is positive or other criteria are met, such as the maximum depth of a tree that users need in XGBoost parameters fine-tuning.

Equation A.11 is used for finding the best split at any node and it only depends on g_i, and h_i (i.e., the first and second order gradient) of the training loss and the regularization parameter γ. Therefore, as long as the first and second-order gradient is provided, XGBoost can optimize any custom loss function.

XGBoost performs better than other tree boosting algorithms due to (i) having the regularization term for preventing the over-fitting, (ii) downscaling of each new tree by a constant parameter τ to reduce the impact of a single tree on the final model, i.e., it gives the future trees more space to improve the model while reducing the impact of the current tree. Moreover, (iii) XGBoost supports column sampling, which means each tree is built using a subset of the columns from the training dataset.

Bibliography

Alexander, L., S. Jiang, M. Murga, 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, big Data in Transportation and Traffic Engineering.

Anselin, L., D. A. Griffith (1988) Do spatial effecfs really matter in regression analysis?, *Papers in Regional Science*, 65(1), pp. 11–34.

Antipova, A., C. Wilmot (2012) Alternative approaches for reducing congestion in baton rouge, louisiana, *Journal of transport geography*, 24, pp. 404–410.

Antoniou, C., E. Chaniotakis, C. Katrakazas, A. Tirachini (2020) A better tomorrow: towards human-oriented, sustainable transportation systems, *European Journal of Transport and Infrastructure Research*, 20(4), pp. 354–361.

Arentze, T., H. Timmermans (2004) A learning-based transportation oriented simulation system, *Transportation Research Part B: Methodological*, 38(7), pp. 613–633, cited By 280.

Ashok, K., M. E. Ben-Akiva (2002) Estimation and prediction of time-dependent origin-destination flows with a stochastic mapping to path flows and link flows, *Transportation science*, 36(2), pp. 184–198.

Atluri, G., A. Karpatne, V. Kumar (2018) Spatio-temporal data mining: A survey of problems and methods, *ACM Computing Surveys (CSUR)*, 51(4), pp. 1–41.

Axhausen, K. W., T. Gärling (1992) Activity-based approaches to travel analysis: conceptual frameworks, models, and research problems, *Transport reviews*, 12(4), pp. 323–341.

Bachir, D., G. Khodabandelou, V. Gauthier, M. El Yacoubi, J. Puchinger (2019) Inferring dynamic origin-destination flows by transport mode using mobile phone data, *Transportation Research Part C: Emerging Technologies*, 101, pp. 254 – 275.

Becker, R., C. Volinsky, R. Cáceres, K. Hanson, S. Isaacman, J. M. Loh, M. Martonosi, J. Rowland, S. Urbanek, A. Varshavsky (2013) Human mobility characterization from cellular network data, *Communications of the ACM*, 56(1), p. 74.

Behara, K. N., A. Bhaskar, E. Chung (2020a) A novel approach for the structural comparison of origin-destination matrices: Levenshtein distance, *Transportation Research Part C: Emerging Technologies*, 111, pp. 513 – 530.

Behara, K. N. S., A. Bhaskar, E. Chung (2020b) Geographical window based structural similarity index for origin-destination matrices comparison, *Journal of Intelligent Transportation Systems*, 0(0), pp. 1–22.

Bengfort, B., R. Bilbro, N. Danielsen, L. Gray, K. McIntyre, P. Roman, Z. Poh, et al. (2018) Yellowbrick, URL http://www.scikit-yb.org/en/latest/.

Bianchi, F. M., A. Rizzi, A. Sadeghian, C. Moiso (2016) Identifying user habits through data mining on call data records, *Engineering Applications of Artificial Intelligence*, 54, pp. 49 – 61.

Bishop, C. M., N. M. Nasrabadi (2006) *Pattern recognition and machine learning*, vol. 4, Springer.

Bonnel, P., E. Hombourger, A.-M. Olteanu-Raimond, Z. Smoreda (2015) Passive mobile phone dataset to construct origin-destination matrix: potentials and limitations, *Transportation Research Procedia*, 11, pp. 381–398.

Botev, Z. I., J. F. Grotowski, D. P. Kroese (2010) Kernel density estimation via diffusion, *Ann. Statist.*, 38(5), pp. 2916–2957.

Bruna, J., W. Zaremba, A. Szlam, Y. LeCun (2013) Spectral networks and locally connected networks on graphs, *arXiv preprint arXiv:1312.6203*.

Brunsdon, C., A. S. Fotheringham, M. E. Charlton (1996) Geographically weighted regression: a method for exploring spatial nonstationarity, *Geographical analysis*, 28(4), pp. 281–298.

Burkhard, O., R. Ahas, E. Saluveer, R. Weibel (2017) Extracting regular mobility patterns from sparse cdr data without a priori assumptions, *Journal of Location Based Services*, 11(2), pp. 78–97.

Burkhard, O., H. Becker, R. Weibel, K. W. Axhausen (2020) On the requirements on spatial accuracy and sampling rate for transport mode detection in view of a shift to passive signalling data, *Transportation Research Part C: Emerging Technologies*, 114, pp. 99–117.

Calabrese, F., G. Di Lorenzo, L. Liu, C. Ratti (2011) Estimating origin-destination flows using mobile phone location data, *IEEE Pervasive Computing*, 10(4), pp. 36–44.

Calabrese, F., M. Diao, G. Di Lorenzo, J. Ferreira Jr, C. Ratti (2013) Understanding individual mobility patterns from urban sensing data: A mobile phone trace example, *Transportation research part C: emerging technologies*, 26, pp. 301–313.

Camps-Valls, G. (2006) *Kernel methods in bioengineering, signal and image processing*, Igi Global.

Cardozo, O. D., J. C. García-Palomares, J. Gutiérrez (2012) Application of geographically weighted regression to the direct forecasting of transit ridership at station-level, *Applied Geography*, 34, pp. 548–558.

Centraal Bureau voor de Statistiek (CBS) (2017) Gegevens per postcode, https://www.cbs.nl/nl-nl/dossier/nederland-regionaal/geografische-data/gegevens-per-postcode, accessed: 1 February 2023.

Cervero, R. (1996) Mixed land-uses and commuting: Evidence from the american housing survey, *Transportation research part a: policy and practice*, 30(5), pp. 361–377.

Chen, C., L. Bian, 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.

Chen, G., S. Hoteit, A. C. Viana, M. Fiore, C. Sarraute (2018) Enriching sparse mobility information in call detail records, *Computer Communications*, 122, pp. 44 – 58.

Chen, T., C. Guestrin (2016) Xgboost: A scalable tree boosting system, in: *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pp. 785–794.

Cheng, S., F. Lu, P. Peng, S. Wu (2019) Multi-task and multi-view learning based on particle swarm optimization for short-term traffic forecasting, *Knowledge-Based Systems*, 180, pp. 116–132.

Cheng, T., J. Haworth, B. Anbaroglu, G. Tanaksaranond, J. Wang (2013) Spatiotemporal data mining.

Chung, J., C. Gulcehre, K. Cho, Y. Bengio (2014) Empirical evaluation of gated recurrent neural networks on sequence modeling, *arXiv preprint arXiv:1412.3555*.

Cohn, R., E. Holm (2021) Unsupervised machine learning via transfer learning and k-means clustering to classify materials image data, *Integrating Materials and Manufacturing Innovation*, pp. 1–14.

Defferrard, M., X. Bresson, P. Vandergheynst (2016) Convolutional neural networks on graphs with fast localized spectral filtering, *Advances in neural information processing systems*, 29.

Demchenko, Y., P. Grosso, C. De Laat, P. Membrey (2013) Addressing big data issues in scientific data infrastructure, in: *2013 International conference on collaboration technologies and systems (CTS)*, IEEE, pp. 48–55.

Demirbas, K. (1988) Maximum a posteriori approach to object recognition with distributed sensors, *IEEE Transactions on Aerospace and Electronic Systems*, 24(3), pp. 309–313.

Demissie, M. G., S. Phithakkitnukoon, L. Kattan (2019) Trip distribution modeling using mobile phone data: Emphasis on intra-zonal trips, *IEEE Transactions on Intelligent Transportation Systems*, 20(7), pp. 2605–2617.

Deng, J., W. Dong, R. Socher, L.-J. Li, K. Li, L. Fei-Fei (2009) Imagenet: A large-scale hierarchical image database, in: *2009 IEEE conference on computer vision and pattern recognition*, Ieee, pp. 248–255.

Deng, M., W. Yang, Q. Liu, R. Jin, F. Xu, Y. Zhang (2018) Heterogeneous space–time artificial neural networks for space–time series prediction, *Transactions in GIS*, 22(1), pp. 183–201.

Diao, Z., X. Wang, D. Zhang, Y. Liu, K. Xie, S. He (2019) Dynamic spatial-temporal graph convolutional neural networks for traffic forecasting, in: *Proceedings of the AAAI conference on artificial intelligence*, vol. 33, pp. 890–897.

Dong, H., M. Wu, X. Ding, L. Chu, L. Jia, Y. Qin, X. Zhou (2015) Traffic zone division based on big data from mobile phone base stations, *Transportation Research Part C: Emerging Technologies*, 58, pp. 278–291.

Eftekhar, Z., S. Behrouzi, Panchamy Krishnakumari, A. Pel, H. van Lint (2024a) The codes associated with the publication: The role of spatial features and adjacency in data-driven short-term prediction of trip production: An exploratory study in the netherlands, URL https://data.4tu.nl/datasets/51fa919d-bc31-4e55-92ac-6fc67ff50fcc/1.

Eftekhar, Z., A. Pel, H. van Lint (2024b) The input data associated with the publication: The role of spatial features and adjacency in data-driven short-term prediction of trip production: An exploratory study in the netherlands, URL https://data.4tu.nl/datasets/835e7093-024c-4bed-b7e5-28f1582f5998/1.

Ermagun, A., D. Levinson (2018) Spatiotemporal traffic forecasting: review and proposed directions, *Transport Reviews*, 38(6), pp. 786–814.

Fekih, M., T. Bellemans, Z. Smoreda, P. Bonnel, A. Furno, S. Galland (2021a) A data-driven approach for origin–destination matrix construction from cellular network signalling data: a case study of lyon region (france), *Transportation*, 48, pp. 1671–1702.

Fekih, M., L. Bonnetain, A. Furno, P. Bonnel, Z. Smoreda, S. Galland, T. Bellemans (2021b) Potential of cellular signaling data for time-of-day estimation and spatial classification of travel demand: a large-scale comparative study with travel survey and land use data, *Transportation Letters*, pp. 1–19.

Fekih, M., L. Bonnetain, A. Furno, P. Bonnel, Z. Smoreda, S. Galland, T. Bellemans (2022) Potential of cellular signaling data for time-of-day estimation and spatial classification of travel demand: a large-scale comparative study with travel survey and land use data, *Transportation Letters*, 14(7), pp. 787–805.

Ferguson, E. (1990) Transportation demand management planning, development, and implementation, *Journal of the American Planning Association*, 56(4), pp. 442–456.

Fotheringham, A. S., M. E. Charlton, C. Brunsdon (1998) Geographically weighted regression: a natural evolution of the expansion method for spatial data analysis, *Environment and planning A*, 30(11), pp. 1905–1927.

Friedman, J. H. (2002) Stochastic gradient boosting, *Computational statistics & data analysis*, 38(4), pp. 367–378.

Fu, X., G. Yu, Z. Liu (2021) Spatial–temporal convolutional model for urban crowd density prediction based on mobile-phone signaling data, *IEEE Transactions on Intelligent Transportation Systems*, 23(9), pp. 14661–14673.

Gärling, T., D. Eek, P. Loukopoulos, S. Fujii, O. Johansson-Stenman, R. Kitamura, R. Pendyala, B. Vilhelmson (2002) A conceptual analysis of the impact of travel demand management on private car use, *Transport Policy*, 9(1), pp. 59–70.

Ghasri, M., T. H. Rashidi, S. T. Waller (2017) Developing a disaggregate travel demand system of models using data mining techniques, *Transportation Research Part A: Policy and Practice*, 105, pp. 138–153.

Groves, R. M. (2006) Nonresponse rates and nonresponse bias in household surveys, *Public Opinion Quarterly*, 70(5), pp. 646–675.

Guo, S., Y. Lin, N. Feng, C. Song, H. Wan (2019) Attention based spatial-temporal graph convolutional networks for traffic flow forecasting, in: *Proceedings of the AAAI conference on artificial intelligence*, vol. 33, pp. 922–929.

Hajek, J. J. (1977) Optimal sample size of roadside-interview origin-destination surveys, *TRB Transportation Research Board*.

Handy, S. (1996) Methodologies for exploring the link between urban form and travel behavior, *Transportation Research Part D: Transport and Environment*, 1(2), pp. 151–165.

Hartigan, J. A., M. A. Wong (1979) Algorithm as 136: A k-means clustering algorithm, *Journal of the royal statistical society. series c (applied statistics)*, 28(1), pp. 100–108.

He, H., E. A. Garcia (2009) Learning from imbalanced data, *IEEE Transactions on Knowledge and Data Engineering*, 21(9), pp. 1263–1284.

Hochreiter, S., J. Schmidhuber (1997) Long short-term memory, *Neural computation*, 9(8), pp. 1735–1780.

Hong, J.-H., S.-B. Cho (2008) A probabilistic multi-class strategy of one-vs.-rest support vector machines for cancer classification, *Neurocomputing*, 71(16), pp. 3275–3281, advances in Neural Information Processing (ICONIP 2006) / Brazilian Symposium on Neural Networks (SBRN 2006).

Huang, H., Y. Cheng, R. Weibel (2019) Transport mode detection based on mobile phone network data: A systematic review, *Transportation Research Part C: Emerging Technologies*, 101, pp. 297–312.

Hubert, L., P. Arabie (1985) Comparing partitions, *Journal of classification*, 2(1), pp. 193–218.

Iglovikov, V., A. Shvets (2018) Ternausnet: U-net with vgg11 encoder pre-trained on imagenet for image segmentation, *arXiv preprint arXiv:1801.05746*.

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

Jiang, S., J. Ferreira, 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.

Jones, M. C., J. S. Marron, S. J. Sheather (1992) Progress in data-based bandwidth selection for kernel density estimation, Tech. rep., North Carolina State University. Dept. of Statistics.

Kaur, T., T. K. Gandhi (2020) Deep convolutional neural networks with transfer learning for automated brain image classification, *Machine Vision and Applications*, 31(3), pp. 1–16.

Khalesian, M., A. Furno, L. Leclercq (2024) Improving deep-learning methods for area-based traffic demand prediction via hierarchical reconciliation, *Transportation Research Part C: Emerging Technologies*, 159, p. 104410.

Kipf, T. N., M. Welling (2016) Semi-supervised classification with graph convolutional networks, *arXiv preprint arXiv:1609.02907*.

Krishnakumari, P., H. van Lint, T. Djukic, O. Cats (2019) A data driven method for od matrix estimation, *Transportation Research Procedia*, 38, pp. 139–159.

Kruskal, J. B. (1964) Nonmetric multidimensional scaling: a numerical method, *Psychometrika*, 29(2), pp. 115–129.

Kuwahara, M., E. C. Sullivan (1987) Estimating origin-destination matrices from roadside survey data, *Transportation Research Part B: Methodological*, 21(3), pp. 233–248.

Levin, M. W., H. Smith, S. D. Boyles (2019) Dynamic four-step planning model of empty repositioning trips for personal autonomous vehicles, *Journal of Transportation Engineering, Part A: Systems*, 145(5), p. 04019015.

Li, A., K. W. Axhausen (2019) Comparison of short-term traffic demand prediction methods for transport services, *Arbeitsberichte Verkehrs-und Raumplanung*, 1447.

Li, S., X. Liang, M. Zheng, J. Chen, T. Chen, X. Guo (2023) How spatial features affect urban rail transit prediction accuracy: a deep learning based passenger flow prediction method, *Journal of Intelligent Transportation Systems*, pp. 1–12.

Liang, Y., S. Ke, J. Zhang, X. Yi, Y. Zheng (2018) Geoman: Multi-level attention networks for geo-sensory time series prediction., in: *IJCAI*, vol. 2018, pp. 3428–3434.

Lin, J.-J., T.-Y. Shin (2008) Does transit-oriented development affect metro ridership? evidence from taipei, taiwan, *Transportation Research Record*, 2063(1), pp. 149–158.

Lisboa, P. J., T. A. Etchells, I. H. Jarman, S. J. Chambers (2013) Finding reproducible cluster partitions for the k-means algorithm, *BMC bioinformatics*, 14(Suppl 1), p. S8.

Liu, K., Z. Chen, T. Yamamoto, L. Tuo (2022) Exploring the impact of spatiotemporal granularity on the demand prediction of dynamic ride-hailing, *IEEE Transactions on Intelligent Transportation Systems*.

Liu, S. Y., S. Liu, Y. Tian, Q. L. Sun, Y. Y. Tang (2021) Research on forecast of rail traffic flow based on arima model, in: *Journal of Physics: Conference Series*, vol. 1792, IOP Publishing, p. 012065.

Liu, X., S. Cheng, X. Zhang, X. Yang, T. B. Nguyen, S. Lee (2012) Unsupervised segmentation in 3d planar object maps based on fuzzy clustering, in: *2012 Eighth International Conference on Computational Intelligence and Security*, pp. 364–368.

Lo, H.-P., C.-P. Chan (2003) Simultaneous estimation of an origin–destination matrix and link choice proportions using traffic counts, *Transportation Research Part A: Policy and Practice*, 37(9), pp. 771 – 788.

Ma, X., J. Zhang, C. Ding, Y. Wang (2018) A geographically and temporally weighted regression model to explore the spatiotemporal influence of built environment on transit ridership, *Computers, Environment and Urban Systems*, 70, pp. 113–124.

Maat, K., H. Timmermans (2006) Influence of land use on tour complexity: a dutch case, *Transportation Research Record*, 1977(1), pp. 234–241.

Mamei, M., N. Bicocchi, M. Lippi, S. Mariani, F. Zambonelli (2019) Evaluating origin–destination matrices obtained from cdr data, *Sensors*, 19(20), p. 4470.

Meppelink, J., J. Van Langen, A. Siebes, M. Spruit (2020) Beware thy bias: Scaling mobile phone data to measure traffic intensities, *Sustainability*, 12(9), p. 3631.

Mohanty, S., A. Pozdnukhov (2020) Dynamic origin-destination demand estimation from link counts, cellular data and travel time data, *Transportation Research Procedia*, 48, pp. 1722 – 1739, recent Advances and Emerging Issues in Transport Research – An Editorial Note for the Selected Proceedings of WCTR 2019 Mumbai.

Ning, C., Z. Hongyi (2016) An optimizing algorithm of non-linear k-means clustering, *International Journal of Database Theory and Application*, 9(4), pp. 97–106.

OpenStreetMap contributors, . (2017) Planet dump retrieved from https://planet.osm.org, URL https://www.openstreetmap.org.

Ortúzar, J. d. D., L. G. Willumsen (2011a) *Assignment*, chap. 1, John Wiley & Sons, Ltd, pp. 1–27.

Ortúzar, J. d. D., L. G. Willumsen (2011b) *Assignment*, chap. 7, John Wiley & Sons, Ltd, pp. 227–268.

Ortúzar, J. d. D., L. G. Willumsen (2011c) *Trip Distribution Modelling*, chap. 5, John Wiley & Sons, Ltd, pp. 175–206.

Pan, Z., Y. Liang, W. Wang, Y. Yu, Y. Zheng, J. Zhang (2019) Urban traffic prediction from spatio-temporal data using deep meta learning, in: *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pp. 1720–1730.

Pel, A., M. Bliemer, S. Hoogendoorn (2011) Modelling traveller behaviour under emergency evacuation conditions, *European Journal of Transport and Infrastructure Research*, 11(2).

Pendyala, R. M., Y.-C. Chiu, P. Waddell, M. Hickman, K. C. Konduri, B. Sana (2010) The design of an integrated model of the urban continuum—location choices, activity-travel behavior, and dynamic traffic patterns, in: *the 12th World Congress of Transport Research, Lisbon, Portugal*.

Poteraş, C. M., M. C. Mihăescu, M. Mocanu (2014) An optimized version of the k-means clustering algorithm, in: *2014 Federated Conference on Computer Science and Information Systems*, IEEE, pp. 695–699.

Qi, Y., S. Ishak (2014) A hidden markov model for short term prediction of traffic conditions on freeways, *Transportation Research Part C: Emerging Technologies*, 43, pp. 95–111.

Qian, X., S. V. Ukkusuri, C. Yang, F. Yan (2020) Short-term demand forecasting for on-demand mobility service, *IEEE Transactions on Intelligent Transportation Systems*.

Raadsen, M. P., M. C. Bliemer, M. G. Bell (2020) Aggregation, disaggregation and decomposition methods in traffic assignment: historical perspectives and new trends, *Transportation Research Part B: Methodological*, 139, pp. 199–223.

Rajabzadeh, Y., A. H. Rezaie, H. Amindavar (2017) Short-term traffic flow prediction using time-varying vasicek model, *Transportation Research Part C: Emerging Technologies*, 74, pp. 168–181.

Rich, J., S. L. Mabit (2012) A long-distance travel demand model for europe, *European Journal of Transport and Infrastructure Research*, 12(1).

Rousseeuw, P. J. (1987) Silhouettes: a graphical aid to the interpretation and validation of cluster analysis, *Journal of computational and applied mathematics*, 20, pp. 53–65.

Rubin, V., T. Lukoianova (2013) Veracity roadmap: Is big data objective, truthful and credible?, *Advances in Classification Research Online*, 24(1), p. 4.

Russell, S., P. Norvig (2016) Artificial intelligence: a modern approach (global 3rd edition), *Essex: Pearson*, pp. 122–125.

Schneider, C. M., C. Rudloff, D. Bauer, M. C. González (2013) Daily travel behavior: lessons from a week-long survey for the extraction of human mobility motifs related information, in: *Proceedings of the 2nd ACM SIGKDD international workshop on urban computing*, pp. 1–7.

Scott, D. W. (2012) *Multivariate Density Estimation and Visualization*, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 549–569.

Scott, D. W. (2015) *Multivariate density estimation: theory, practice, and visualization*, John Wiley & Sons.

Seo, Y., M. Defferrard, P. Vandergheynst, X. Bresson (2018) Structured sequence modeling with graph convolutional recurrent networks, in: *International conference on neural information processing*, Springer, pp. 362–373.

Shamos, M. I., D. Hoey (1975) Closest-point problems, in: *16th Annual Symposium on Foundations of Computer Science (sfcs 1975)*, IEEE, pp. 151–162.

Sheather, S. J., M. C. Jones (1991) A reliable data-based bandwidth selection method for kernel density estimation, *Journal of the Royal Statistical Society: Series B (Methodological)*, 53(3), pp. 683–690.

Shen, X., Y. Zhou, S. Jin, D. Wang (2020) Spatiotemporal influence of land use and household properties on automobile travel demand, *Transportation Research Part D: Transport and Environment*, 84, p. 102359.

Silverman, B. W. (1986) *Density estimation for statistics and data analysis*, vol. 26, CRC press.

Simonoff, J. S. (2012) *Smoothing methods in statistics*, Springer Science & Business Media.

Smola, A. J., B. Schölkopf, K.-R. Müller (1998) The connection between regularization operators and support vector kernels, *Neural Networks*, 11(4), pp. 637 – 649.

Srinivasan, D., M. C. Choy, R. L. Cheu (2006) Neural networks for real-time traffic signal control, *IEEE Transactions on intelligent transportation systems*, 7(3), pp. 261–272.

Srivastava, A. K., N. Safaei, S. Khaki, G. Lopez, W. Zeng, F. Ewert, T. Gaiser, J. Rahimi (2022) Winter wheat yield prediction using convolutional neural networks from environmental and phenological data, *Scientific Reports*, 12(1), pp. 1–14.

Stopher, P. R., S. Greaves, P. Bullock (2003) Simulating household travel survey data: application to two urban areas, in: *82nd Annual Meeting of the Transportation Research Board, Washington, DC*, Citeseer.

Sun, H., Y. Chen, J. Lai, Y. Wang, X. Liu (2021) Identifying tourists and locals by k-means clustering method from mobile phone signaling data, *Journal of Transportation Engineering, Part A: Systems*, 147(10), p. 04021070.

Sun, H., H. X. Liu, H. Xiao, R. R. He, B. Ran (2003) Use of local linear regression model for short-term traffic forecasting, *Transportation Research Record*, 1836(1), pp. 143–150.

Szegedy, C., V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna (2016) Rethinking the inception architecture for computer vision, in: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2818–2826.

Szocska, M., P. Pollner, I. Schiszler, T. Joo, T. Palicz, M. McKee, A. Asztalos, L. Bencze, M. Kapronczay, P. Petrecz, et al. (2021) Countrywide population movement monitoring using mobile devices generated (big) data during the covid-19 crisis, *Scientific reports*, 11(1), pp. 1–9.

Tavassoli, A., A. Alsger, M. Hickman, M. Mesbah (2016) How close the models are to the reality? comparison of transit origin-destination estimates with automatic fare collection data, in: *Proc. 38th Australas. Transp. Res. Forum (ATRF)*, pp. 1–15.

Thakuriah, P. (2001) Urban transportation planning: A decision-oriented approach, *Journal of Transportation Engineering*, 127(5), pp. 454–454.

Timmermans, H., T. A. Arentze (2011) Transport models and urban planning practice: Experiences with albatross, *Transport Reviews*, 31(2), pp. 199–207.

Tolouei, R., S. Psarras, R. Prince (2017) Origin-destination trip matrix development: Conventional methods versus mobile phone data, *Transportation research procedia*, 26, pp. 39–52.

Toole, J. L., S. Colak, B. Sturt, L. P. Alexander, A. Evsukoff, M. C. González (2015) The path most traveled: Travel demand estimation using big data resources, *Transportation Research Part C: Emerging Technologies*, 58, pp. 162 – 177, big Data in Transportation and Traffic Engineering.

van der Vooren, A., F. Alkemade (2012) Managing the diffusion of low emission vehicles, *IEEE Transactions on Engineering Management*, 59(4), pp. 728–740.

van Elteren, T. (2018) *A comparative study of human engineered features and learned features in deep convolutional neural networks for image classification*, Ph.D. thesis.

Van Gansbeke, W., S. Vandenhende, S. Georgoulis, M. Proesmans, L. Van Gool (2020) Scan: Learning to classify images without labels, in: *European Conference on Computer Vision*, Springer, pp. 268–285.

Wand, M. P., M. C. Jones (1994) *Kernel smoothing*, Crc Press.

Wang, C., D. Chen, L. Hao, X. Liu, Y. Zeng, J. Chen, G. Zhang (2019) Pulmonary image classification based on inception-v3 transfer learning model, *IEEE Access*, 7, pp. 146533–146541.

Wang, D., Y. Yang, S. Ning (2018) Deepstcl: A deep spatio-temporal convlstm for travel demand prediction, in: *2018 international joint conference on neural networks (IJCNN)*, IEEE, pp. 1–8.

Wang, M.-H., S. D. Schrock, N. V. Broek, T. Mulinazzi (2013) Estimating dynamic origin-destination data and travel demand using cell phone network data, *International Journal of Intelligent Transportation Systems Research*, 11(2), pp. 76–86.

Wang, R., D. B. Work, R. Sowers (2016) Multiple model particle filter for traffic estimation and incident detection, *IEEE Transactions on Intelligent Transportation Systems*, 17(12), pp. 3461–3470.

Wang, X., Y. Shang, G. Li (2024) Dtm-gcn: A traffic flow prediction model based on dynamic graph convolutional network, *Multimedia Tools and Applications*, 83, pp. 12345–12360.

Wang, Y., G. Correia, E. de Romph, B. F. Santos (2020) Road network design in a developing country using mobile phone data: An application to senegal, *IEEE Intelligent Transportation Systems Magazine*, 12(2), pp. 36–49.

Welling, M., T. N. Kipf (2016) Semi-supervised classification with graph convolutional networks, in: *J. International Conference on Learning Representations (ICLR 2017)*.

Winter, K., O. Cats, K. Martens, B. van Arem (2020) Relocating shared automated vehicles under parking constraints: assessing the impact of different strategies for on-street parking, *Transportation*, cited By 1.

Wu, Z., S. Pan, G. Long, J. Jiang, C. Zhang (2019) Graph wavenet for deep spatial-temporal graph modeling, *arXiv preprint arXiv:1906.00121*.

Xie, Y., K. Zhao, Y. Sun, D. Chen (2010) Gaussian processes for short-term traffic volume forecasting, *Transportation Research Record*, 2165(1), pp. 69–78.

Xiong, X., K. Ozbay, L. Jin, C. Feng (2020) Dynamic origin–destination matrix prediction with line graph neural networks and kalman filter, *Transportation Research Record*, 2674(8), pp. 491–503.

Xu, D., Y. Wang, P. Peng, S. Beilun, Z. Deng, H. Guo (2020) Real-time road traffic state prediction based on kernel-knn, *Transportmetrica A: Transport Science*, 16(1), pp. 104–118.

Xu, X., A. Chen, C. Yang (2017) An optimization approach for deriving upper and lower bounds of transportation network vulnerability under simultaneous disruptions of multiple links, *Transportation research procedia*, 23, pp. 645–663.

Xu, Z., Z. Lv, J. Li, H. Sun, Z. Sheng (2022) A novel perspective on travel demand prediction considering natural environmental and socioeconomic factors, *IEEE Intelligent Transportation Systems Magazine*, 15(1), pp. 136–159.

Yabe, T., N. K. Jones, P. S. C. Rao, M. C. Gonzalez, S. V. Ukkusuri (2022) Mobile phone location data for disasters: A review from natural hazards and epidemics, *Computers, Environment and Urban Systems*, 94, p. 101777.

Yair, E., A. Gersho (1990) Maximum a posteriori decision and evaluation of class probabilities by boltzmann perceptron classifiers, *Proceedings of the IEEE*, 78(10), pp. 1620–1628.

Yan, H., X. Ma (2021) Learning dynamic and hierarchical traffic spatiotemporal features with transformer, *arXiv preprint arXiv:2104.05163*.

Yang, H., K. Xie, K. Ozbay, Y. Ma, Z. Wang (2018a) Use of deep learning to predict daily usage of bike sharing systems, *Transportation research record*, 2672(36), pp. 92–102.

Yang, J.-S. (2005) Travel time prediction using the gps test vehicle and kalman filtering techniques, in: *Proceedings of the 2005, American Control Conference, 2005.*, IEEE, pp. 2128–2133.

Yang, Z., M. L. Franz, S. Zhu, J. Mahmoudi, A. Nasri, L. Zhang (2018b) Analysis of washington, dc taxi demand using gps and land-use data, *Journal of Transport Geography*, 66, pp. 35–44.

Yeung, K. Y., D. R. Haynor, W. L. Ruzzo (2001) Validating clustering for gene expression data, *Bioinformatics*, 17(4), pp. 309–318.

Yu, B., H. Yin, Z. Zhu (2017) Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting, *arXiv preprint arXiv:1709.04875*.

Zhai, J., J. Qi, C. Shen (2022) Binary imbalanced data classification based on diversity oversampling by generative models, *Information Sciences*, 585, pp. 313–343.

Zhang, J., Y. Zheng, D. Qi (2017) Deep spatio-temporal residual networks for citywide crowd flows prediction, in: *Proceedings of the AAAI conference on artificial intelligence*, vol. 31.

Zhang, J., Y. Zheng, J. Sun, D. Qi (2019) Flow prediction in spatio-temporal networks based on multitask deep learning, *IEEE Transactions on Knowledge and Data Engineering*, 32(3), pp. 468–478.

Zhang, S., H. Zheng, H. Su, B. Yan, J. Liu, S. Yang (2021) Gacan: Graph attention-convolution-attention networks for traffic forecasting based on multi-granularity time series, *arXiv preprint arXiv:2110.14331*.

Zhang, Y., Y. Liu (2009) Traffic forecasting using least squares support vector machines, *Transportmetrica*, 5(3), pp. 193–213.

Zhang, Y., X. Qin, S. Dong, B. Ran (2010) Daily od matrix estimation using cellular probe data, in: *89th Annual Meeting Transportation Research Board*, vol. 9.

Zhang, Z. (2021) Resnet-based model for autonomous vehicles trajectory prediction, in: *2021 IEEE International Conference on Consumer Electronics and Computer Engineering (ICCECE)*, IEEE, pp. 565–568.

Zhao, L., Y. Song, C. Zhang, Y. Liu, P. Wang, T. Lin, M. Deng, H. Li (2019) T-gcn: A temporal graph convolutional network for traffic prediction, *IEEE transactions on intelligent transportation systems*, 21(9), pp. 3848–3858.

Zhao, T., Z. Huang, W. Tu, B. He, R. Cao, J. Cao, M. Li (2022) Coupling graph deep learning and spatial-temporal influence of built environment for short-term bus travel demand prediction, *Computers, Environment and Urban Systems*, 94, p. 101776.

Zhao, Z., H. N. Koutsopoulos, J. Zhao (2020) Discovering latent activity patterns from transit smart card data: A spatiotemporal topic model, *Transportation Research Part C: Emerging Technologies*, 116, p. 102627.

Zhao, Z., S.-L. Shaw, Y. Xu, F. Lu, J. Chen, L. Yin (2016) Understanding the bias of call detail records in human mobility research, *International Journal of Geographical Information Science*, 30(9), pp. 1738–1762.

Zhou, Y., Q. Yuan, F. Ding, M. Chen, C. Yang, T. Guo (2022) Demand, mobility, and constraints: Exploring travel behaviors and mode choices of older adults using a facility-based framework, *Journal of transport geography*, 102, p. 103368.

Zilske, M., K. Nagel (2014) Studying the accuracy of demand generation from mobile phone trajectories with synthetic data, *Procedia Computer Science*, 32, pp. 802 – 807, the 5th International Conference on Ambient Systems, Networks and Technologies (ANT-2014), the 4th International Conference on Sustainable Energy Information Technology (SEIT-2014).

Zilske, M., K. Nagel (2015) A simulation-based approach for constructing all-day travel chains from mobile phone data, *Procedia Computer Science*, 52, pp. 468–475, available Open Access publishedVersion at https://depositonce.tu-berlin.de//handle/11303/7626.

Glossary

The following abbreviations are used in this thesis:

AP	Afternoon Peak
ARIMA	Autoregressive Integrated Moving Average
CBD	Central Bureau of Statistics
CDR	Call Detailed Record
CNN	Convolutional Neural Network
COV	Coefficient of Variation
DCNN	Deep Convolutional Neural Networks
FR	Fortunate Range
GCN	Graph Convolutional Neural Network
GRU	Gated Recurrent Unit
GSM	Global System for Mobile Communication
GSSI	Geographical Window-based Structural Similarity Index
GTWR	Geographically and Temporally Weighted Regression
HCA	Hierarchical Clustering Analysis
KA	Kernel-based Approach
KNN	K-nearest Neighbor
KDE	Kernel Density Estimation
LSTM	Long Short-term Memory
LS-SVM	Least Squares Support Vector Machine
MAE	Mean Absolute Error
MAP	Maximum a-posteriori
MAPE	Mean Absolute Percentage Error
MP	Morning Peak
MSE	Mean Squared Error
MZ	Mobility Zone
OD	Origin-destination
OSM	Open Street Map
OVR	One-vs.-Rest
PLU	Periodic Location Update
PI	Polling Interval
POI	Point of Interest
RMSE	Root Mean Squared Error
SI	Silhouette Index
SMOTE	Synthetic Minority Oversampling Technique
TAZ	Traffic Analysis Zone
XGBoost	eXtreme Gradient Boosting

Summary

Urban mobility and transportation systems are inherently complex and characterized by dynamic patterns that evolve over time. Effective management, planning, and optimization of these systems necessitate a deep understanding of spatial and temporal travel demand patterns. This thesis contributes to the field of transportation science by developing and applying state-of-the-art data-driven methods to explore the spatial and temporal patterns of travel demand within large-scale metropolitan networks. Specifically, it focuses on addressing critical challenges associated with estimating and predicting travel demand by leveraging the potential of GSM mobile phone data.

At the core of this research lies the understanding that the quality and temporal resolution of input data are pivotal in steering the efficacy of transportation planning and policy-making. Central to this investigation is the formulation of an innovative, data-driven framework tailored for the analysis of GSM data, intending to refine the accuracy in the construction of origin-destination (OD) matrices. This novel approach meticulously examines the impact of polling intervals (PI) on the temporal fidelity of GSM data, addressing the inherent challenges in discerning stay durations and the subsequent effects on travel demand estimation. Through the strategic manipulation of data collection intervals, this study not only proposes an advanced methodology for interpreting sparse GSM data but also identifies an ”optimal PI” that significantly enhances the structural accuracy of OD matrices. The integration of Kernel-based learning from minimal travel diary samples underscores the feasibility of reconstructing reliable travel patterns, marking a significant leap forward in the utilization of mobile phone data for transportation analysis. By delineating a pathway toward optimizing GSM data handling, this research contributes substantially to the foundation of data-driven transportation planning, ensuring the derivation of more accurate and robust travel demand models.

Building upon the foundational insights derived from analyzing GSM data for estimating travel demand, we advance our understanding by distinguishing the spatial-temporal travel demand patterns. It transitions from the foundational analysis of raw GSM data to a focused exploration of how spatial-temporal heterogeneity influences trip production across different traffic analysis zones (TAZs). It emphasizes the importance of recognizing spatial-temporal variations in trip production (and attraction), which are crucial for a reliable prediction and estimation of travel demand. Through a detailed exploration of the variability in trip production at various times of the day and days of the week, significant patterns emerge among TAZs in the Netherlands, providing critical insights for transportation planning and policy-making. By linking temporal patterns of travel production with spatial urbanization characteristics, this study offers valuable perspectives on mobility, crucial for transportation analysis and the formulation of effective policies. It represents a significant step towards refining our approach to modeling (and predicting) travel demand, ensuring transportation systems are more responsive to the needs of urban and rural areas.

Next, we delve into one of the complexities of forecasting travel demand by analyzing predictive accuracies at these different scales. This exploration is crucial for understanding the spatial-temporal dependence and heterogeneity inherent in trip production, aspects often overlooked in conventional forecasting approaches. The chapter uses a method that combines graph convolutional neural networks (GCN) with long short-term memory networks (LSTM), an ap-

proach aimed at addressing the challenge of spatial heterogeneity by creating a nationwide graph that reflects the adjacency of mobility zones (MZ).

The key focus of the chapter is on the multi-scale aspect of travel demand prediction, where the research underscores the importance of understanding how spatial adjacency and the scale of analysis influence the accuracy of trip production forecasts. This approach is significant as it moves beyond the traditional one-size-fits-all models, offering a more nuanced understanding of how travel demand prediction can vary across different spatial scales—from provinces to municipalities to 4-digit postal code zones. Therefore, it demonstrates that the choice of spatial scale for analysis is not merely a technical decision but a critical factor that can significantly influence the effectiveness of travel demand models. This insight is crucial for transportation planners and policymakers who aim to develop targeted strategies for different urban and regional contexts.

The exploration of travel demand patterns in the context of socio-spatial characteristics refines our understanding of urban mobility; Hence, we delve into the sensitivity of travel demand to the interaction of demographic and land-use attributes and we explore the intricate relationships that govern trip production predictions across varying spatial scales. The research underscores the importance of a data-driven approach in analyzing the layers of complexity inherent in predicting trip production, providing a methodological framework that leverages an extensive range of socio-spatial data. This approach is pivotal in unraveling the impact of various demographic factors, such as population density, average age, income levels, and employment status, alongside collective land-use characteristics on the predictability of travel demand patterns. The analysis embarks on a journey through different spatial scales, examining how the specific socio-spatial makeup of an area contributes to or detracts from the predictability of trip production. Through this lens, the research brings to light the variegated effects of demographic and land-use factors across spatial dimensions, offering insights into the differential impact these variables have on travel demand forecasting.

The last segment of the study employs sophisticated analytical techniques to investigate the prediction errors associated with trip production forecasts. By identifying and evaluating the dominant patterns of these prediction residuals, the research delves deeper into the socio-spatial foundations that contribute to discrepancies in travel demand forecasting. This meticulous analysis paves the way for a more refined understanding of trip production heterogeneity and the socio-spatial determinants that shape it, thereby enhancing the accuracy of demand prediction models.

To conclude, this thesis explores the intricate landscape of travel demand prediction, addressing how data quality, spatio-temporal context, spatial scale, and socio-spatial characteristics can enhance forecasting accuracy. Key insights reveal the importance of temporal data resolution, the predictive power of land-use characteristics across spatial scales, and the integration of spatial adjacency in prediction models. Despite its contributions to transport planning and demand estimation, the research acknowledges limitations related to data representativeness and methodological approaches, suggesting avenues for future studies to refine predictive models further. By marrying data-driven methodologies with socio-spatial interactions, this work lays a foundational blueprint for advancing urban mobility analysis and policy-making, aiming for more efficient, inclusive, and responsive transportation planning.

Samenvatting

Stedelijke mobiliteit en transportsystemen zijn van nature complex en worden gekenmerkt door dynamische patronen die in de loop van de tijd evolueren. Effectief beheer, planning en optimalisatie van deze systemen vereisen een diepgaand begrip van de ruimtelijke en temporele patronen van vervoersvraag. Deze thesis draagt bij aan het vakgebied van transportwetenschap door het ontwikkelen en toepassen van geavanceerde datagedreven methoden om de ruimtelijke en temporele patronen van vervoersvraag binnen grootschalige metropolitane netwerken te verkennen. Specifiek richt het zich op het aanpakken van kritieke uitdagingen die geassocieerd zijn met het schatten en voorspellen van vervoersvraag door het benutten van het potentieel van GSM mobiele telefoongegevens.

In de kern van dit onderzoek ligt het begrip dat de kwaliteit en temporele resolutie van invoergegevens cruciaal zijn in het sturen van de effectiviteit van transportplanning en beleidsvorming. Centraal in dit onderzoek staat de formulering van een innovatief, datagedreven kader op maat voor de analyse van GSM-gegevens, met als doel de nauwkeurigheid in de constructie van herkomst-bestemmingsmatrices te verfijnen. Deze nieuwe benadering onderzoekt nauwgezet de impact van pollingintervallen (PI) op de temporele betrouwbaarheid van GSM-gegevens, waarbij de inherente uitdagingen in het onderscheiden van verblijfsduren en de daaropvolgende effecten op de schatting van vervoersvraag worden aangepakt. Door de strategische manipulatie van gegevensverzamelingsintervallen stelt deze studie niet alleen een geavanceerde methodologie voor voor het interpreteren van schaarse GSM-gegevens, maar identificeert ook een "optimaal P" dat de structurele nauwkeurigheid van OD-matrices aanzienlijk verbetert. De integratie van op Kernel-gebaseerd leren uit minimale reisdagboeksamples onderstreept de haalbaarheid van het reconstrueren van betrouwbare reispatronen, wat een aanzienlijke stap voorwaarts markeert in het gebruik van mobiele telefoongegevens voor transportanalyse. Door een pad uit te stippelen naar het optimaliseren van de GSM-gegevensverwerking, levert dit onderzoek een aanzienlijke bijdrage aan de basis van datagedreven transportplanning, waarbij zorg wordt gedragen voor de afleiding van nauwkeurigere en robuustere vervoersvraagmodellen.

Voortbouwend op de fundamentele inzichten verkregen uit de analyse van GSM-gegevens voor het schatten van vervoersvraag, verdiepen we ons begrip door de ruimtelijk-temporele vervoersvraagpatronen te onderscheiden. Het gaat over van de fundamentele analyse van ruwe GSM-gegevens naar een gerichte verkenning van hoe ruimtelijk-temporele heterogeniteit de tripproductie beïnvloedt over verschillende verkeersanalysezones (TAZ's). Het benadrukt het belang van het herkennen van ruimtelijk-temporele variaties in tripproductie (en attractie), die cruciaal zijn voor een betrouwbare voorspelling en schatting van vervoersvraag. Door een gedetailleerde verkenning van de variabiliteit in tripproductie op verschillende tijden van de dag en dagen van de week, komen significante patronen naar voren onder TAZ's in Nederland, wat cruciale inzichten biedt voor transportplanning en beleidsvorming. Door temporele patronen van reisproductie te koppelen aan ruimtelijke verstedelijkingskenmerken, biedt deze studie waardevolle perspectieven op mobiliteit, cruciaal voor transportanalyse en het formuleren van effectieve beleidsmaatregelen. Het vertegenwoordigt een significante stap naar het verfijnen van onze aanpak voor het modelleren (en voorspellen) van vervoersvraag, zodat transportsystemen responsiever zijn voor de behoeften van stedelijke en landelijke gebieden.

Vervolgens duiken we in een van de complexiteiten van het voorspellen van vervoersvraag

door de voorspellingsnauwkeurigheden op deze verschillende schalen te analyseren. Deze verkenning is cruciaal voor het begrijpen van de ruimtelijk-temporele afhankelijkheid en heterogeniteit die inherent zijn aan tripproductie, aspecten die vaak over het hoofd worden gezien in conventionele voorspellingsbenaderingen. Het hoofdstuk gebruikt een methode die grafische convolutieneurale netwerken (GCN) combineert met lange kortetermijngeheugennetwerken (LSTM), een aanpak gericht op het aanpakken van de uitdaging van ruimtelijke heterogeniteit door het creëren van een landelijk grafiek die de aangrenzendheid van verkeersanalysezones (TAZ's) weerspiegelt.

De belangrijkste focus van het hoofdstuk ligt op het multi-schaal aspect van vervoersvraagvoorspelling, waarbij het onderzoek het belang onderstreept van het begrijpen hoe ruimtelijke aangrenzendheid en de schaal van analyse de nauwkeurigheid van tripproductievoorspellingen beïnvloeden. Deze aanpak is significant omdat het verder gaat dan de traditionele one-size-fits-all modellen, en biedt een meer genuanceerd begrip van hoe vervoersvraagvoorspelling kan variëren over verschillende ruimtelijke schalen - van provincies tot gemeenten tot 4-cijferige postcodezones. Daarom toont het aan dat de keuze van ruimtelijke schaal voor analyse niet slechts een technische beslissing is, maar een kritieke factor die de effectiviteit van vervoersvraagmodellen aanzienlijk kan beïnvloeden. Dit inzicht is cruciaal voor transportplanners en beleidsmakers die gerichte strategieën willen ontwikkelen voor verschillende stedelijke en regionale contexten.

De verkenning van vervoersvraagpatronen in de context van socio-ruimtelijke kenmerken verfijnt ons begrip van stedelijke mobiliteit; daardoor duiken we in de gevoeligheid van vervoersvraag voor de interactie van demografische en grondgebruikskenmerken en verkennen we de complexe relaties die tripproductievoorspellingen regeren over verschillende ruimtelijke schalen. Het onderzoek benadrukt het belang van een datagedreven aanpak bij het analyseren van de lagen van complexiteit die inherent zijn aan het voorspellen van tripproductie, en biedt een methodologisch kader dat een uitgebreid scala aan socio-ruimtelijke gegevens benut. Deze aanpak is cruciaal bij het ontrafelen van de impact van verschillende demografische factoren, zoals bevolkingsdichtheid, gemiddelde leeftijd, inkomensniveaus en werkstatus, naast collectieve grondgebruikskenmerken op de voorspelbaarheid van vervoersvraagpatronen. De analyse begint aan een reis door verschillende ruimtelijke schalen, onderzoekend hoe de specifieke socio-ruimtelijke samenstelling van een gebied bijdraagt aan of afbreuk doet aan de voorspelbaarheid van tripproductie. Door deze lens brengt het onderzoek de gevarieerde effecten van demografische en grondgebruiksfactoren over ruimtelijke dimensies aan het licht, en biedt inzichten in de differentiële impact van deze variabelen op vervoersvraagvoorspelling.

Het laatste segment van de studie maakt gebruik van geavanceerde analytische technieken om de voorspellingsfouten te onderzoeken die geassocieerd zijn met tripproductievoorspellingen. Door het identificeren en evalueren van de dominante patronen van deze voorspellingsresiduen, duikt het onderzoek dieper in de socio-ruimtelijke fundamenten die bijdragen aan discrepanties in vervoersvraagvoorspelling. Deze nauwgezette analyse baant de weg voor een verfijnder begrip van tripproductieheterogeniteit en de socio-ruimtelijke determinanten die deze vormgeven, waardoor de nauwkeurigheid van vraagvoorspellingsmodellen wordt verbeterd.

Ter afsluiting verkent deze thesis het complexe landschap van vervoersvraagvoorspelling, waarbij wordt ingegaan op hoe datakwaliteit, ruimtelijk-temporele context, ruimtelijke schaal en socio-ruimtelijke kenmerken de voorspellingsnauwkeurigheid kunnen verbeteren. Belangrijke inzichten onthullen het belang van temporele dataresolutie, de voorspellende kracht van grondgebruikskenmerken over ruimtelijke schalen heen, en de integratie van ruimtelijke aangrenzendheid in voorspellingsmodellen. Ondanks de bijdragen aan transportplanning en vraagschatting, erkent het onderzoek beperkingen met betrekking tot gegevensrepresentativiteit en methodologische benaderingen, en suggereert het wegen voor toekomstige studies om voorspellingsmodellen verder te verfijnen. Door datagedreven methodologieën te combineren met socio-ruimtelijke interacties, legt dit werk een fundamentele blauwdruk voor het bevorderen van stedelijke mobiliteitsanalyse en beleidsvorming, met als doel efficiëntere, inclusievere en responsievere transportplanning.

About the author

Zahra Eftekhar was born on November 30, 1992, in Lar, Fars, Iran, a place celebrated for its rich culture and historical monuments. Her early years were marked by a move to the province capital, Shiraz, where she embarked on her educational journey. It was in Shiraz that Zahra pursued her bachelor's degree from 2012 to 2016, majoring in Civil Engineering at Shiraz University. Her academic prowess was evident early on when she ranked among the top 100 in the country for civil engineering in 2016. This achievement paved the way for her to delve into transport engineering and planning at the prestigious Sharif University of Technology (SUT) for her master's program, which she completed in December 2018. Her master's thesis, "Estimation of OD matrix using mobile phone data," laid the groundwork for her future research endeavors.

In March 2019, Zahra embarked on her PhD journey at Dittlab (Data Analytics and Traffic Simulation Laboratory), within the Department of Transport & Planning at Delft University of Technology. Her research, an integral part of the MiRRORs project and funded by the National Data Warehouse (NDW) of the Netherlands, focuses on unraveling spatial-temporal patterns of travel demand. Applying her expertise in travel demand, clustering analysis, and machine learning, Zahra has developed innovative methodologies and frameworks. These contributions are instrumental in identifying and comparing travel patterns, as well as elucidating their association with socio-spatial features.

Beyond her academic and research pursuits, Zahra is deeply committed to leveraging her analytical and data science skills across various domains. Her passion for data transcends her research, aiming to make a tangible impact in diverse fields. Zahra's work embodies a unique blend of technical excellence and a fervent desire to contribute meaningful insights beyond the confines of her discipline.

Publications

Journal papers

- **Eftekhar, Z.**, Pel, A., & van Lint, H. (2023). Effects of Periodic Location Update Polling Interval on the Reconstructed Origin–Destination Matrix: A Dutch Case Study Using a Data-Driven Method. *Transportation Research Record*, 2677(9), 292–313. https://doi-org.tudelft.idm.oclc.org/10.1177/03611981231158638 (published)

- **Eftekhar, Z.**, Pel, A., & van Lint, H. . (2023). A Cluster Analysis of Temporal Patterns of Travel Production in the Netherlands: Dominant within-day and day-to-day patterns and their association with Urbanization Levels. *European Journal of Transport and Infrastructure Research*, 23(3), 1–29.
https://doi.org/10.18757/ejtir.2023.23.3.6499 (published)

- **Eftekhar, Z.**, Behrouzi, S., Krishnakumari, P., Pel, A., & van Lint, H. . (2025). The Role of Spatial Features and Adjacency in Data-driven Short-term Prediction of Trip Production:An Exploratory Study in the Netherlands, *IEEE Transactions on Intelligent Transportation Systems* (third round of review)

Peer-reviewed Conference Contributions

- **Eftekhar, Z.**, Pel, A., & van Lint, H. (2023). Characterizing the Temporal Patterns of Travel Production: A Bird's Eye View of the Urbanization Levels in The Netherlands. *hEART 2022: 10th Symposium of the European Association for Research in Transportation*, Leuven, Belgium.

- **Eftekhar, Z.**, Behrouzi, S., Krishnakumari, P., Pel, A., & van Lint, H. (2023). The Effects of Considering the Spatial Adjacency of Mobility Zones on Predicting the Trip Production: an Exploratory Study in the Netherlands. *102th Annual Meeting of the Transporation Research Board*, Washington D.C., USA.

- **Eftekhar, Z.**, Pel, A., & van Lint, H.. A Cluster Analysis of Temporal Patterns of Travel Production in the Netherlands: Dominant within-day and day-to-day patterns and their association with Urbanization Levels. *24th TRAIL Congress, TRAIL Research School*, Utrecht, The Netherlands.

- **Eftekhar, Z.**, Pel, A., & van Lint, H. (2021). Kernel-based Approach to Reconstruct Travel Diaries from GSM Records. *ISTDM 2021: International Symposium on Transportation Data & Modelling.*

TRAIL Thesis Series

The following list contains the most recent dissertations in the TRAIL Thesis Series, For a complete overview of more than 400 titles, see the TRAIL website: www.rsTRAIL.nl.

The TRAIL Thesis Series is a series of the Netherlands TRAIL Research School on transport, infrastructure and logistics.

Eftekhar, Z., *Exploring the Spatial and Temporal Patterns in Travel Demand: A data-driven approach*, T2025/9, June 2025, TRAIL Thesis Series, the Netherlands

Reddy, N., *Human Driving Behavior when Interacting with Automated Vehicles and the Implications on Traffic Efficiency*, T2025/8, May 2025, TRAIL Thesis Series, the Netherlands

Durand, A., *Lost in Digitalisation? Navigating public transport in the digital era*, T2025/7, May 2025, TRAIL Thesis Series, the Netherlands

Dong, Y., *Safe, Efficient, and Socially Compliant Automated Driving in Mixed Traffic: Sensing, Anomaly Detection, Planning and Control*, T2025/6, May 2025, TRAIL Thesis Series, the Netherlands

Droffelaar, I.S. van, *Simulation-optimization for Fugitive Interception*, T2025/5, May 2025, TRAIL Thesis Series, the Netherlands

Fan, Q., *Fleet Management Optimisation for Ride-hailing Services: from mixed traffic to fully automated environments*, T2025/4, April 2025, TRAIL Thesis Series, the Netherlands

Hagen, L. van der, *Machine Learning for Time Slot Management in Grocery Delivery*, T2025/3, March 2025, TRAIL Thesis Series, the Netherlands

Schilt, I.M. van, *Reconstructing Illicit Supply Chains with Sparse Data: a simulation approach*, T2025/2, January 2025, TRAIL Thesis Series, the Netherlands

Ruijter, A.J.F. de, *Two-Sided Dynamics in Ridesourcing Markets*, T2025/1, January 2025, TRAIL Thesis Series, the Netherlands

Fang, P., *Development of an Effective Modelling Method for the Local Mechanical Analysis of Submarine Power Cables*, T2024/17, December 2024, TRAIL Thesis Series, the Netherlands

Zattoni Scroccaro, P., *Inverse Optimization Theory and Applications to Routing Problems*, T2024/16, October 2024, TRAIL Thesis Series, the Netherlands

Kapousizis, G., *Smart Connected Bicycles: User acceptance and experience, willingness to pay and road safety implications*, T2024/15, November 2024, TRAIL Thesis Series, the Netherlands

Lyu, X., *Collaboration for Resilient and Decarbonized Maritime and Port Operations*, T2024/14, November 2024, TRAIL Thesis Series, the Netherlands

Nicolet, A., *Choice-Driven Methods for Decision-Making in Intermodal Transport: Behavioral heterogeneity and supply-demand interactions*, T2024/13, November 2024, TRAIL Thesis Se-

ries, the Netherlands

Kougiatsos, N., *Safe and Resilient Control for Marine Power and Propulsion Plants*, T2024/12, November 2024, TRAIL Thesis Series, the Netherlands

Uijtdewilligen, T., *Road Safey of Cyclists in Dutch Cities*, T2024/11, November 2024, TRAIL Thesis Series, the Netherlands

Liu, X., *Distributed and Learning-based Model Predictive Control for Urban Rail Transit Networks*, T2024/10, October 2024, TRAIL Thesis Series, the Netherlands

Clercq, G. K. de, *On the Mobility Effects of Future Transport Modes*, T2024/9, October 2024, TRAIL Thesis Series, the Netherlands

Dreischerf, A.J., *From Caveats to Catalyst: Accelerating urban freight transport sustainability through public initiatives*, T2024/8, September 2024, TRAIL Thesis Series, the Netherlands

Zohoori, B., Zohoori, B., Model-based Risk Analysis of Supply Chains for Supporting Resilience, T2024/7, October 2024, TRAIL Thesis Series, the Netherlands, T2024/7, October 2024, TRAIL Thesis Series, the Netherlands

Poelman, M.C., *Predictive Traffic Signal Control under Uncertainty: Analyzing and Reducing the Impact of Prediction Errors*, T2024/6, October 2024, TRAIL Thesis Series, the Netherlands

Berge, S.H., *Cycling in the age of automation :Enhancing cyclist interaction with automated vehicles through human-machine interfaces*, T2024/5, September 2024, TRAIL Thesis Series, the Netherlands

Wu, K., *Decision-Making and Coordination in Green Supply Chains with Asymmetric Information*, T2024/4, July 2024, TRAIL Thesis Series, the Netherlands

Wijnen, W., *Road Safety and Welfare*, T2024/3, May 2024, TRAIL Thesis Series, the Netherlands

Caiati, V., *Understanding and Modelling Individual Preferences for Mobility as a Service*, T2024/2, March 2024, TRAIL Thesis Series, the Netherlands

Vos, J., *Drivers' Behaviour on Freeway Curve Approach*, T2024/1, February 2024, TRAIL Thesis Series, the Netherlands

Geržinič, N., *The Impact of Public Transport Disruptors on Travel Behaviour*, T2023/20, December 2023, TRAIL Thesis Series, the Netherlands

Dubey, S., *A Flexible Behavioral Framework to Model Mobilityon-Demand Service Choice Preference*, T2023/19, November 2023, TRAIL Thesis Series, the Netherlands

TRAIL

Summary

Urban mobility systems are intrinsically complex, requiring robust data-driven insights for effective planning. This thesis uses GSM mobile phone data to examine how data quality, spatio-temporal resolution, and socio-spatial factors influence travel demand in metropolitan networks. By linking land-use and demographic attributes to trip production patterns, it offers valuable guidance for tailoring transportation policies to local needs, ultimately fostering more efficient and responsive transport systems.

About the Author

Zahra Eftekhar conducted her PhD research in the Department of Transport & Planning at Delft University of Technology. She holds an MSc degree in Civil Engineering with a specialization in Transportation Planning and Engineering. Her research interests include the application of machine learning to understand travel patterns.

TRAILResearch School ISBN 978-90-5584-364-0

Erasmus Universiteit Rotterdam
Radboud University
rijksuniversiteit groningen
TU Delft
UNIVERSITY OF TWENTE.
TU/e Eindhoven University of Technology