Year-on-year analysis of multi-modal digital travel diaries
Temporal, spatial and modal traveler profiles

Sipetas, Charalampos; Geržinič, Nejc; Huang, Zhiren; Cats, Oded; Mladenović, Miloš N.

DOI
10.1016/j.tra.2025.104734

Publication date
2026

Document Version
Final published version

Published in
Transportation Research Part A: Policy and Practice

Transportation Research Part A 203 (2026) 104734

Contents lists available at ScienceDirect

Transportation Research Part A

journal homepage: www.elsevier.com/locate/tra

Year-on-year analysis of multi-modal digital travel diaries: Temporal, spatial and modal traveler profiles

Charalampos Sipetas [a,b,*], Nejc Geržinič [c], Zhiren Huang [d], Oded Cats [c], Miloš N. Mladenović [a]

[a] Department of Built Environment, Aalto University, Espoo, Finland
[b] Division of Engineering, New York University Abu Dhabi, Abu Dhabi, United Arab Emirates
[c] Department of Transport and Planning, Delft University of Technology, Delft, the Netherlands
[d] Department of Computer Science, Aalto University, Espoo, Finland

A R T I C L E I N F O

Keywords:
Traveler segmentation
Multi-modality
Demand profiles
Smartphone data
Traveler trajectories

A B S T R A C T

Understanding multi-modal urban mobility patterns is essential for effective planning and policy-making. Traditional data sources, such as infrequent surveys or smart card records, often lack the temporal, spatial, and modal comprehensiveness required to fully capture the complexity of multi-modal travel behavior. Emerging mobility data sources are instrumental in capturing these patterns and in enabling additional insights. This study leverages a digitally collected trajectory-level dataset (i.e., TravelSense) obtained from a smartphone application operated by the public transport authority of Helsinki, Finland. Unlike conventional public transport data, TravelSense provides insights into modal choices alongside temporal and spatial travel characteristics. In order to analyze mobility patterns and explore the capabilities of this novel dateset, a Latent Profile Analysis is employed to classify travelers based on these attributes over a week-long period, with profiles compared across three consecutive years (2022, 2023, and 2024). Findings reveal that while spatial travel patterns remain relatively stable, temporal and modal patterns exhibit greater variability. A distinct shift is observed between 2022 and subsequent years, likely reflecting post-pandemic behavioral changes. Key traveler groups identified include exclusive active mode users (13 % annually) and non-private car users, whose share declined from 38 % in 2022 to approximately 20 % in 2023 and 2024. Study findings offer valuable input for shaping evidence-based mobility policies, particularly those aiming to support sustainable travel behavior and adapt to evolving urban mobility needs through enhanced multi-modality. TravelSense enables detailed analysis of temporal, spatial, and modal travel patterns, underscoring the value of novel data for multi-modal transport research.

1. Introduction

Studying behavioral patterns of travelers is essential for shaping policies that support sustainable mobility (Molin et al., 2016). Most recent studies up to date emphasize on spatio-temporal aspects of human mobility (Cats, 2024). These aspects describe when and where people travel, often revealed through patterns such as morning peak travel or directional flows to specific destinations. However, many existing approaches analyze time and space separately or rely on aggregated data, which can overlook individual-level variation and evolving travel behavior. Rapid technological advancements nowadays enable the automated collection of detailed

* Corresponding author.
E-mail address: charalampos.sipetas@aalto.fi, charalampos.sipetas@nyu.edu (C. Sipetas).

https://doi.org/10.1016/j.tra.2025.104734

Available online 4 November 2025

traveler trajectories, providing insights into multi-modal behaviour alongside spatio-temporal dynamics (Huang et al., 2022). Such information is beneficial for shaping a range of critical decisions, for example, related with resource allocation, infrastructure management, service design and ticket pricing. Alongside the technological developments, the need for capturing changes in mobility patterns over time through longitudinal approaches has been highlighted in the literature (Briand et al., 2017). The longitudinal focus on understanding human mobility is even more important in complex multi-modal networks, that can enable different modal combinations and a range of spatio-temporal patterns.

The importance of understanding travel mode choices is increasingly recognized in recent literature (Varga et al., 2025), with growing emphasis on policies that promote sustainable modal shifts through multi-modality (An et al., 2022). Multi-modality, the use of multiple transport modes within or across trips, is vital for sustainable mobility, yet often perceived as burdensome compared to private car use (Fu et al., 2024). This highlights the need to deepen our understanding of mode choice behavior and multi-modal travel patterns (Huang et al., 2024). However, multi-modality in travel behaviour has traditionally been overlooked in the literature, primarily due to challenges in data acquisition and the complexities of developing advanced models (Clifton and Muhs, 2012).

A conventional method for gathering information on traveler movements within a transport network is through surveys (Stopher and Greaves, 2007). However, information is collected infrequently due to high costs, leading to concerns about outdated information. An accurate and comprehensive understanding of individuals' travel modality styles requires observations over multiple days to capture both the fluctuations and the consistency in their travel behavior (Kuhnimhof et al., 2006; Nobis, 2007; Buehler and Hamre, 2015; Iogansen et al., 2025). Automated data collection sources, such as smart cards, allow up-to-date information about human mobility and have proven invaluable for revealing travel patterns mostly in terms of spatio-temporal aspects (Cats, 2024). Studying multi-modality through automatically collected sources, in most cases requires advanced inference models that integrate multiple data sources to reveal multi-modal journey patterns (Seaborn et al., 2009; Gordon et al., 2013; Munizaga and Palma, 2012; Barry et al., 2009). Recent studies have combined conventional questionnaires with GPS-based travel diaries to approach multi-modality (Iogansen et al., 2025). As the investigation of multi-modality heavily depends on data and technological advancements constantly introduce novel data sources, it is critical to explore their potentials into understanding human mobility from a temporal, spatial and modal standpoint.

While existing data-driven studies have contributed valuable insights into travel behavior, many rely on infrequently collected survey data or focus narrowly on temporal and/or spatial dimensions using mostly smart card records (Cats, 2024). These approaches often lack the capacity to capture the full complexity of temporal, spatial and modal aspects of increasingly dynamic and integrated urban transport systems. In this study, we utilize the novel TravelSense dataset in order to analyse multi-modal mobility within the Helsinki region, to better understand traveler behaviour and assist critical decision-making for strategy and policy development. TravelSense data emerged in the recent years as a promising source of information for detailed multi-modal traveler trajectories in the Helsinki region transport network. Its first pilot was launched during 2020. As a data source, TravelSense depends on the operation of a ticketing application for smartphones, which collects the respective information after user consent ensuring privacy and anonymity. Information included in the generated dataset refers to temporal, spatial and modal aspects of mobility, resembling digital travel diaries. Among other locations, it is noted that a similar smartphone-based data source that collects data from multi-modal journeys while ensuring user privacy has been piloted in Barcelona, Spain (CIT UPC, 2023).

Creating traveler profiles based on their activities is a common approach to a methodical investigation of human mobility patterns. In relation, Latent Profile Analysis (LPA) is a family of statistical models that can be used for identifying latent subpopulations within a population based on a certain set of variables (Spurk et al., 2020). Such a clustering technique is used in other mobility-related studies for revealing traveling patterns (Ton et al., 2020; Gutiérrez et al., 2020; McBride et al., 2018). In this study, LPA is utilized for generating traveler profiles across three variable categories, namely temporal, spatial, and modal. In addition, a cross-categorical analysis is performed to offer combinatorial insights of temporal-spatial-modal activities. Finally, a longitudinal analysis is performed to identify the changes in the profiles over the span of three years.

With this in mind, the objectives of this study are threefold:

1. We analyse mobility patterns within urban transport networks, highlighting the importance of mode-related travel patterns in addition to spatio-temporal attributes that are commonly met when automated data sources are considered. We also investigate the changes in these profiles over time, through analysing information from three consecutive years. The insights from such a study are expected to be relevant also for networks beyond Helsinki region, which serves as the case study here.
2. We explore and demonstrate the capabilities of a novel data source that refers to digitally collected travel diaries in studies that profile travelers based on temporal, spatial, and modal attributes of their trips. Such dynamic sources of information are expected to become more common among transport authorities as technology evolves and mobility standards consolidate, potentially leading to part or full replacement of time and labor consuming travel surveys.
3. We reveal and discuss practice and policy-related impacts and implications resulting from the TravelSense technology within the field of transport demand profiling. Our expectation is that as such digitally collected travel diaries become more common, transport stakeholders should be prepared for their associated challenges and opportunities to addressing critical mobility issues and enhancing transport network performance.

The rest of this article is structured as follows. Section 2 presents the literature review in the field of data sources and demand profiling. Section 3 describes the TravelSense technology. Section 4 explains the methodology implemented in this study to generate traveler profiles and analyse year-to-year mobility patterns. Section 5 presents the data and variables for the Helsinki case study. Section 6 reports the results of implementing the proposed methodology for demand profile analysis. Section 7 discusses the findings of this study with emphasis on policy and practice implications. Finally, Section 8 provides the conclusions of this study.

2. Literature review

2.1. Data sources

Understanding multi-modal travel behaviours is essential for effective transportation planning and transport demand profiling. To gain this understanding, modern ICT infrastructure provides diverse mobility data sources (Zheng et al., 2015; Welch and Widita, 2019; Cats, 2024), including smart card transactions, sensor data (such as Bluetooth and WiFi signals), and mobile phone data such as call detail records (CDR) and signaling data. While these sources provide valuable insights, each of them also possess inherent shortcomings. For instance, some focus on a single transport mode, while others cover multiple modes, but lack detailed and mode-specific information.

Automated fare control (AFC) systems are widely deployed in public transport networks, allowing passengers to access services using smart cards or other forms of contactless payment. Studies have leveraged smart card data for various applications, such as estimating route choice patterns (Zhao et al., 2016), classifying passenger behaviour over time (Briand et al., 2017), and validating multi-modal transit models (Dixit et al., 2024). However, smart card data are limited to recorded interactions within the public transport system, offering no information on pre-boarding or post-alighting movements. Additionally, since data generation occurs only during tap events, uncertainties arise regarding destination stops–especially in systems without mandatory tap-out requirements. To mitigate these issues, inference techniques have been proposed (Trépanier et al., 2007; Zhao et al., 2007; Seaborn et al., 2009; Zheng et al., 2018), typically based on assumptions regarding trip continuity and return patterns. Nonetheless, in proof-of-payment systems, particularly for season ticket holders who may not tap in or out, destination inference remains a significant challenge.

To address the limitations associated with smart card data, continuous sensing through Bluetooth beacons and WiFi access points has been explored. Kostakos et al. (2013) deployed Bluetooth transceivers on buses but found that only about 12 % of passengers carried Bluetooth enabled devices. Similarly, Gonzales et al. (2018) installed wireless device detectors at metro stations in Boston to track passenger waiting times, but detection latencies hindered the accuracy of short-duration measurements. Tu et al. (2019) have used WiFi-enabled buses to infer boarding stops and partial destinations, though these systems are typically deployed only at select stops or vehicles. Moreover, the growing prevalence of MAC address randomization (Nitti et al., 2020; Jee et al., 2023) by mobile operating systems has reduced the effectiveness of WiFi-based tracking, complicating continuous journey monitoring and raising privacy concerns.

Mobile phone data offer broader insights into overall mobility patterns compared to transport system-specific smart card or sensor data, making them widely used for origin-destination (O-D) estimation (Iqbal et al., 2014; Jiang et al., 2017). While signaling data provide superior temporal resolution for trip identification over CDRs (Huang et al., 2018; Wang and Chen, 2018; Cheng et al., 2024), a major limitation across mobile phone data is the difficulty in reliably inferring transport modes. This challenge stems primarily from the coarse spatial positioning and relatively infrequent updates of cell tower events. To address this, researchers have employed methodologies such as matrix factorization (Graells-Garrido et al., 2018) and Bayesian inference (Bachir et al., 2019) to differentiate between transport modes. Alternative approaches have explored GPS trajectory mining from various mobile applications, including travel platforms (Zheng et al., 2008), disaster alert systems (Pang et al., 2018), and navigation services (Wang et al., 2021). Feature engineering and machine learning techniques (Zheng et al., 2008; Huang et al., 2020; Wang et al., 2021) have significantly improved mode classification accuracy. However, GPS signals are prone to fail in complex environments such as underground metro systems (Zheng, 2011). As a result, GPS/phone-based travel diaries (Stopher and Shen, 2011; Houston et al., 2014) are commonly used in travel surveys to collect detailed trip chain data, including home and workplace locations, transport modes, and route numbers. However, manual entry requirements limit their scalability.

Given the limitations of individual data sources, recent research has emphasized data fusion approaches in achieving a more holistic understanding of mobility patterns. Graells-Garrido et al. (2023) fused mobile phone application usage data with official surveys and census data using matrix factorization to create updated human mobility patterns. Liu et al. (2025) developed a fusion model combining mobile signal, smart card, and train operation data to adjust inferred passenger travel paths using journey time distributions. Similarly, Sipetas et al. (2020) integrated train operations data with object detection techniques to analyze crowding patterns. Additionally, Krueger et al. (2023) combined household travel survey data with ride-hailing trip records in a unified modeling framework, correcting for sample bias and thereby providing more reliable estimates of ride-sourcing demand and mode choice behaviour.

Other studies have also applied data fusion methods to different combinations of transport-related data sources. Roncoli et al. (2023) fused Automatic Passenger Counting (APC) and Automatic Vehicle Location (AVL) data to improve the quality of real-time information on passenger volumes within public transport networks. Beyond operational benefits, the method also contributes to understanding travel behavior and patterns, for example, by offering insights into passengers' responses to crowding conditions. Kusakabe and Asakura (2014) proposed a data fusion method that relies on smart card data and person trip survey data for estimating behavioral attributes. Papacharalampous et al. (2016) developed an approach to fuse a variety of data referring to multi-modal flows (e.g., camera-based pedestrian counts, car parking utilization data, counts of arena spectators) in the frame of a large-scale urban event in Amsterdam, Netherlands. Among others, their goal was to identify behavioral differences between transport modes and support planning decisions related to parking allocation and public transport scheduling during major events. By leveraging the strengths of multiple data sources, these fusion techniques provide a more comprehensive and robust representation of urban mobility dynamics.

2.2. Transport demand profiling

Clustering is often utilized for analysing travel demand, as it provides more detailed insights into the travel behaviour of distinctive user groups which have similar travel patterns. Focusing on studies that utilized automated data collection sources, the existing literature mostly includes studies that cluster travelers based on their temporal or spatio-temporal characteristics through smart card data. The majority of such studies consider particularly the temporal aspect of travel demand, with similar patterns emerging in multiple studies. Cats and Ferranti (2022) analysed the travel demand data in Stockholm and clustered it based on the time-of-day travel patterns and the day-of-the-week travel patterns. The former reveals five distinct groups. The largest (37 %) and most typical is the 2-peak pattern with a morning and afternoon peak in demand, most likely indicating travel to and from work. The second (31 %) exhibits a similar pattern, although with less pronounced peaks and comparatively more travel during midday. The third (15 %) is a 3-peak traveler group, with a third peak in the late evening (around 10pm), possibly after a social activity. Next (11 %) is a 2-peak traveler group with an earlier morning peak, before 6am, and also a slightly longer tail later into the evening. Finally, the last group (6 %) exhibits no clear peak behaviour, with fairly equally distributed demand from early morning to late evening, with the highest concentration in the afternoon (around 4pm).

Similar pattern combinations are also reported by other studies. In their study on Shenzhen, Lin et al. (2022) find 35 % of travelers align with the 2-peak pattern, namely in the morning and afternoon, with a smaller group within that starting somewhat earlier in the morning. Similarly to Cats and Ferranti (2022) they observed a cluster with 3-peaks, including a late evening peak, although their group is smaller at only 12 %. The 2-peak group was also observed in Gatineau, representing almost half of all travelers (Agard et al., 2006). Curiously, in addition to this group, they also find 14 % of travelers travel with 2 peaks, but in the morning and midday, before the afternoon peak. For the rest, another 15 % have an even distribution of trips over the day and a quarter seem to not travel much at all. Briand et al. (2016) carried out an analysis in Rennes. In their case, a 2-peak pattern can be for only 22 %. Also, while they do report 3-peak travelers (16 %) as well, the third peak is midday, not late evening, like reported by Cats and Ferranti (2022) and Lin et al. (2022). Most striking in Rennes may be a group with only an afternoon peak, representing almost 20 % of travelers. This resembles the last group of Cats and Ferranti (2022), albeit constituting a much larger share of the population. A study on Sejong City (Lee et al., 2022) reports a full 20 clusters, although differences between certain clusters are minimal, leading to groups of patterns, largely resembling previously cited research, with 2-peak (morning, afternoon) and throughout-the-day distributed patterns dominating the results. It is interesting to see that multiple studies report the typical 2-peak morning-afternoon pattern, with other patterns being less consistent between them. Additionally, the size of the respective groups also seems to vary significantly between studies. This can largely be attributed to differences in contexts, such as the size of the city, available modes, share and characteristics of the population using public transport etc.

A few studies also went further, analysing changes in the behaviour of travelers when measured at multiple points in time and the migrations between clusters or rather the formation of new clusters. According to Lin et al. (2022), having analysed 7 years of data (2011 to 2017) in Shenzhen, while the clusters stay largely similar year-to-year, the size of each cluster changes. This shows that behavioural patterns are consistent, while travelers change which one they belong to based on their current activity schedule. Similarly, Briand et al. (2017) find individuals largely stay within the same cluster or move to a similar one. Examining a much smaller time scale of 12 weeks Agard et al. (2006) also reports similar findings. He et al. (2021) carried out a clustering analysis of travel demand based on temporal and spatial characteristics in Ottawa. The results of spatial clustering primarily reveal different origin points for daily commuters into the city center of Ottawa. A full overview of spatial and temporal patterns observed in public transport demand can be found in the literature review of Cats (2024).

Focusing now on non-automated data sources, the majority of studies that approach demand clustering utilize travel surveys. Some of these studies also include the modal aspect in their approaches. In (Molin et al., 2016), the authors utilize survey data drawn from a national panel in the Netherlands to investigate multi-modal travel groups and attitudes. In their findings, the authors report that the majority of people (65 %) are multi-modal, with a somewhat higher usage of car (27 %), bicycle (24 %) or public transport (14 %). Only 17 % are found to be heavily uni-modal, making almost exclusive use of the car. And while behaviour of people tends to change, Kroesen and van Cranenburgh (2016) utilize data from the German panel and find that no specific user group is more or less predisposed to either and that travelers of all sorts are equally likely to remain in a specific pattern or switch. In (Ton et al., 2020), the authors use data from the Netherlands mobility panel, combined with a companion survey, to investigate daily mobility patterns and their relationship with attitudes towards modes. Their findings indicate that most travelers follow multi-modal mobility patterns in their daily activities and tend to have a more favorable attitude toward the transport modes they use compared to those they do not. Olafsson et al. (2016) explored modality styles in Denmark through survey data on transport modes and travel purposes. The study emphasizes on cycling in multi-modal transport behaviours. The findings suggest that travel behaviour is predominantly multi-modal with cycling taking place in all the identified modality styles.

Wang et al. (2022) utilized survey data from low-income communities of Michigan to investigate the travelers' preference towards shared mobility. Their study considers ride-hailing and on-demand services in addition to conventional fixed route buses. The findings suggest three latent segments, referring to shared-mode enthusiasts, shared-mode opponents, and fixed-route transit loyalists. Ride-hailing services are also considered in a recent study by Iogansen et al. (2025) that approaches multi-modal travel patterns through a combination of traditional questionnaires and GPS-based travel diaries. The authors focus on California and identify four traveler groups, referring to drive-alone users, carpoolers, transit users and cyclists.

2.3. Research gap

Contemporary needs of advanced multi-modal networks require up-to-date information about travelers' movements in order to reveal current travel patterns and assist effective decision-making and policy-shaping. Existing literature includes several studies that utilize infrequently collected survey data for profiling demand from a multi-modal perspective. Studies that utilize automated technologies for revealing travel patterns are mostly based on smart card data and refer to either temporal or spatio-temporal aspects. Therefore, there is an apparent need to investigate the potentials of emerging automated data sources into offering holistic views of temporal, spatial and modal mobility patterns that will be aligned with increasing needs of complex multi-modal urban networks for frequent and detailed information of traveler movements.

3. TravelSense data

The Helsinki Region Transport, a public transport authority of Helsinki metropolitan region, abbreviated as HSL (based on Finnish "Helsinki Seudun Liikenne"), operates a comprehensive ticketing and information application for smartphones that allows travelers to select the best travel route for their trip, book the right ticket for their itinerary, and learn about news related to the local network, among others. This application refers to both public transport users and non-users. For example, the latter could use it as a source of information about disruptions in the local network that might affect their trip with a private car or an active mode. Besides its basic functions, this smartphone application allows recording traveler movements, as long as they consent, under a strict framework that ensures user privacy and anonymity. Data collection depends on a set of sensing technologies referring to both network infrastructure and users' portable devices. The output is a data set called "TravelSense", which constitutes a form of digitally collected travel diaries. The information allows trajectory-level details of how local travelers move in a spatio-temporal perspective including also the utilized travel modes. This ticketing application includes all modes of the Helsinki region's public transport network, including bus, tram, subway, train and ferries. All modes except for ferries are included in the current study. The modes extend beyond public transport, including active modes (walking, running, cycling) and private cars. This mode-related information is a feature that makes this source of specific interest for multi-modal applications. In the following we provide details related to the TravelSense data.

3.1. Infrastructure

The data collection infrastructure relies principally on HSL's smartphone app. Through the app, a user's device is able to recognise Bluetooth low energy beacons throughout the public transport network (Fig. 1). Each device is assigned a random ID every day for which it shares data. The app also uses the mobile phone's GPS coordinates to determine users' coarse-grained locations and trajectories. Finally, the app uses the activity recognition modules of the user's mobile phone to determine whether the user is still, walking, cycling, or on board a vehicle.

Therefore, the physical sources (Fig. 1a) used in the data collection can be classified in the following way:

1. **Moving Bluetooth beacons**: These are installed inside public transport vehicles (buses, trams, trains, subways and ferries).
2. **Portable devices**: These are the users' mobile phones, which recognise the user activity and movement and also recognise the Bluetooth beacons throughout the public transport network.

The HSL app (Fig. 1b) logs beacon recognition and pushes this information along with position and activity information to the TravelSense servers. TravelSense data are pivotal in detailing complete trip chains, capturing the nuanced stages of a traveler's journey from start to finish. Fig. 2 illustrates a representative example of a traveler's path from the initial point to the final destination, enriched with the data sources utilized at each stage:

1. The origin is pinpointed as a place of prolonged inactivity, typically indicating a location of long stay, sensed through the mobile phone application.
2. The traveler's walk from the origin to the bus stop involves active sensing by the mobile phone application, depicted by a black bar in the figure.
3. Onboard bus time, off-board bus time and bus vehicle information are logged through interactions with Bluetooth beacons installed on the bus, shown as a light blue bar in Fig. 2.
4. The transition to the subway involves another segment of the walk detected by the mobile phone application (black bar).
5. Subway travel time and vehicle information are captured through Bluetooth beacon interactions within the subway, highlighted by an orange bar in Fig. 2.
6. The final segment of walking to the destination is again tracked by the mobile phone application (black bar).
7. Destination, similar to the origin, is determined as the place of long stay.

This example demonstrates the ability to detail each component of the journey, from walking segments to time spent on various modes of public transportation. The specific measures used to sense and record these travel activities are further described in Huang et al. (2022).

3.2. Privacy protection

The TravelSense app requires users to opt-in for data consent, and it does not collect any location information until users give their permission. The asked permissions consist of two parts: location and activity. The right panel in Fig. 1b shows the screenshot of the

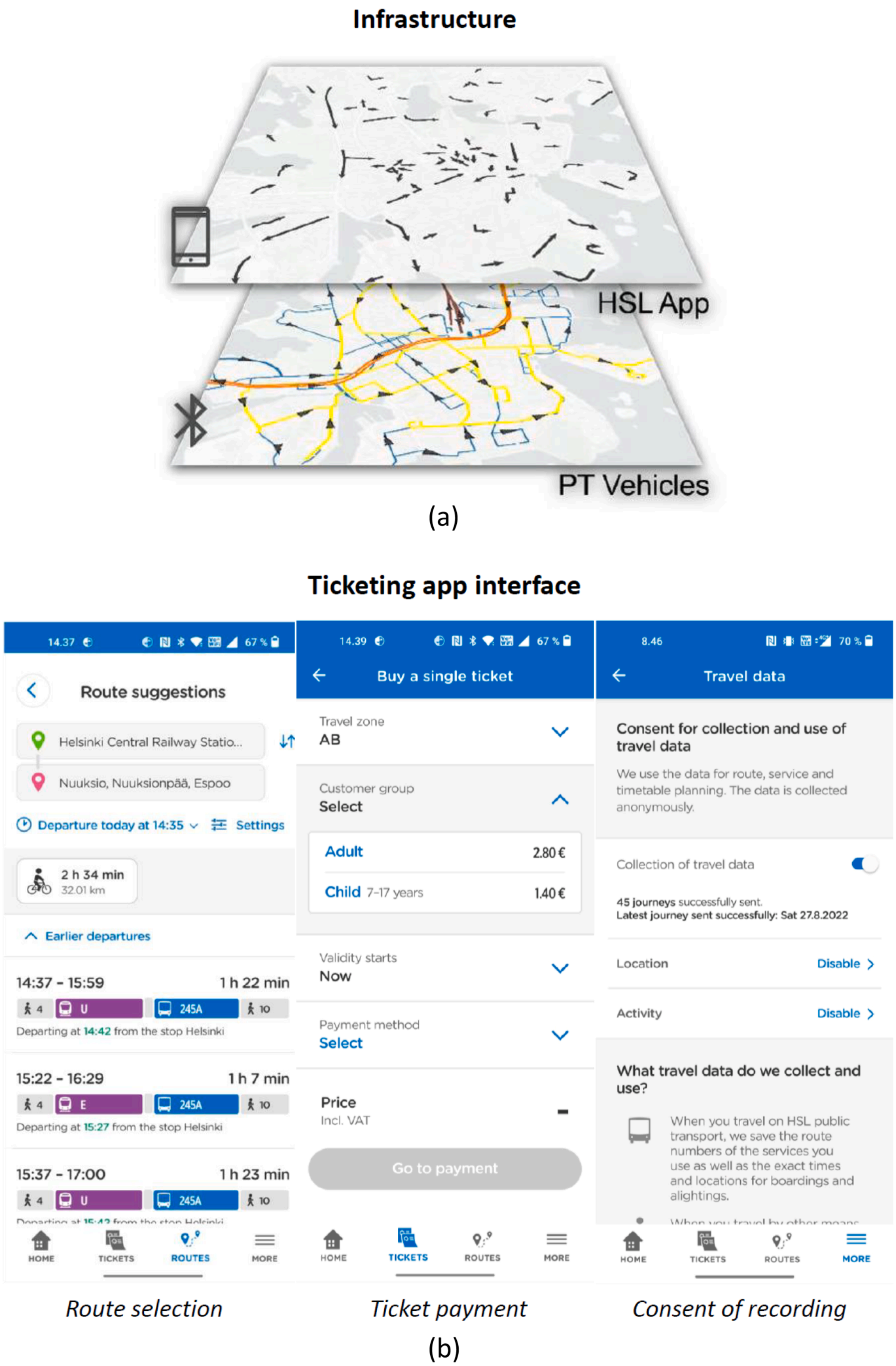

Fig. 1. HSL ticketing app (a) infrastructure; (b) interface screenshots.

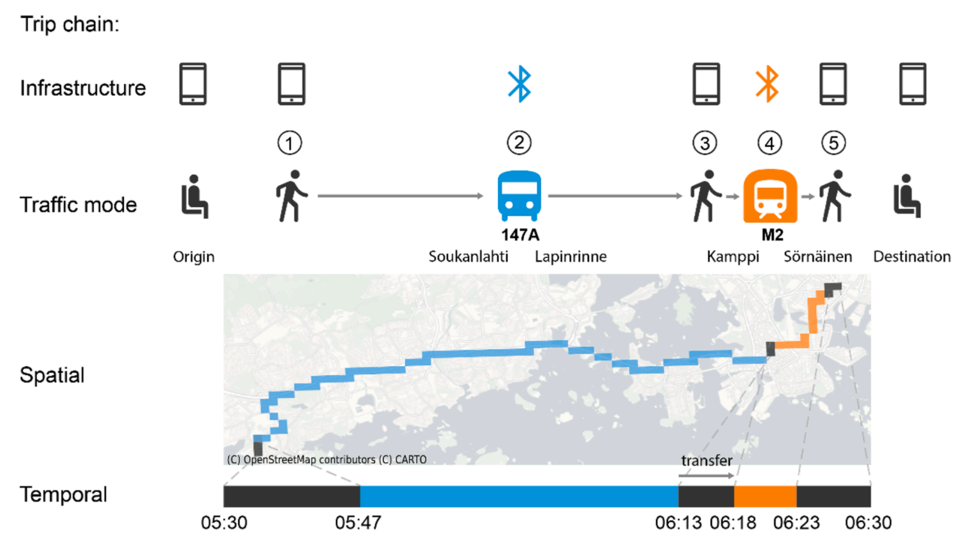

Fig. 2. An illustration of how TravelSense senses a trip chain.

consent for travel data. To ensure anonymity and prevent individual identification, each user's mobile phone is assigned a random ID every day they share data. Within the network, individuals' locations are determined by recognizing Bluetooth beacons located at bus stops (within Helsinki city limits), train and subway stations, and inside vehicles. For locations outside the network, GPS data are used, but actual coordinates are not recorded; instead, they are reported in a coarse-grained manner. Locations of app users are resolved up to grid cells of dimension 250 m × 250 m, which is provided by Statistic Finland. For privacy purposes, the timestamps of journey segments outside the network are obfuscated by rounding to the nearest quarter hour, either backwards or forwards in time.

3.3. *TravelSense in related work*

The first study that refers to TravelSense data is that by Huang et al. (2022). In this study, the authors introduced TravelSense as a data source for sensing mobility patterns, with emphasis on the details of the sensing technology and its potential contribution to studies that target multi-modality. The first application of TravelSense data in a study related to multi-modality is made in Sipetas et al. (2024). This study employs TravelSense technology to quantify stop-level transfers within Helsinki region's transport network, as part of a methodology that assesses multi-modal network performance in terms of transfers and network connectivity. On-going studies with TravelSense data include an approach for detecting changes in multi-modal travel patterns during special events (Huang et al., 2023a), fusing TravelSense with APC data for constructing O-D matrices for train operations (Huang et al., 2023b), and revealing spatial mobility patterns through temporal demand profiles (Espinosa Mireles de Villafranca et al., 2024). To the best of the authors knowledge, there are no other studies published in literature that have used TravelSense data.

4. Methodology

In order to achieve our research objectives, we develop a methodology for revealing multi-modal mobility patterns within complex public transport networks utilizing a single data source that resembles digitally collected travel diaries. Our approach is based on a commonly established profiling technique, the LPA. The generated profiles are aspired to include all three fundamental aspects of human mobility, including temporal, spatial and modal patterns. Changes in profiles across three consecutive years are also analysed to investigate the dynamic nature of human mobility. In order to achieve the study goals, a set of steps are required, which are presented in detail as follows. The overall workflow of the proposed methodology is illustrated in Fig. 3.

4.1. *Selection of variables*

The first step of the proposed method refers to identifying the variables that can be extracted from the available dataset and are relevant for the purpose of this study. In existing literature on profiling users based on smart card data, a common approach is to divide data according to whether they refer to activity or spatial information (El Mahrsi et al., 2017; Briand et al., 2017; Manley et al., 2018; Gutiérrez et al., 2020). This study targets at investigating also the critical aspect of modal choices and overall multi-modality in traveling. This information is available in digitally collected travel diaries, like TravelSense. Therefore, we introduce the modal variables and, in overall, we focus on the three categories of temporal, spatial, and modal variables.

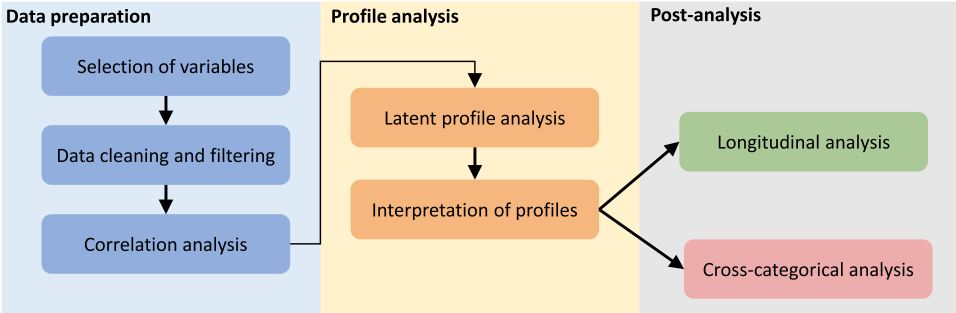

Fig. 3. Flowchart of the proposed methodology.

4.2. *Data cleaning and filtering*

In order to clean and filter the raw data, the first step is to determine logical criteria for removing noisy records. Such records include data that do not follow common logic, as for example records with zero travel time and long travel distance. The TravelSense raw dataset might include noisy data resulting, for example, from technical issues associated with the various sensing technologies involved with the data collection. In addition, records corresponding to inactive users should be detected and removed. Depending on the selected variables for creating profiles, it is up to those implementing the proposed methodology to identify such logical criteria and remove the detected records. It is also critical to remove extreme values that serve as outliers. In this study, we consider as outliers all values of each variable that lie outside the range of values $[Q1 - m \times IQR, Q3 + m \times IQR]$, where $Q1$, is the first quartile, $Q3$ is the third quartile, IQR the interquartile range defined as $Q1 - Q3$, and m a multiplier. We consider $m = 3$ which is a non-strict case, which allows to retain more extreme values in the dataset.

4.3. *Correlation analysis*

This study prioritizes obtaining explainable results that can be clearly interpreted, providing valuable insights into multi-modal traveler profiles. In this direction, correlation matrices are constructed for each category of variables separately. Between two highly correlated variables (i.e., correlation greater than 0.50 in absolute value) the one that supports better the interpretability of results is kept. In this way, we ensure redundancy reduction, simplicity and interpretability of the results.

The R programming language provides a built-in function, *cor()*, for computing correlations, with Pearson's method as the default (based on linearity assumptions). To ensure robust and comprehensive results, we also examine two additional methods, namely Spearman and Kendall, that do not rely on linearity assumptions. All three approaches are well-established and widely used in the literature (Okoye and Hosseini, 2024).

4.4. *Latent profile analysis*

After a subset of variables for each category is retained, LPA is performed for each category separately. The R package utilized in the current study refers to *tidyLPA* (Rosenberg et al., 2018), which allows to construct the latent profiles using a systematic data approach (Wickham, 2014). More specifically, we consider the four *mclust* models from *tidyLPA*, namely EEI (i.e., equal variances, equal independent covariances), EEE (i.e., equal variances, equal covariances), VVI (i.e., variable variances, independent covariances), and VVV (i.e., variable variances, variable covariances). These models differ in relation to the assumptions they make about the variances and covariances of the profiles. In the proposed framework, we test and evaluate all four models with a number of clusters ranging from two to seven (i.e., 24 models in total) and we evaluate the results through well-established statistical criteria. More specifically, we consider the following criteria available through *tidyLPA*:

- **Akaike information criterion (AIC)** (Akaike, 1987): This index is based on the log-likelihood and the number of estimated parameters, with the sample size not directly included in the AIC estimation formula. In *tidyLPA*, the lower the value, the better the performance.
- **Bayesian information criterion (BIC)** (Schwarz, 1978): This index is based on the log-likelihood, the number of estimated parameters, and the sample size. In *tidyLPA*, the lower the value, the better the performance.
- **Posterior classification probabilities**: Indicate whether the observations are classified in the correct class. In *tidyLPA*, the minimum and the maximum probabilities are reported as "prob_min" and "prob_max", respectively. The higher the values, the better the performance.
- **Entropy** (Celeux and Soromenho, 1996): Quantifies the ability of the model to generate well-separated profiles and is calculated as the weighted average of posterior classification probabilities. In *tidyLPA*, the higher the value, the better the performance.
- **p-values from the Bootstrapped likelihood ratio test (BLRT)** (McLachlan et al., 2019): Indicate for each k-profile model whether adding the k^{th} profile adds significantly to model fit compared to the model with k-1 profiles. In *tidyLPA*, this p-value is reported as "BLRT_p". The lower the value, the better the k-profile model compared to the model with k-1 profiles.

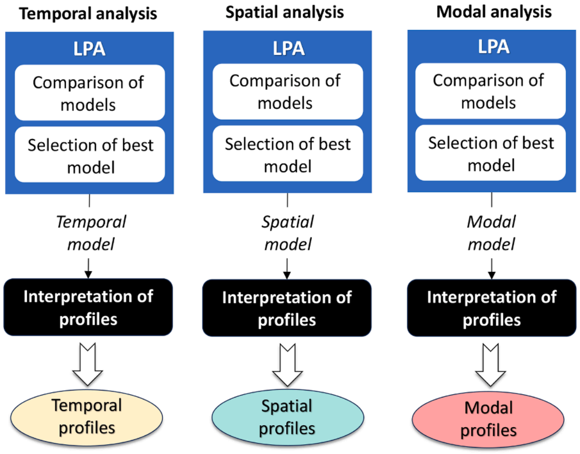

Fig. 4. The demand profiles estimation scheme.

In addition to the above, *tidyLPA* also reports "n_min" and "n_max" which are the percent of observations that are included in the smallest and in the biggest profile, respectively. Literature includes different approaches for evaluating the candidate models. For example, in Spurk et al. (2020) the authors state that 78.3 % of the studies they reviewed utilized BIC, while AIC was less frequently applied (i.e., 58.7 %). In (Ton et al., 2020), the authors use the criterion of minimum class size equal to 8 % of the sample size, among others. In the following, we consider models that result in n_min greater than 10 %, to ensure that our profiles represent significant proportions of the local travelers. Our proposed evaluation framework is well-aligned with other transportation studies, such as Gutiérrez et al. (2020), in which a combination of the aforementioned quantified criteria, personal judgement and domain expertise are used for selecting the best models that generate meaningful profiles with invaluable insights about spatio-temporal multi-modal mobility.

4.5. Interpretation of profiles

After LPA determines the probability of each record belonging to each profile, descriptive statistics per variable and per profile can be calculated. We consider a set of metrics in interpreting the profiles using boxplots that include mean value, median value, interquartile range, outliers, and whiskers. In our analysis we mostly focus on mean values and interquartile range to draw conclusions regarding the main characteristics of each profile. All other metrics are also considered per case. It is important to identify these characteristics by also examining how each profile compares with the others in terms of each variable included in the LPA. Each profile is given a name which reflects the characteristics that make the profile stand out compared to the others. The output of this step is a set of profiles per category, with each profile being assigned a name and a description of main characteristics. The overall methodological scheme for the demand profiles estimation is summarized in Fig. 4.

4.6. Longitudinal analysis

The longitudinal analysis in this study refers to the comparison of demand profiles that are obtained across different time periods, however, without monitoring the same traveler over time due to the nature of the utilized data. This analysis is performed for each category of variables separately. The profiles obtained for one week of a certain year, are compared with the profiles generated for one week of the following year. The comparison is performed in terms of profiles' descriptive statistics and content.

It is important to account for differences caused by seasonal or other expected factors, hence, the compared time periods should be considered equivalent in terms of travel behaviours. For example, a simple way to achieve this is to consider weeks that belong to the same month of each year. Identifying profiles that remain persistent across multiple time periods implies that relevant decision and policy-makers can account for such profiles with high certainty while shaping mobility plans. Observed changes in profiles demonstrate trends in mobility that should be further investigated and evaluated by stakeholders while planning for the future.

4.7. Cross-categorical analysis

Cross-categorical analysis examines the relationships among profiles from different categories, providing deeper insights into the behaviours and characteristics of public transport travelers. Specifically, we apply the chi-square test, to test if there are statistically significant correlations between profiles of different categories. The null hypothesis of the chi-square test is that the data is randomly drawn and thus no correlations exist between variables; or categories in the case of this research. The expected number of travelers in a profile combination is thus equal to the product of belonging to those two profiles separately. Since we have three different categories, we design multiple contingency tables, one for each of the mode profiles. We focus on the modal profiles because this category represents the novel addition that TravelSense data allow us to introduce into the literature about traveler demand profiling through a single data set.

We then calculate the deviations between expected and the actual numbers of travelers in each combination of temporal, spatial and modal profiles. More specifically, we consider that for a total of N observations the expected number of travelers that belong to modal profile M_k (of size N_{M_k}), spatial profile S_j (of size N_{S_j}) and temporal profile T_i (of size N_{T_i}), is calculated as:

$$E_{kji} = N_{M_k} \times \left(\frac{N_{T_i}}{N} \times \frac{N_{S_j}}{N} \right) \tag{1}$$

where $\frac{N_{T_i}}{N}$ and $\frac{N_{S_j}}{N}$ represent the probability of a traveler belonging to temporal profile T_i and spatial profile S_j, respectively. Then, the percent deviation between the actual and expected number of modal profile M_k travelers that belong to the combination of temporal profile T_i and spatial profile S_j is calculated as:

$$D_{kji}(\%) = 100 \times \frac{A_{kji} - E_{kji}}{E_{kji}} \tag{2}$$

Positive values of D_{kji} imply that we have an over-representation of travelers (i.e, more travelers than expected) from modal profile M_k to the combination of temporal profile T_i and spatial profile S_j. Negative values of D_{kji} imply that we have an under-representation of travelers (i.e, fewer travelers than expected) from modal profile M_k to the combination of temporal profile T_i spatial profile S_j. This approach for conducting cross-categorical analysis allows us to detect noteworthy patterns which might become a subject of further investigation.

5. Case study setup

This study utilises TravelSense data from April 2022, April 2023 and April 2024 in order to obtain the demand profiles. April in Helsinki is considered a month that represents profiles that are not affected by extreme weather conditions. Easter holidays often occur during April, so special attention is needed to avoid the affected time periods. The temporal basis of the analysis is set to be one week of records, which is expected to capture as many travelers as possible. Since TravelSense does not allow the notion of whether the same traveler is captured over multiple days, it is unknown how many patterns that correspond to one traveler are repeated over multiple days (e.g., every day travel from home to work in the morning) in a weekly dataset. In order to include regular traveling patterns, only workdays are considered. More specifically, the weekdays that are considered for the analysis are the following:

- April 2022: Monday, 04 April 2022 to Friday, 08 April 2022
- April 2023: Monday, 17 April 2023 to Friday, 21 April 2023
- April 2024: Monday, 08 April 2024 to Friday, 12 April 2024

5.1. Selection of variables

TravelSense data allow for the quantification of a wide selection of variables that are relevant to a spatio-temporal and modal analysis of public transport trips. The variables given below are quantified on a daily basis and then combined in a weekly dataset for further analysis. It is reminded that for privacy purposes it is impossible to monitor the trajectories of the same traveler across multiple days. The LPA is performed using scaled and log transformed values of the selected variables.

5.1.1. Temporal variables

The temporal variables identified for this analysis are mostly associated with traveler's activities during distinct time periods within a day and are as follows:

1. *TotalTravelTime*: Total travel time (minutes) by the traveler in one day.
2. *NumberOfTrips*: Number of trips that the traveler performs in one day.
3. *NumberOfTripChains*: Number of trip chains for each traveler in one day.
4. *NumberOfTripsStart_0_6*: Number of trips that the traveler starts between 00:00am and 06:00am in one day.
5. *NumberOfTripsStart_6_10*: Number of trips that the traveler starts between 06:00am and 10:00am in one day.
6. *NumberOfTripsStart_10_14*: Number of trips that the traveler starts between 10:00am and 14:00pm in one day.
7. *NumberOfTripsStart_14_18*: Number of trips that the traveler starts between 14:00pm and 18:00pm in one day.
8. *NumberOfTripsStart_18_24*: Number of trips that the traveler starts between 18:00pm and 24:00pm in one day.

5.1.2. Spatial variables

The spatial variables in this study refer to the total length of the performed trips per day, as well as to the number of trips the traveler performs across the different fare zones daily. The zones are shown in Fig. 5. More specifically, the fare system in Helsinki defines values according to the combination of zones that travelers cover in a trip. The fare per traveler is not recorded directly

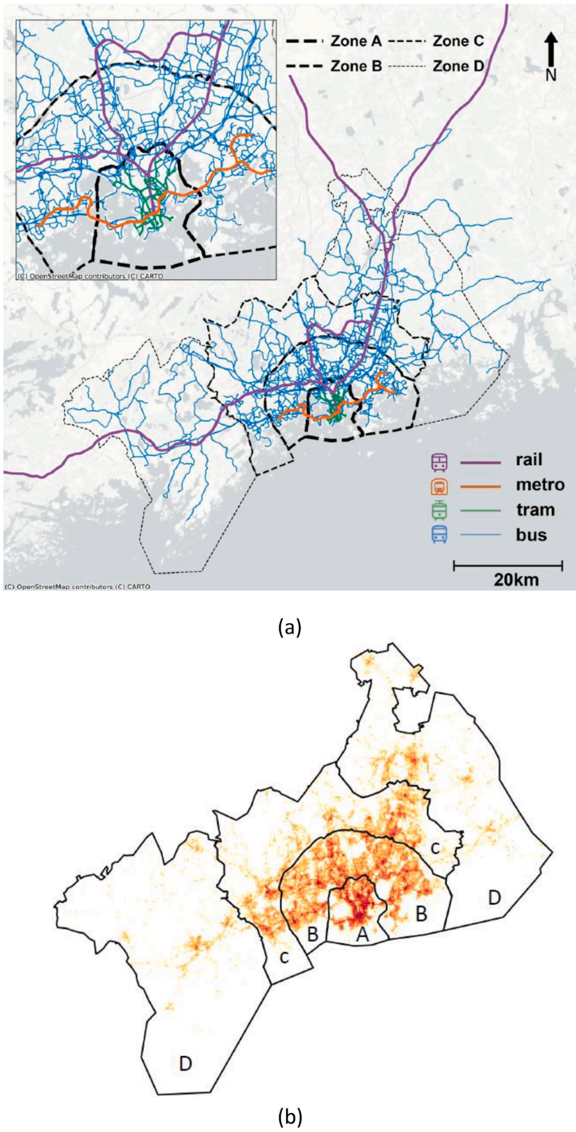

Fig. 5. Map of the Helsinki public transport network including a) public transport network (Source: Sipetas et al. (2024)), and b) TravelSense-based demand distribution for April 2022 (darker colors represent higher demand density).

in the TravelSense dataset. However, the zone combination that corresponds to each trip can be easily inferred. Considering the value of such information for shaping policies related with fare and zone management, in this study we do not consider each zone independently, but we study the combination of zones that corresponds to different fare values. This approach also allows to include origin-destination information at a zone level in a methodic way and without considering all possible zone combinations, which would increase the number of variables and thus the complexity of our analysis significantly. There are seven types of fares that HSL offers and refer to the combinations of zones “AB”, “ABC”, “ABCD”, “BC”, “BCD”, “CD”, “D”. Zone A includes the center of Helsinki, while zone D is the farthest from the city center. Zone B is surrounding zone A and together they can be considered as the “core zone” of urban activity in the area, as indicated also from the demand distribution shown in Fig. 5b. The identified variables are as follows:

1. *TotalTravelDistance*: Total distance traveled (km) by the traveler in one day.
2. *NumberOfTripsAB*: Number of trips the traveler performs with fare type AB in one day (i.e., trips within zone A, trips from zone A to zone B and vice-versa, and trips within zone B).
3. *NumberOfTripsABC*: Number of trips the traveler performs with fare type ABC in one day (i.e., trips from zone A to zone C and vice-versa).
4. *NumberOfTripsABCD*: Number of trips the traveler performs with fare type ABCD in one day (i.e., trips from zone A to zone D and vice-versa).

Table 1
Number of records during data cleaning.

No. records	April 2022	April 2023	April 2024
Raw dataset	152,495	278,299	345,016
After removal of non-logical records	116,486	226,308	286,440
After removal of outliers	112,902	221,064	27,6776

5. *NumberOfTripsBC*: Number of trips the traveler performs with fare type BC in one day (i.e., trips from zone B to zone C and vice-versa, and trips within zone C).
6. *NumberOfTripsBCD*: Number of trips the traveler performs with fare type BCD in one day (i.e., trips from zone B to zone D and vice-versa).
7. *NumberOfTripsCD*: Number of trips the traveler performs with fare type CD in one day (i.e., trips from zone C to zone D and vice-versa).
8. *NumberOfTripsD*: Number of trips the traveler performs with fare type D in one day (i.e., trips within zone D).

5.1.3. Modal variables

The modal variables identified here include all public transport modes in Helsinki (except for ferries), active modes and private cars, aiming at offering invaluable multi-modal insights to this analysis. The modal variables are as follows:

1. *NumberOfModes*: The number of modes the traveler uses in one day.
2. *NumberOfModesPerTripchain*: The number of modes the traveler uses per trip chain in one day.
3. *NumberOfTripsWithActive*: The total number of trips with active modes (i.e., walk, run, bike) the traveler performs in one day.
4. *NumberOfTripsWithPrivate*: The number of trips with private vehicles the traveler performs in one day.
5. *NumberOfTripsWithBus*: The number of trips with bus the traveler performs in one day.
6. *NumberOfTripsWithTram*: The number of trips with tram the traveler performs in one day.
7. *NumberOfTripsWithSubway*: The number of trips with subway the traveler performs in one day.
8. *NumberOfTripsWithTrain*: The number of trips with train the traveler performs in one day.

5.2. Data cleaning

The cleaning process included the removal of outliers to ensure a higher quality of profiling. We first identified records that corresponded to inactive travelers, by assuming that these travelers were associated with zero travel time and/or travel distance and/or number of travel zones and/or number of modes per day. Hence, the identified records were removed. In addition, we implemented the IQR-based outlier detection method for some of the variables in each category. More specifically, since in some variables we had only few records with non-zero values, our goal was to ensure that these values will not be eliminated. Therefore, the variables that were used for outlier detection were the following: *TotalTravelTime*, *NumberOfTrips*, *NumberOfTripChains*, *TotalTravelDistance*, *NumberOfModes*, *NumberOfModesPerTripChain*. The variables not included in the outliers detection were the ones that refer to the pure temporal, spatial and modal distribution of travelers (i.e., the ones that refer to how many trips travelers made in each time period, in each fare type, and with each mode). As previously mentioned, the multiplier set for calibrating the outlier detection method was equal to 3, which means that the process was not strict, therefore, only highly extreme records were excluded. The number of records for April 2022, April 2023, and April 2024 at each stage of data cleaning are shown in Table 1. According to this table, the number of records increase every year.

Boxplots for the total number of trips per traveler for the clean datasets are shown in Fig. 6. According to these boxplots, April 2022 is associated with significantly lower number of recorded trips during the five weekdays studied here, when compared to the following years. April 2023 and April 2024 present almost identical statistics in terms of number of trips.

5.3. Correlation analysis

Fig. 7 presents the correlations among all the variables of each category for April 2022 using the Pearson's method. Similar patterns are observed for April 2023 and April 2024, whether Pearson's, Spearman's or Kendall's methods are used.

More specifically, the correlation check revealed that among the temporal variables, *TotalTravelTime*, *NumberOfTrips*, and *NumberOfTripChains* are highly correlated with each other (i.e., correlations of 0.50 or higher in absolute value) according to all three methods. Hence, *NumberOfTrips* and *NumberOfTripChains* are removed from the list of selected variables for this analysis.

For the spatial variables, correlations are generally below 0.50 in absolute value. A minor exception occurs between *NumberOfTripsAB* and *NumberOfTripsBC*, where Spearman shows moderate negative correlations (between -0.50 and -0.60), while Pearson and Kendall remain below 0.50 in absolute value. Both variables were retained as they represent distinct trip relations and the effect is not consistent across methods.

For the modal variables, no significant correlations are found across methods and years, apart from 2022, where *NumberOfModes* and *NumberOfTripsActive* reached 0.53 with Spearman (versus 0.48 with Pearson and 0.46 with Kendall). Both variables were kept, as the effect is limited to a single method and year, and each reflects conceptually distinct aspects of modal behavior.

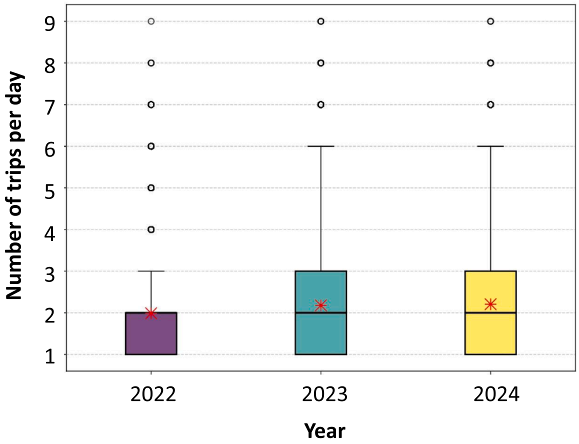

Fig. 6. Total number of trips per day per traveler for the three years of this study.

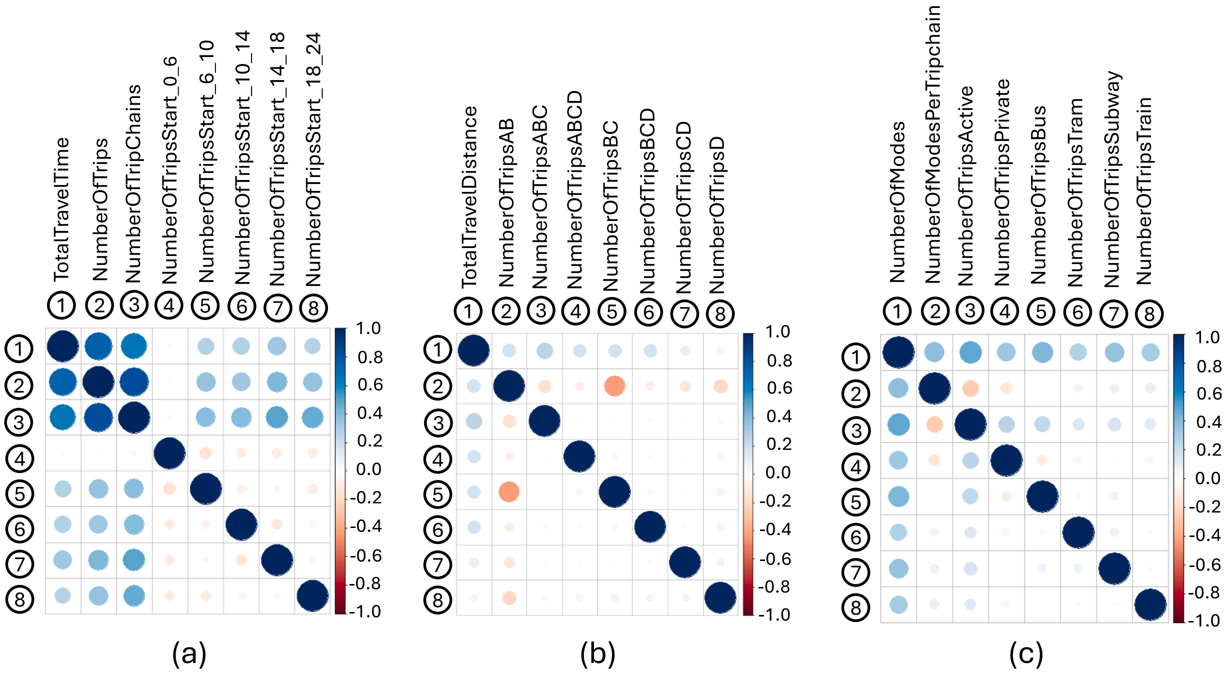

Fig. 7. Correlation matrices of a) temporal, b) spatial, and 3) modal variables for April 2022 using Pearson's method.

6. Results

6.1. Latent profile analysis best models identification

The results for the selected as best LPA models are shown in Tables 2, 3 and 4, for temporal, spatial and modal variables, respectively. The EEI model is selected as the best for temporal variables in all three years. For temporal variables of April 2022, five profiles are selected as the best fit, while for April 2023 and April 2024, four profiles are selected. Regarding the LPA for the spatial variables, for all three years the best model is EEI with three profiles. For the modal variables, the best model is EEE with four profiles for all three years.

As mentioned in the Methodology section, models with a minimum percentage of records greater or equal than 10 % are considered to ensure that each profile represents a meaningful percentage of travelers. In all evaluations performed, the model and number of profiles associated with the lowest BIC and AIC are selected as the best. All other metrics were also reviewed to confirm the selection. In our selection of best models, each variable category per year is analyzed independently. As a result, the optimal number of profiles may vary between years. For example, we identify five profiles for temporal variables during 2022 and four profiles for these variables in later years. It is highlighted here that LPA uses probabilistic estimation to determine profile membership. In our study, the mean

Table 2
Latent profile analysis selected models for temporal variables.

Year	Model	Profiles	AIC	BIC	Entropy	prob _min	prob _max	n _min	n _max	BLRT _p
2022	EEI	5	128860.12	129245.49	0.98	0.96	1	0.14	0.31	0.01
2023	EEI	4	438952.39	439292.49	1	1	1	0.17	0.3	0.01
2024	EEI	4	559306.31	559653.83	1	1	1	0.16	0.32	0.01

Table 3
Latent profile analysis selected models for spatial variables.

Year	Model	Profiles	AIC	BIC	Entropy	prob _min	prob _max	n _min	n _max	BLRT _p
2022	EEI	3	1794771.98	1795099.55	1	1	1	0.25	0.4	0.01
2023	EEI	3	3580446.5	3580796.91	0.99	1	1	0.24	0.45	0.01
2024	EEI	3	4479988.9	4480346.95	0.99	1	1	0.25	0.46	0.01

Table 4
Latent profile analysis selected models for modal variables.

Year	Model	Profiles	AIC	BIC	Entropy	prob _min	prob _max	n _min	n _max	BLRT _p
2022	EEE	4	1336287.11	1336971.14	1	1	1	0.12	0.38	0.01
2023	EEE	4	2672018.8	2672750.54	0.99	0.99	1	0.14	0.4	0.01
2024	EEE	4	3425985.51	3426733.21	0.98	0.98	1	0.13	0.45	1

probability of all observations belonging to the assigned profile of each one of the three categories is almost equal to 1 for all three time periods considered here.

6.2. Profile analysis and interpretation

This section includes the profile analysis and interpretation for the temporal, spatial and modal profiles of April 2022. The results for April 2023 and April 2024 are presented in detail in Appendices A, B, and C, for temporal, spatial, and modal profiles, respectively.

6.2.1. Temporal profiles

Boxplots for each variable and profile for temporal variables are shown in Fig. 8. With these statistics we can analyse the results of the LPA and therefore interpret the generated profiles. The profile interpretation is also enhanced by the histograms shown in Fig. 9. Therefore, the obtained temporal profiles for April 2022 are the following:

- **Profile 1 - Morning peak travelers with moderate travel times (31 %):** These travelers have moderate daily travel times and make a few trips between midnight and 06.00. Along with profile 5, they record the highest number of trips during the morning peak (06.00–10.00), while they do not travel between 10.00 - 14.00.
- **Profile 2 - Midday travelers with moderate travel times (22 %):** These travelers also have moderate travel times and do not travel during the morning peak (06.00–10.00). They make a few trips between midnight and 06.00 but are most active between 10.00 and 14.00, similar to profile 5.
- **Profile 3 - Afternoon peak travelers with short travel times (17 %):** This profile has the shortest daily travel times, making trips primarily between midnight - 06.00 and 14.00–18.00. During these periods, they have the highest number of trips compared to other profiles, with limited travel in the evening (18.00-midnight).
- **Profile 4 - Evening travelers with short travel times (16 %):** These travelers have relatively short daily travel times and travel exclusively between 14.00 - midnight. They are most active during 18.00 - midnight, making more trips during this period than any other profile.
- **Profile 5 - All day travelers with long travel times (14 %):** This profile has the longest daily travel times and makes trips throughout the entire day. They are particularly active during the morning peak (06.00–10.00), similar to profile 1.

6.2.2. Spatial profiles

Considering the boxplots of Fig. 10 and stacked bar chart of Fig. 11, the obtained spatial profiles for April 2022 are described as follows:

- **Profile 1 - Long distance core zone travelers (40 %):** This profile, along with Profile 3, covers the longest daily travel distances. They make trips that correspond to all zones, though with varying frequency. Notably, they record a significantly higher number of trips with fare type AB compared to other profiles.

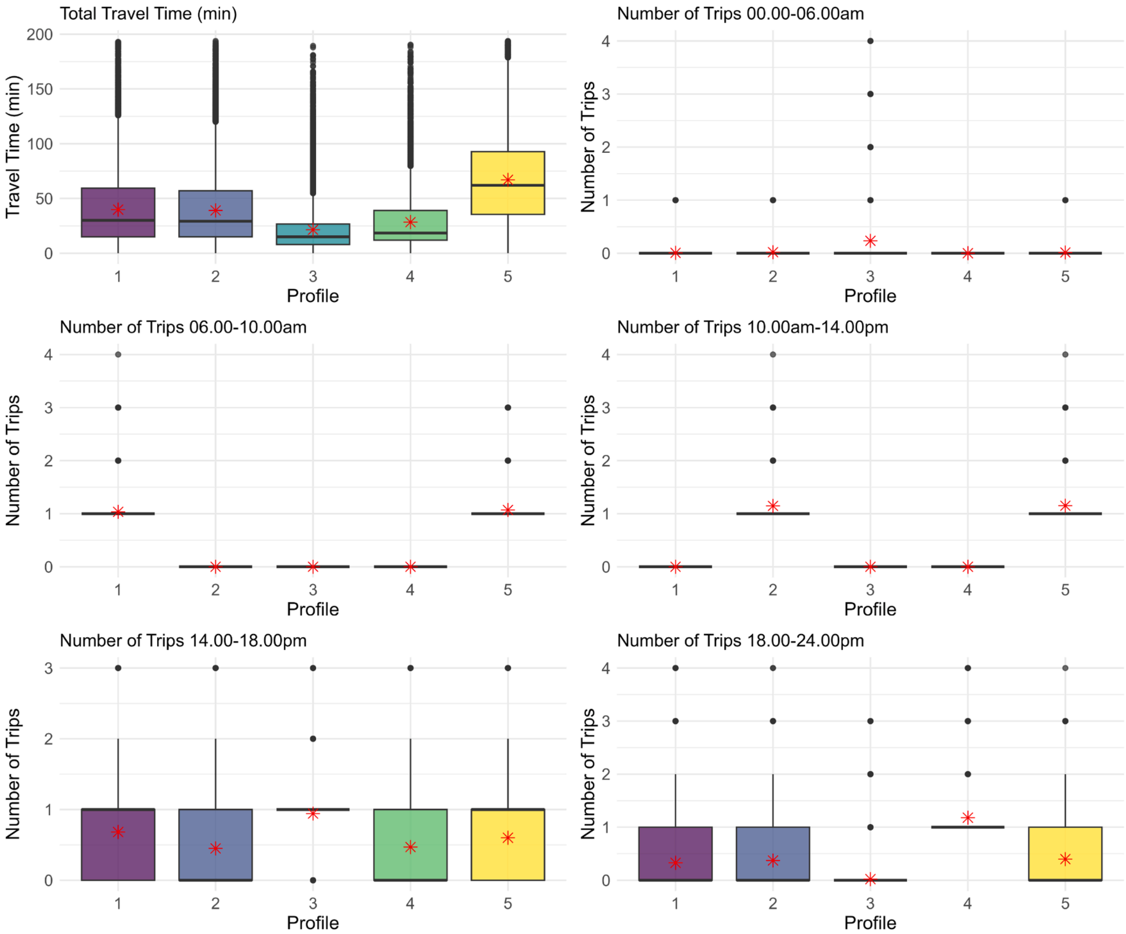

Fig. 8. Boxplots of temporal profiles for April 2022.

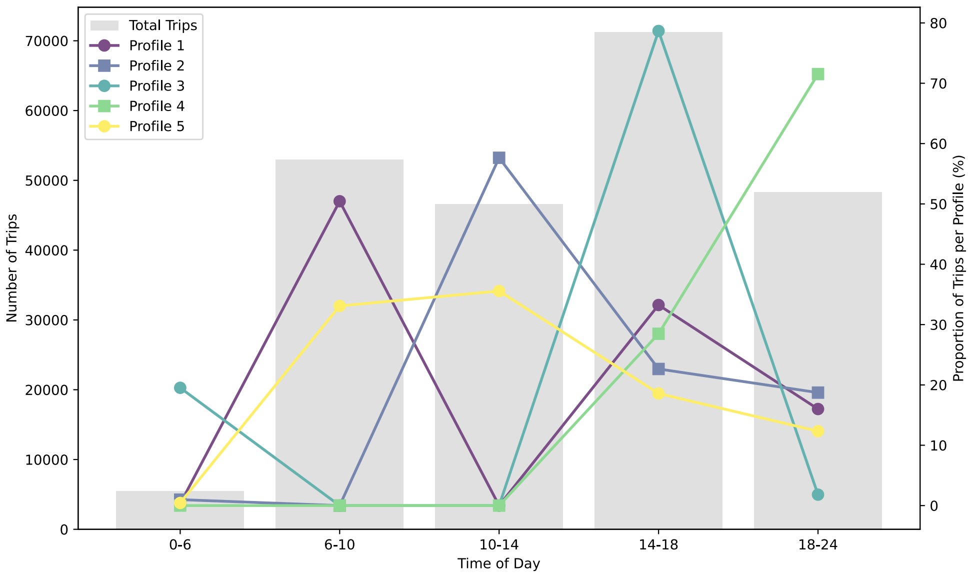

Fig. 9. Histogram of number of trips per traveler and time period for April 2022.

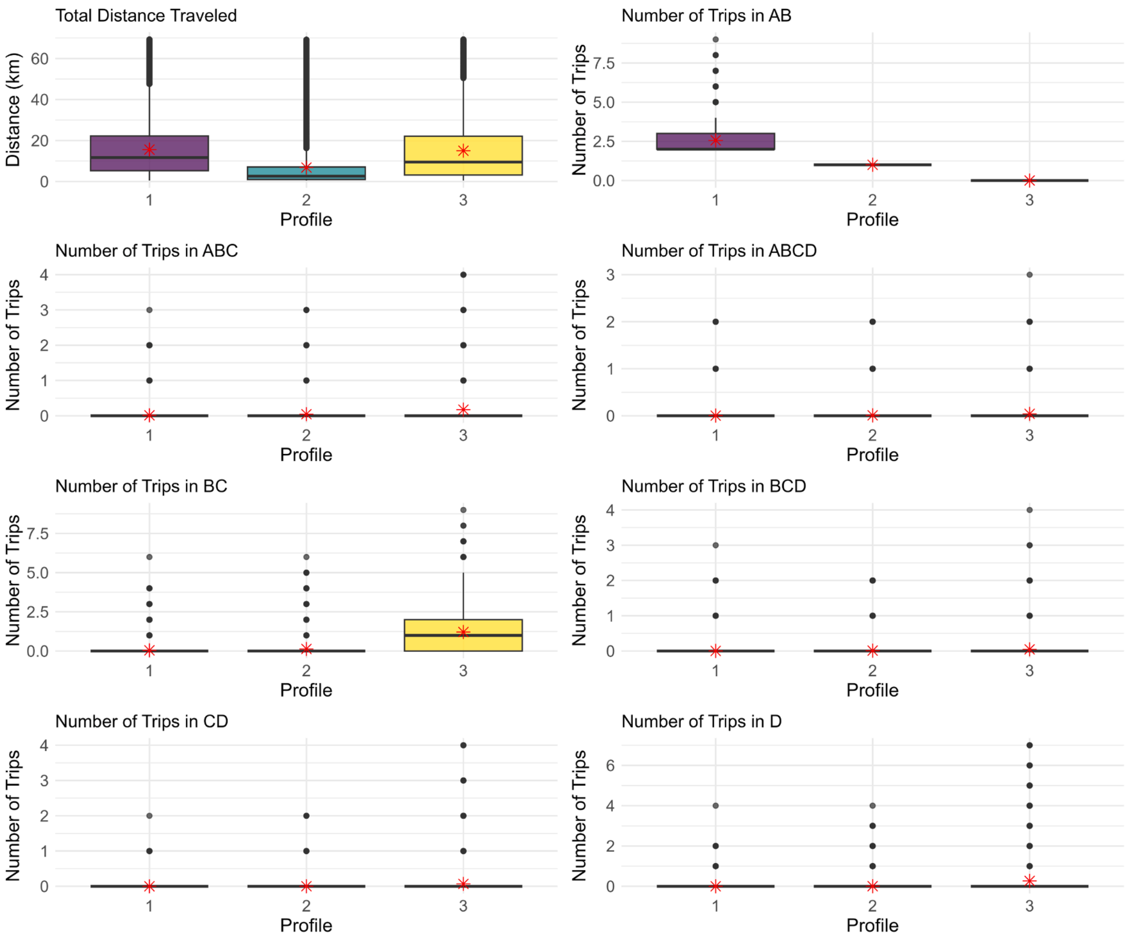

Fig. 10. Boxplots of spatial profiles for April 2022.

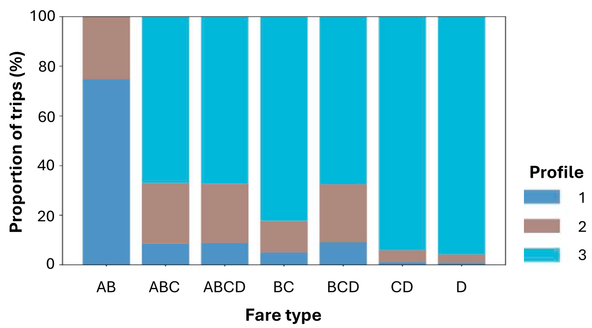

Fig. 11. Stacked bar chart with proportion of spatial profiles per fare type for April 2022.

- **Profile 2 - Short distance travelers with a single core zone trip (35 %):** These travelers have the shortest total travel distances. While they make trips in all zones, their activity in some areas is minimal. A defining characteristic of this profile is that they consistently make exactly one trip per day with fare type AB.
- **Profile 3 - Long distance travelers beyond the core zone (25 %):** Similar to profile 1, this profile covers long distances. However, unlike other profiles, they do not make any trips with fare type AB. Across all other zone combinations they record the highest average number of trips compared to other profiles.

6.2.3. Modal profiles

Boxplots for each variable and profile for modal variables are shown in Fig. 12. Based on them, the obtained modal profiles for April 2022 are as follows:

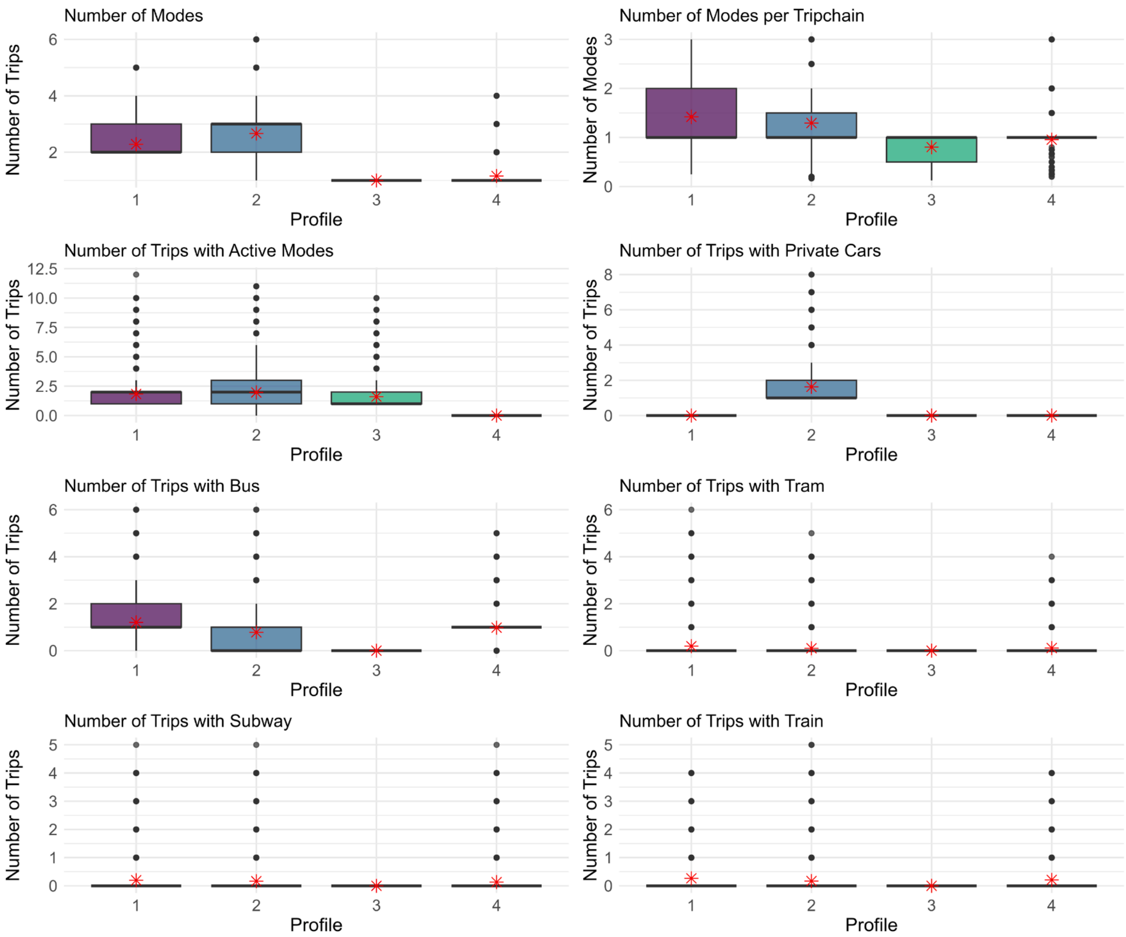

Fig. 12. Boxplots of modal profiles for April 2022.

- **Profile 1 - Highly multi-modal travelers without private cars (38 %):** These travelers use a relatively high number of transport modes per day and per trip chain, similar to profile 2. They utilize all available modes except for private cars.
- **Profile 2 - Highly multi-modal travelers using all modes (38 %):** Similar to profile 1, these travelers exhibit high multi-modality, using a variety of transport modes per day and per trip chain. Unlike any other group, they are the only profile that also includes private cars in their travel choices.
- **Profile 3 - Active mode travelers (13 %):** These travelers choose only one mode per day for traveling and more specifically, the active modes (i.e., walking, running, and cycling).
- **Profile 4 - Short multi-modality public transport users (11 %):** These travelers rely almost entirely on a single transport mode per day, with low multi-modality in their trip chains. They do not use active modes or private cars for their daily commutes.

6.3. Longitudinal analysis

Figs. 13, 14 and 15 summarize the longitudinal analysis across the three years studied here for the temporal, spatial and modal profiles, respectively. The solid arrows indicate profiles that are identical in terms of statistics resulting from the boxplots in the profiles analysis and interpretation. The dashed arrows indicate profiles that present similarities in the profile analysis and interpretation but are not identical. According to Fig. 13, April 2022 temporal profiles are distinct from April 2023 and April 2024, while the two latter are identical. The most robust are the profiles about moderate travel times morning peak travelers (dominant in all three time periods) and long travel times all day travelers (least represented profile in all three time periods). From April 2023 to April 2024, not only the profiles remained the same, but the percentage of travelers that belongs in each one of them is almost unchanged. The spatial profiles (Fig. 14) remained unchanged over time, not only in terms of content but also in terms of percentage of travelers that is represented by each one of them. Although most of the modal profiles (Fig. 15) remain stable over time, it is observed that there is a change in the percentage of travelers that belong to each profile. It is interesting that the role of bus mode is enhanced from April 2022 to April 2023, shaping a profile that most likely replaced the discontinued profile of low multi-modality public transport users.

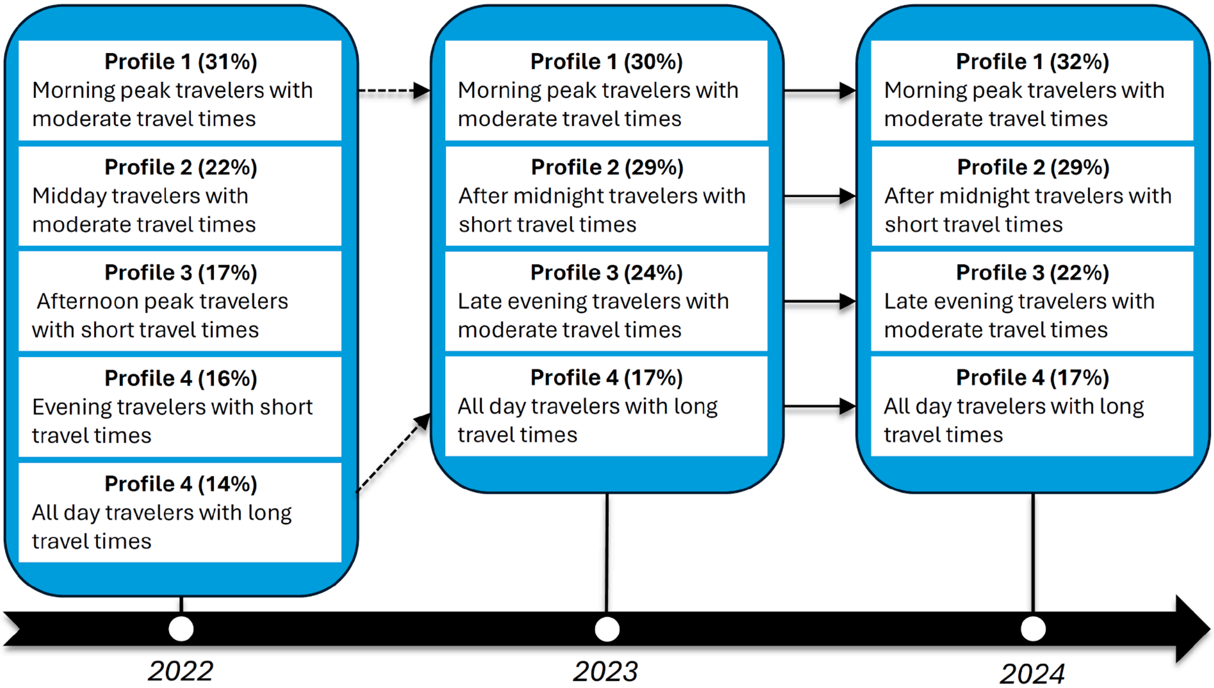

Fig. 13. Longitudinal analysis for temporal profiles.

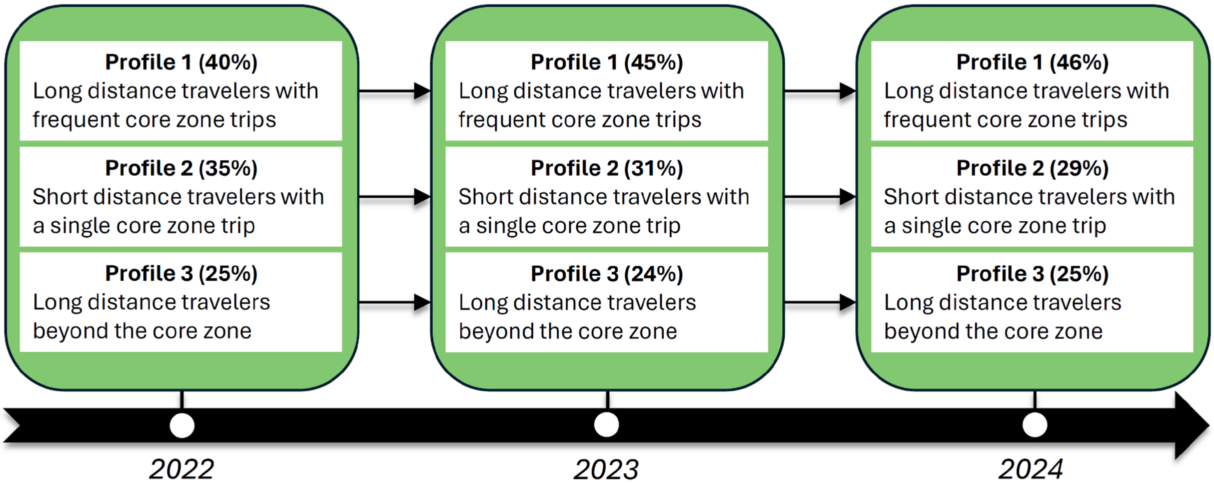

Fig. 14. Longitudinal analysis for spatial profiles.

6.4. *Cross-categorical analysis*

The cross-categorical analysis is summarized in Fig. 16 that presents the deviation between the expected and actual number of travelers that belongs to each combination of temporal and spatial profiles for each modal profile for April 2022 (Fig. 16a), 2023 (Fig. 16b), and 2024 (Fig. 16c). Darker green colors indicate combinations of temporal-spatial-modal profiles with high positive deviation (actual greater than expected). Darker orange colors indicate combinations with high negative deviation (actual lower than expected). It is apparent that April 2023 and April 2024 present the same patterns in terms of extreme positive and negative deviations per modal profile, however, it is reminded that modal profile 2 of 2023 became profile 3 in 2024, and vice-versa. Therefore, the two profiles have switched their spatio-temporal patterns between the two years. The figure reveals that high multi-modality travelers make trips that are more than expected related with morning peak activities in the core zone. Their trips are associated more than expected with moderate travel times and long travel distances. Low multi-modality travelers are less than expected associated with all day activities with long travel times and distances. In overall, low multi-modality travelers present lower deviations (either positive or negative) than high multi-modality travelers across the three years.

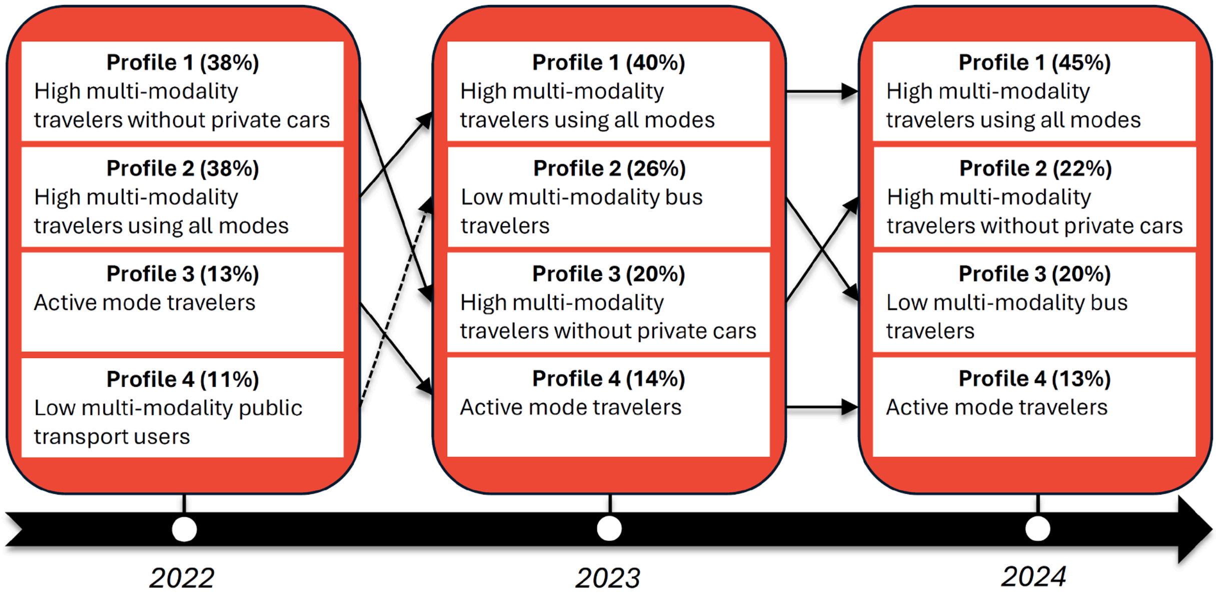

Fig. 15. Longitudinal analysis for modal profiles.

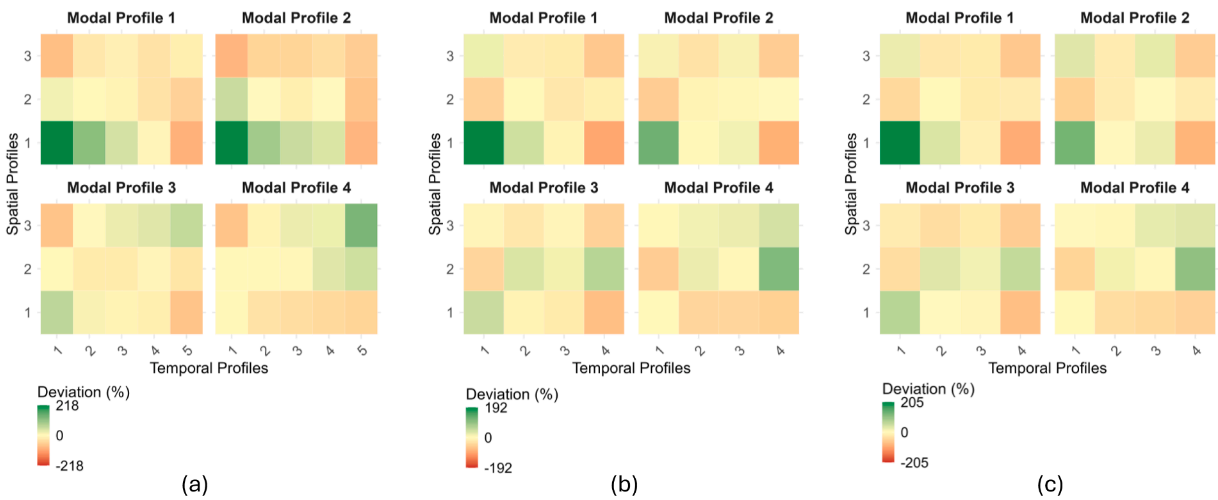

Fig. 16. Deviation between the expected and actual number of travelers in each combination of temporal and spatial profiles for each modal profile for a) April 2022, b) April 2023, and c) April 2024.

7. Discussion

7.1. Key findings

The demand profile analysis conducted in this study revealed a variety of multi-modal mobility characteristics within urban transport networks. The temporal profiles represented travelers that make trips exclusively during certain time periods, either during expected (e.g., morning peak) or less expected hours (e.g., after midnight). In accordance to existing literature in the field of demand profiling (Cats and Ferranti, 2022; Lin et al., 2022; Agard et al., 2006; Briand et al., 2016), our study identified a profile with one peak during morning peak hours and one peak during afternoon peak hours. This is temporal profile 1 in this study, which represents approximately 30 % of travelers for each year. This percentage is also similar to the ones reported in other studies (e.g., in Cats and Ferranti, 2022 and Lin et al., 2022).

The obtained spatial profiles include information about the zone combinations in which travelers move, distinguishing mainly between patterns within or beyond the core zone (represented by fare type AB). The profile that includes specifically one trip of fare type AB per day was of particular interest. Notwithstanding, the significant majority of Helsinki region travelers (more than 65 %) belongs to the two profiles associated with traveling high distances daily (more than 17km per day on average). The profile that includes travelers beyond the core zone (and likely representing also suburban and rural areas residents), requires special attention from transport operators in order to ensure that these travelers are well-connected and have smooth access to the public transport network.

Regarding modal profiles, the rail-bound public transport modes (i.e., tram, subway, train) did not shape any particular profile. Instead bus mode did, at least for April 2023 and April 2024. The active modes (referring to walking, running, and cycling) shaped a profile of travelers that choose no other mode. These travelers have downloaded the ticketing app, however, they do not travel with public transport, at least during weekdays. The concept of having pure unimodal profiles is not new in literature. Molin et al. (2016) also identified a uni-modal profile in the Netherlands, which however, referred to private cars. In our study all private car users are included also in one single profile. However, this profile, unlike the active modes profile, is also choosing other modes for traveling. Therefore, it most likely refers to travelers that have regular access to private cars, including taxis and ride-hailing services. These observations serve also as a potential indication of who is likely to use a smartphone ticketing application and to allow being tracked. The high multi-modality identified in Molin et al. (2016) for the travelers in the Netherlands is also confirmed for the Helsinki region travelers, with more than 60 % of them belonging to a profile of high multi-modality in three years studied here. According to studies, multi-modality in traveling is increasing in several industrialized countries (Kuhnimhof et al., 2012).

The longitudinal analysis showed that the temporal variables were the least robust, at least from 2022 to 2023. This is also apparent from Fig. 6, that shows much lower trips during April 2022. A solid assumption is that COVID-19 still had effects on travelers' movements during April 2022, thus triggering these differences. Spatial profiles were found to be the most stable across the three years. Their high robustness in Helsinki is also confirmed for travelers in Shenzen according to Lin et al. (2022). Regarding modal profiles they were also found to be mostly stable. The "low multi-modality public transport users" of April 2022 were replaced by "low multi-modality bus travelers" in 2023, demonstrating that the bus mode gained popularity. This could also be attributed to the COVID-19 effects that could have discouraged travelers from choosing modes with smaller space, such as buses. Changes in modal behaviour from year to year can be justified in locations with high multi-modality by existing studies which support that multi-modal individuals tend to change their mode use over time (Heinen and Ogilvie, 2016).

The year-to-year analysis of travel patterns in Helsinki should also be viewed through the lens of infrastructure changes. More specifically, Helsinki public transport network is constantly undergoing improvement interventions that lead to changes in its operational characteristics. For example, the Helsinki metro was extended during 2023. Although, there are no infrastructure changes that can be associated with the different patterns observed in our analysis period, the overall dynamic nature of the Helsinki network serves as a strong proof of how robust travel patterns, and more generally travel habits, are for the local travelers. Not only the profiles from year to year present many similarities, but also the combinatorial view across the three categories, as revealed by the cross-categorical analysis, shows that combined mobility patterns are relatively robust from April 2022 to April 2024.

7.2. Implications for digital travel diary development and deployment

From a technical standpoint, TravelSense depends on a variety of sensors which are expected to effectively communicate and collaborate with each other in order to collect the raw dataset. Similar to other sources' data collection inefficiencies met in literature (e.g., misreports in self-report travel diaries investigated in Stopher and Shen, 2011, or incomplete APC datasets addressed in Roncoli et al., 2023), this complex system leads to several technical errors that are reflected in the raw dataset. As demonstrated in this study, such errors require careful detection and removal before the data can be meaningfully analyzed. Potential users of TravelSense-type datasets should be prepared to address these challenges before utilizing such data in their projects. In addition, the raw formatting of TravelSense is not readily usable for transportation applications, since advanced data processing methods are required to transform raw sensor outputs into usable trajectory-level information. This underscores the need for expertise in data handling and processing among researchers and practitioners working with such datasets.

In this study, TravelSense allowed us to quantify a wide range of candidate variables for conducting LPA. Eight variables per category (i.e., temporal, spatial, modal) were selected as important for revealing the mobility patterns of local travelers. The correlation analysis resulted in removing two of them from the temporal variables to avoid redundancy and over-representation, while also ensuring a smoother interpretation of results. Some of the variables were strictly continuous, while others represented counts. All variables were log transformed and normalised, to be better fitted to a mixture model like LPA. Similar to Gutiérrez et al. (2020), LPA generated eligible clusters for the case of EEI and EEE models incorporated within *tidyLPA* in R.

Privacy concerns are increasingly gaining attention in recent literature, leading to the development of studies for ensuring data protection (Du et al., 2025). Traveler privacy and data anonymisation is a priority in the frame of TravelSense data as well, therefore, spatio-temporal information beyond the public transport network is aggregated. This necessary data aggregation presents certain limitations for studies requiring highly detailed location and time-specific information. In addition, privacy guidelines do not allow us to monitor the same traveler across multiple days. This is a feature that needs to be taken into consideration when choosing to use a dataset like TravelSense for longitudinal studies. In the current study, we used observations (i.e., trajectories) for travelers within five weekdays of a week per year, without knowing which of these travelers appear in more than one days. Our goal was to capture as many patterns as possible, assuming also that each weekday is equivalent to the others in terms of mobility patterns. For studies that focus on revealing variability of travel behaviour within individuals (e.g., Heinen and Chatterjee, 2015), TravelSense could not be a proper data source, at least when the analysis time horizon extends beyond a single day.

The HSL smartphone ticketing app needs some time to gain proper diffusion among the Helsinki region travelers, like any other innovative technology that is introduced to the public (Rogers, 1962). In addition, travelers need time to feel comfortable to agree on their daily movements to be recorded by the app. It is unknown what percentage of daily travelers are being recorded by the app, yet our study showed that the number of recorded travelers increases year by year. This is a very promising aspect for the future of this type of data sources. However, its quality and utility heavily depend on user acceptance of new technologies and trust in data privacy measures. As such, transport stakeholders should consider public engagement campaigns to inform travelers about the benefits of contributing to this dataset, emphasizing how their participation can enhance the overall travel experience. It should be noted that literature suggests that multi-modal individuals are more open to adopting emerging transport services (Diana, 2010). By assuming an analogy between the attitude towards emerging transport services and emerging transport applications, such as the one used for TravelSense data collection, we can reasonably expect that TravelSense will continue to see increasing adoption among multi-modal travelers in Helsinki over time.

7.3. Implications for transport policy and planning

Commonly, automated data sources that derive from public transport systems allow information that refers only to public transport movements. Although TravelSense represents a sensing technology that derives from a public transport smartphone application, it allows to capture traveler movements beyond the public transport network. Such broad information could be considered similar to smartphone-based data, however, they do not commonly include mode choices (Toole et al., 2015). TravelSense information extends beyond that of commonly used smart card data and is updated more frequently than traditional surveys. To better contextualize this novel dataset, TravelSense provides mobility information within complex networks that resembles digitally collected travel diaries. The literature includes other automatically collected travel diaries, as for example GPS-based travel diaries (Iogansen et al., 2025). TravelSense-style data can be beneficial in a variety of applications within transport networks and can be used for re-approaching existing methods that are developed to address multi-modal mobility. For example, there are many approaches for imputing OD matrices from tap-in only data in public transport networks (e.g., for Boston network in Sánchez-Martínez, 2017). Since such approaches often rely on assumptions about specific travel behaviour of the cardholders, TravelSense-style data can support model validation and offer a robust and reliable insight into travel behaviour.

The current study emphasizes on the potentials of TravelSense data in the field of demand profiling through considering spatio-temporal aspects (commonly viewed through smart card data when automated sources are considered Cats, 2024) and modal choices (commonly met in survey-based studies Molin et al., 2016; Kroesen and van Cranenburgh, 2016; Olafsson et al., 2016). Aligned with literature that supports the need for multi-day observations when studying travel behavior (Kuhnimhof et al., 2006), TravelSense allows tracking mobility patterns every day at trajectory-level. This facilitates its usage within different methodological approaches for revealing demand patterns. In our study, we chose a latent clustering technique for obtaining traveler profiles, an approach met in other studies that focus on revealing mobility patterns (McBride et al., 2018; Ton et al., 2020; Gutiérrez et al., 2020). Through this approach, we are able to uncover specific travel patterns of different travelers and thus better understand their mobility needs. For example, we can see when traditional work commutes take place, when occasional travelers make use of the system, and can potentially uncover a group of travelers that was previously unknown. The data flexibility allows to define variables targeting the support for specific policies. For example, the identification of fare zones in which travelers move, utilized in spatial profiles here, can support ticket pricing strategies.

Key findings of our study highlight the effect of utilizing TravelSense-style data into demand profiling studies and indicate further implications for policy-making. For example, the active modes profile obtained in our study is of high importance, as the literature suggests that active modes receive significant attention by governments worldwide in efforts to achieve a modal shift from motorised to active modes (Ton et al., 2020). Insights on current private car users who are potential exclusive public transport users are crucial for policymakers and transport planners aiming to develop strategies that encourage sustainable mobility by shifting demand toward public transport. Existing literature supports that a key objective of travel behaviour research is to assist policymakers and stakeholders in shaping policies that promote more sustainable mobility choices (Banister, 2008), further highlighting the importance of TravelSense-type data for analysing travel demand.

Through performing the analysis over the span of three years, we are able to assess how the behaviour and mobility needs of travelers change over time, in alignment with existing literature that highlights the importance of longitudinal approaches (Briand et al., 2017). Spatial profiles tend to be more stable, as they are likely primarily influenced by land-use policies, which are traditionally fairly stable and changes occur over longer timespans. Temporal and modal profiles, while also stable, do show some more volatility, especially in the wake of the COVID-19 pandemic, when travel behaviour was drastically affected, as confirmed by relevant studies (Kim and Kwan, 2021). The stability of the data in particular from 2023 to 2024 also informs the policymaker that the travel behaviour of users has since stabilised. The changes in profiles from 2022 to 2023 can motivate the transport authorities to update survey data that are collected before 2023. Our key findings from the longitudinal analysis in this study can serve as an alert for transport authorities and as a trigger for further response. For example, findings such that less travelers belong to the profile that does not use private cars (including taxi and ride-hailing) over the years could raise concerns about a modal swift towards less sustainable modes.

A combined evaluation of data types and their usage in demand profiling is essential, also considering the critical role that data play in decision making nowadays (Provost and Fawcett, 2013). In the future, novel data-driven studies in the field can assist in understanding behaviour for making different types of decisions, including critical high-cost and long-term decisions (e.g., new infrastructure). In overall, the exploration of novel data sources and their combination with traditional sources for generating information of higher quality (e.g., the combination of GPS-based travel diaries with questionnaires in Iogansen et al., 2025) can assist transport stakeholders into a broader guidance towards integrated transport policies. The range of such policies is wide, including congestion pricing, parking management, public transport ticketing, and promoting cycling, among many others.

7.4. Study limitations and future research

While this study provides novel insights into mobility profiling using TravelSense data, several limitations should be acknowledged. First, the temporal scope was restricted to weekdays in April for the years 2022, 2023, and 2024. This choice enabled comparability across years but excludes potential seasonal variation (e.g., winter versus summer travel) as well as differences between weekday and weekend patterns. It is noteworthy that April was selected for our analysis because it generally features moderate weather conditions, enabling us to capture typical mobility patterns without the influence of seasonal extremes. Consequently, the results should be interpreted as representative of typical weekday mobility rather than capturing the variability in travel behaviour across the year.

Second, our study focused primarily on an aggregated analysis of travel profiles. This approach is effective for identifying broad demand patterns but less sensitive to short-term behavioural shifts (e.g., daily variability or differences in trip purposes). Expanding the analysis to disaggregated temporal, spatial, and/or modal resolutions could provide additional insights into how stable or flexible individual travel patterns are over time.

Third, we chose a set of temporal, spatial, and modal variables that allowed for a meaningful interpretation of mobility patterns and served the purposes of the current study. However, the inclusion of different or additional variables could have highlighted alternative aspects of traveler behaviour and potentially led to different profiles. Future work could explore alternative or complementary variables, tailored to support specific policy-shaping and decision-making for ensuring sustainable mobility.

Finally, our analysis did not directly incorporate external explanatory factors such as socio-demographic attributes, land use patterns, or broader economic and policy influences. External factors (e.g., COVID-19 and infrastructure expansion) were considered while interpreting changes in the profiles over years, however, these interpretations should be viewed as indicative rather than causal. A more integrated approach (e.g., linking digital travel diaries to complementary data sources) could strengthen explanatory power in future work.

8. Conclusions

Demand profiling is an invaluable approach to identify and understand the different groups of travelers in a transport network. However, such a challenging task requires a data source that would allow to capture as many mobility aspects as possible within a multi-modal network. In this study, we utilized TravelSense, a smartphone-based data set that resembles digitally collected travel diaries and is available in Helsinki, Finland.

TravelSense is associated with both benefits and challenges that had an impact on our study for demand profiling. The highlights of its benefits refer to information about multi-modal travel patterns at trajectory-level through a single data source. Such capabilities allow its usage in a wide range of studies and applications, including crowding mitigation, pricing strategies, infrastructure asset management, and others. Among the challenges it poses, we need to highlight the inability to monitor a single traveler over multiple days. It is noted that this challenge is related with high standards of travelers' privacy protection, emphasizing on the trade-off between detailed traveler information and traveler anonymity protection. Another potential challenge refers to the nature of the data collection itself. Since travelers need to allow to be monitored, concepts such as acceptance and trust towards emerging technologies are also added in the list of considerations that policy makers should take into account.

Through LPA we obtained profiles across three categories (temporal, spatial, and modal) and three years (2022, 2023, and 2024), giving a holistic overview of human mobility within Helsinki region, the relationships among profiles and the changes over the years. Although results indicate a high robustness of human mobility from year to year, they also indicate a different mobility behaviour during 2022, most likely affected by the COVID-19 pandemic. Several of our key findings are in alignment with existing studies that refer to completely different locations around the world (e.g., China, Sweden, Canada, France) and are based on a different data source (i.e., smart cards). This underscores the similarities of travel patterns resulting from common habits and preferences that people share irrespective of the specifics of geographical locations.

Future work may investigate how mobility patterns potentially vary across seasons of the year (e.g., winter versus summer). Similarly, a phase-based analysis could be conducted by identifying important changes in the network (e.g., tramline extension) and study the changes in profiles, if any, before and after. In addition, digitally collected travel diaries allow the quantification of a variety of variables through which human mobility could be approached. Future work will focus on identifying alternative variables that will allow to shift the focus of demand profiling to a different direction according to a specific goal (e.g., shaping ticket pricing policies, integrating new modes in the network, changing service design), always aiming at supporting the ongoing transformation towards a sustainable urban mobility system.

CRediT authorship contribution statement

Charalampos Sipetas: Writing – original draft, Methodology, Formal analysis, Data curation, Conceptualization; **Nejc Geržinič:** Writing – original draft, Methodology, Formal analysis; **Zhiren Huang:** Writing – original draft, Methodology, Formal analysis, Data curation, Conceptualization; **Oded Cats:** Writing – review & editing, Validation, Supervision, Formal analysis; **Miloš N. Mladenović:** Writing – review & editing, Validation, Supervision, Conceptualization.

Data availability

The authors do not have permission to share data.

Declaration of competing interest

The authors declare that they have no conflict of interest.

Acknowledgments

The authors thank HSL for access to the TravelSense dataset, and for their time and discussions about this study. The work of C. Sipetas was supported by the Research Council of Finland project ALCOSTO (no. 349327) and the EU Horizon Europe project ACUMEN (no. 101103808). Z. Huang is supported by the NetResilience consortium funded by the Strategic Research Council at the Academy of Finland (grant numbers 345188 and 345183) and Guangdong Science and Technology Strategic Innovation Fund (the Guangdong-Hong Kong-Macau Joint Laboratory Program), Project No.: 2020B1212030009. Calculations were performed using computer resources within the Aalto University School of Science "Science-IT" project.

Appendix A. Temporal profiles for April 2023 and April 2024

The obtained temporal profiles for April 2023 (Figs. A.17 and A.18) are the following:

- **Profile 1 - Morning peak travelers with moderate travel times (30 %):** This profile experiences moderate daily travel times. They make very few trips between midnight and 06.00 and, along with profile 4, record the highest number of trips during the morning peak (06.00 - 10.00). However, they do not travel at all between 18.00 and midnight. This profile shares similarities with activity profile 1 from April 2022.

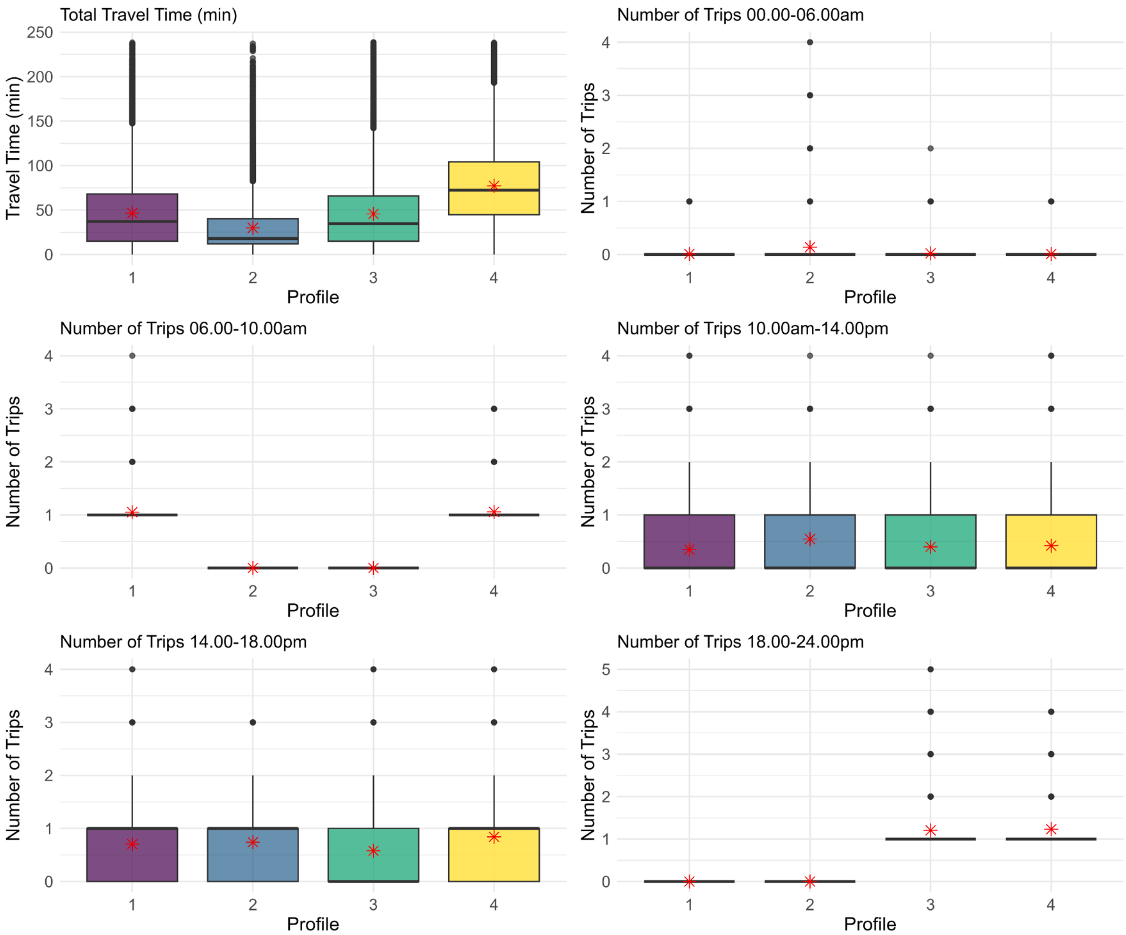

Fig. A.17. Boxplots of temporal profiles for April 2023.

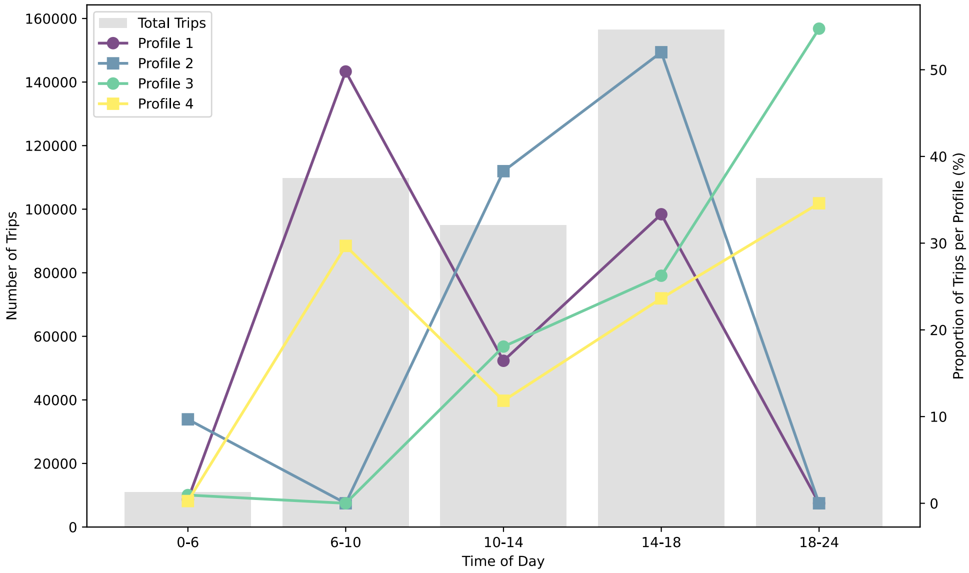

Fig. A.18. Histogram of number of trips per traveler and time period for April 2023.

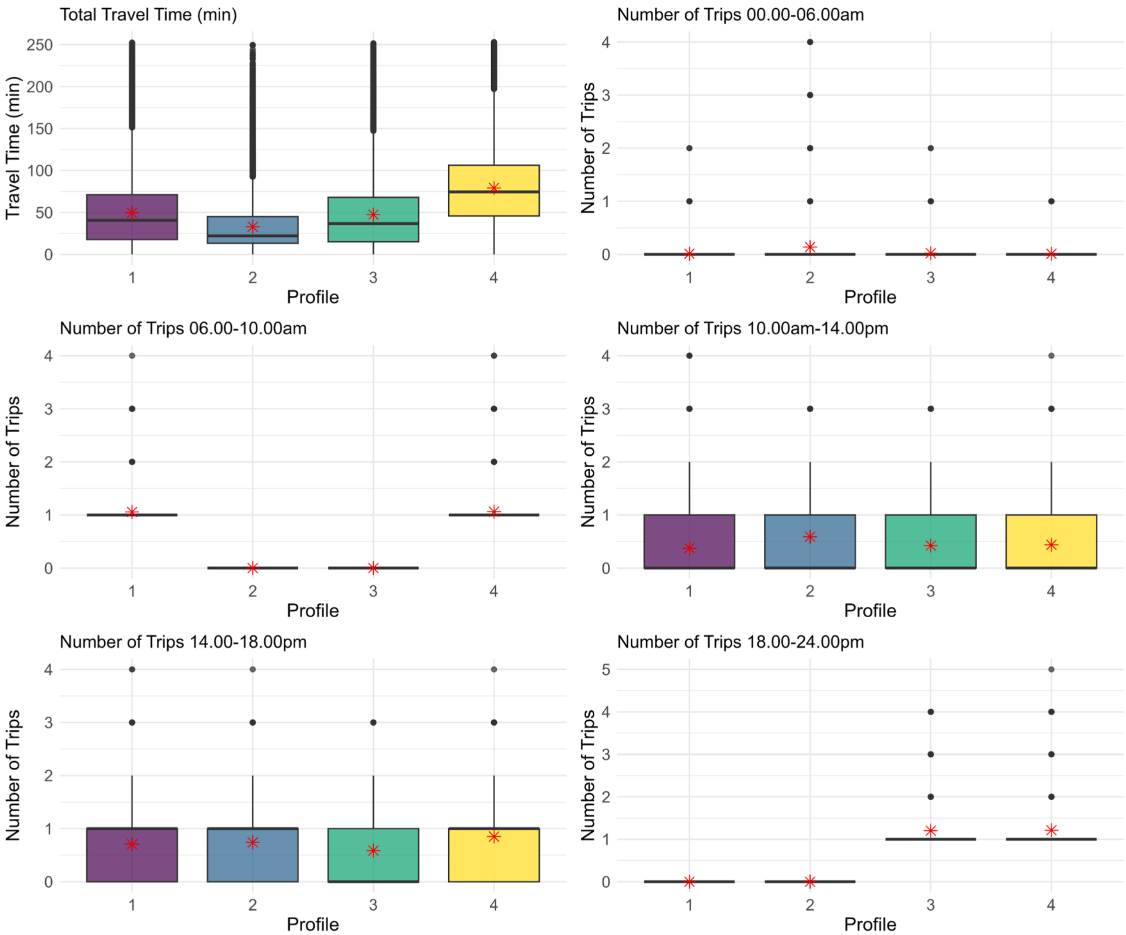

Fig. A.19. Boxplots of temporal profiles for April 2024.

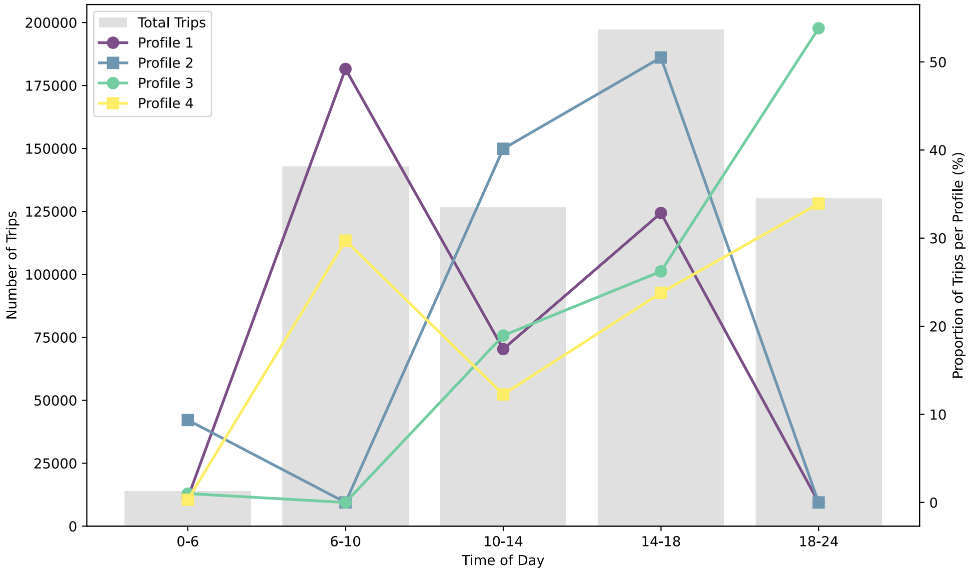

Fig. A.20. Histogram of number of trips per traveler and time period for April 2024.

- **Profile 2 - After-midnight travelers with short travel times (29 %):** These travelers have the shortest total travel times. Their trips occur exclusively between midnight - 06.00 and 10.00 - 18.00. They exhibit the highest number of trips among all profiles during the early morning hours (midnight - 06.00). Additionally, they have the highest trip frequency between 10.00 and 14.00, although the differences from other profiles are relatively small.
- **Profile 3 - Late evening travelers with moderate travel times (24 %):** This group has moderate travel times but does not travel between 06.00 and 10.00. They make a few trips between midnight and 06.00 and, along with profile 4, record the highest number of trips during the evening period (18.00 - midnight).
- **Profile 4 - All day travelers with long travel times (17 %):** These travelers have the longest travel times among all profiles, making trips throughout the entire day. They exhibit the highest trip frequency during the morning peak (06.00 - 10.00), similar to profile 1. Additionally, they record the most trips during the evening period (18.00 - midnight), alongside profile 3. This profile closely resembles temporal profile 5 from April 2022.

The obtained temporal profiles for April 2024 (Figs. A.19 and A.20) are the following:

- **Profile 1 - Morning peak travelers with moderate travel times (32 %):** This profile has the same characteristics as profile 1 of April 2023.
- **Profile 2 - After midnight travelers with short travel times (29 %):** This profile has the same characteristics as profile 2 of April 2023.
- **Profile 3 - Late evening travelers with moderate travel times (22 %):** This profile has the same characteristics as profile 3 of April 2023.
- **Profile 4 - All day travelers with long travel times (17 %):** This profile has the same characteristics as profile 4 of April 2023.

Appendix B. Spatial profiles for April 2023 and April 2024

The obtained spatial profiles for April 2023 (Figs. B.21 and B.22) are described as follows:

- **Profile 1 - Long distance travelers with frequent core zone trips (45 %):** This profile is the same as profile 1 of April 2022.
- **Profile 2 - Short distance travelers with single core zone trip (31 %):** This profile is the same as profile 2 of April 2022.
- **Profile 3 - Long distance travelers beyond the core zone (24 %):** This profile is the same as profile 3 of April 2022.

The obtained spatial profiles for April 2024 (Figs. B.23 and B.23) are described as follows:

- **Profile 1 - Long distance travelers with frequent core zone trips (46 %):** This profile is the same as profile 1 of April 2022 and profile 1 of April 2023.
- **Profile 2 - Short distance travelers with single core zone trip (29 %):** This profile is the same as profile 2 of April 2022 and profile 2 of April 2023.
- **Profile 3 - Long distance travelers beyond the core zone (25 %):** This profile is the same as profile 3 of April 2022 and profile 3 of April 2023.

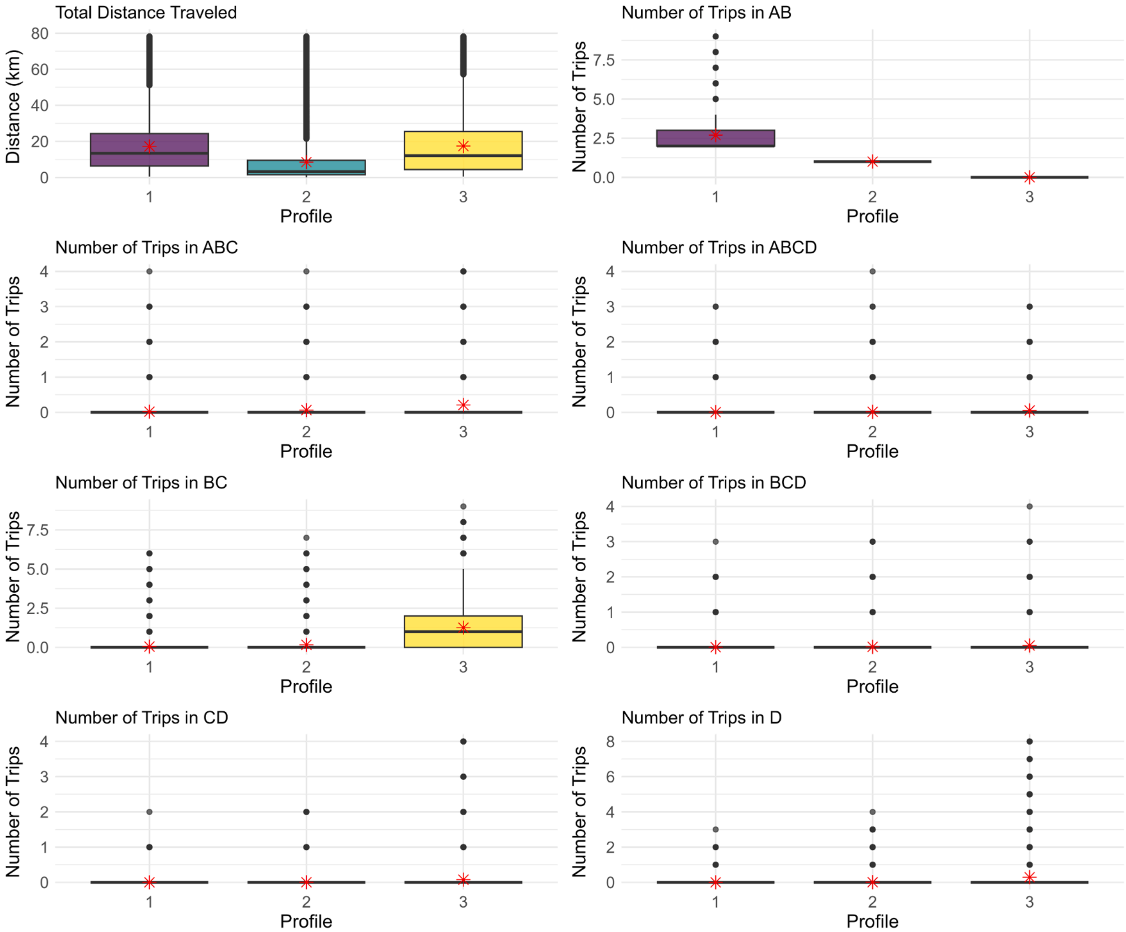

Fig. B.21. Boxplots of spatial profiles for April 2023.

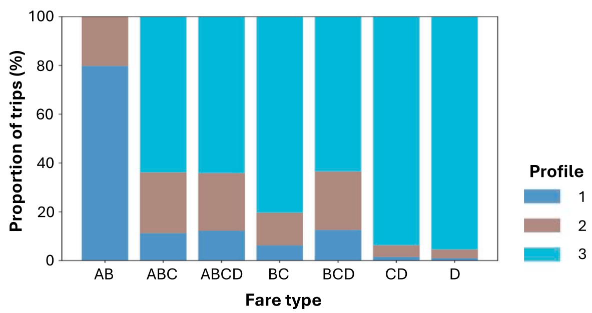

Fig. B.22. Stacked bar chart with proportion of spatial profiles per fare type for April 2023.

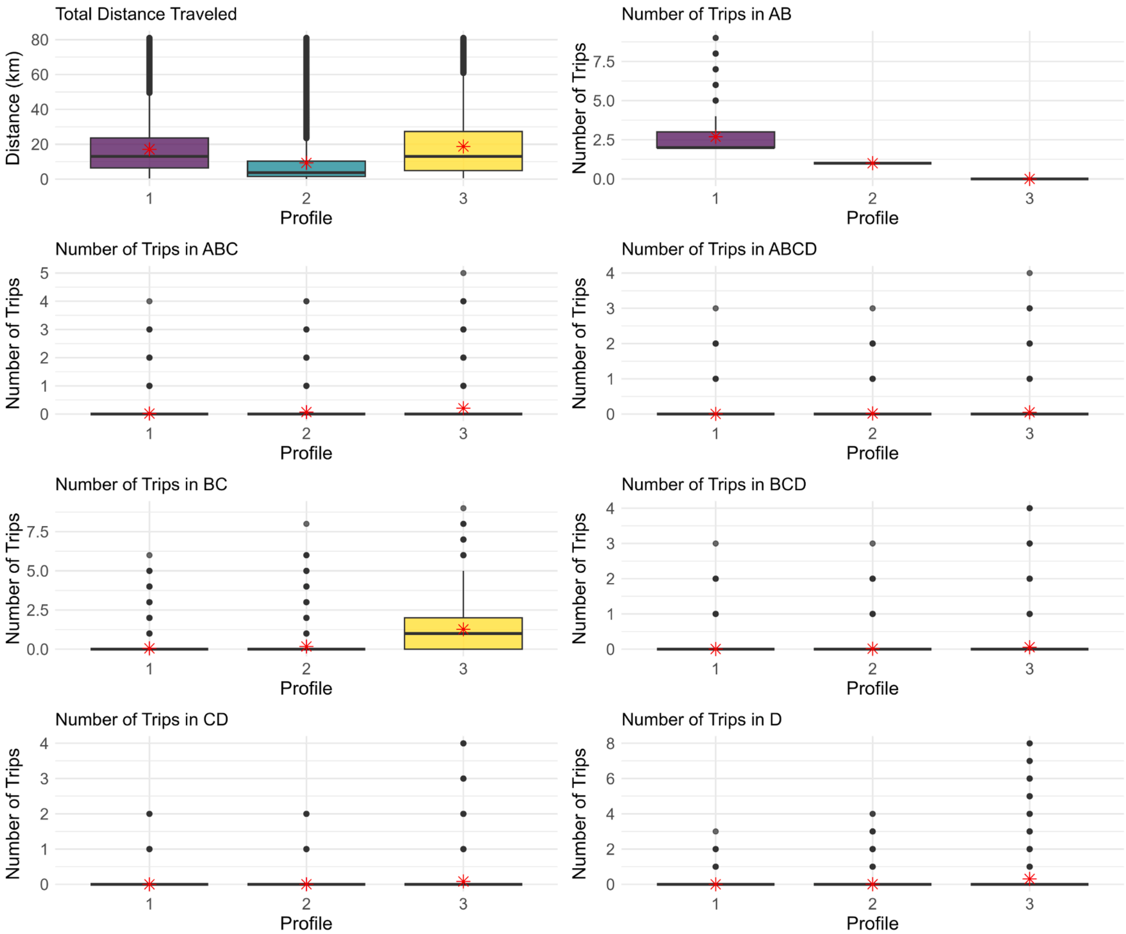

Fig. B.23. Boxplots of spatial profiles for April 2024.

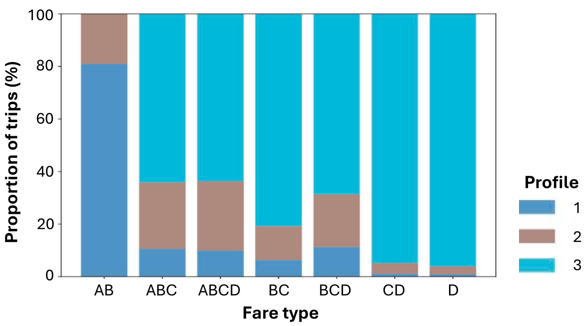

Fig. B.24. Stacked bar chart with proportion of spatial profiles per fare type for April 2024.

Appendix C. Modal profiles for April 2023 and April 2024

The obtained modal profiles for April 2023 (Fig. C.25) are mostly related to those of April 2022 and can be described as follows:

- **Profile 1 - Highly multi-modal travelers using all modes (40 %):** This profile is the same as profile 2 of April 2022.
- **Profile 2 - Low multi-modality bus travelers (26 %):** These travelers present low multi-modality per trip chain, since they only choose active modes and bus for their daily commute. Their distinctive characteristic compared to the other profiles is that they present the highest usage of bus mode.
- **Profile 3 - High multi-modality travelers without private cars (20 %):** This profile is the same as profile 1 of April 2022.
- **Profile 4 - Active mode travelers (14 %):** This profile is the same as profile 3 of April 2022.

The obtained modal profiles for April 2024 (Fig. C.26) are related to those of April 2022 and April 2023, as follows:

- **Profile 1 - High multi-modality travelers using all modes (45 %):** This profile is the same as profile 2 of April 2022 and profile 1 of April 2023.
- **Profile 2 - High multi-modality travelers without private cars (22 %):** This profile is the same as profile 1 of April 2022 and profile 3 of April 2023.
- **Profile 3 - Low multi-modality bus travelers (20 %):** This profile is the same as profile 2 of April 2023.
- **Profile 4 - Active mode travelers (13 %):** This profile is the same as profile 3 of April 2022 and profile 4 of April 2023.

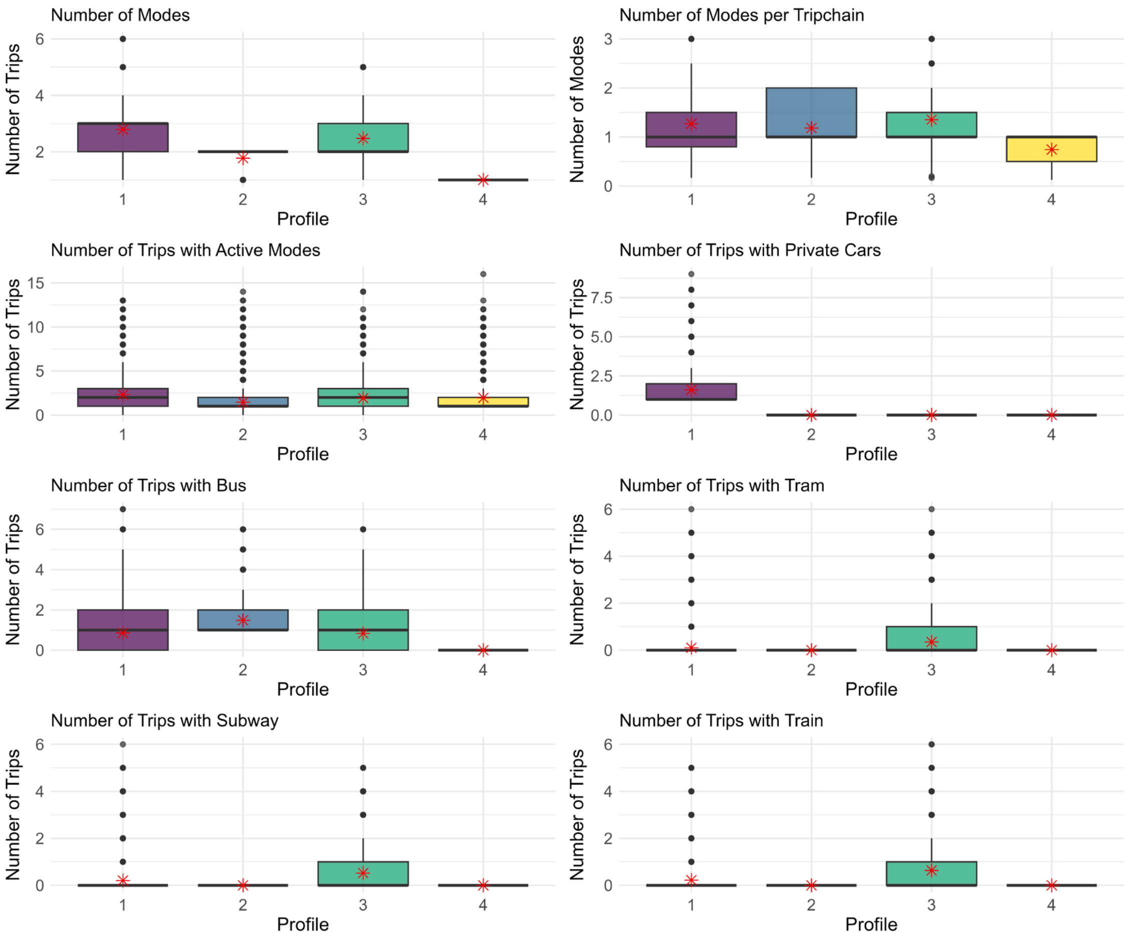

Fig. C.25. Boxplots of modal profiles for April 2023.

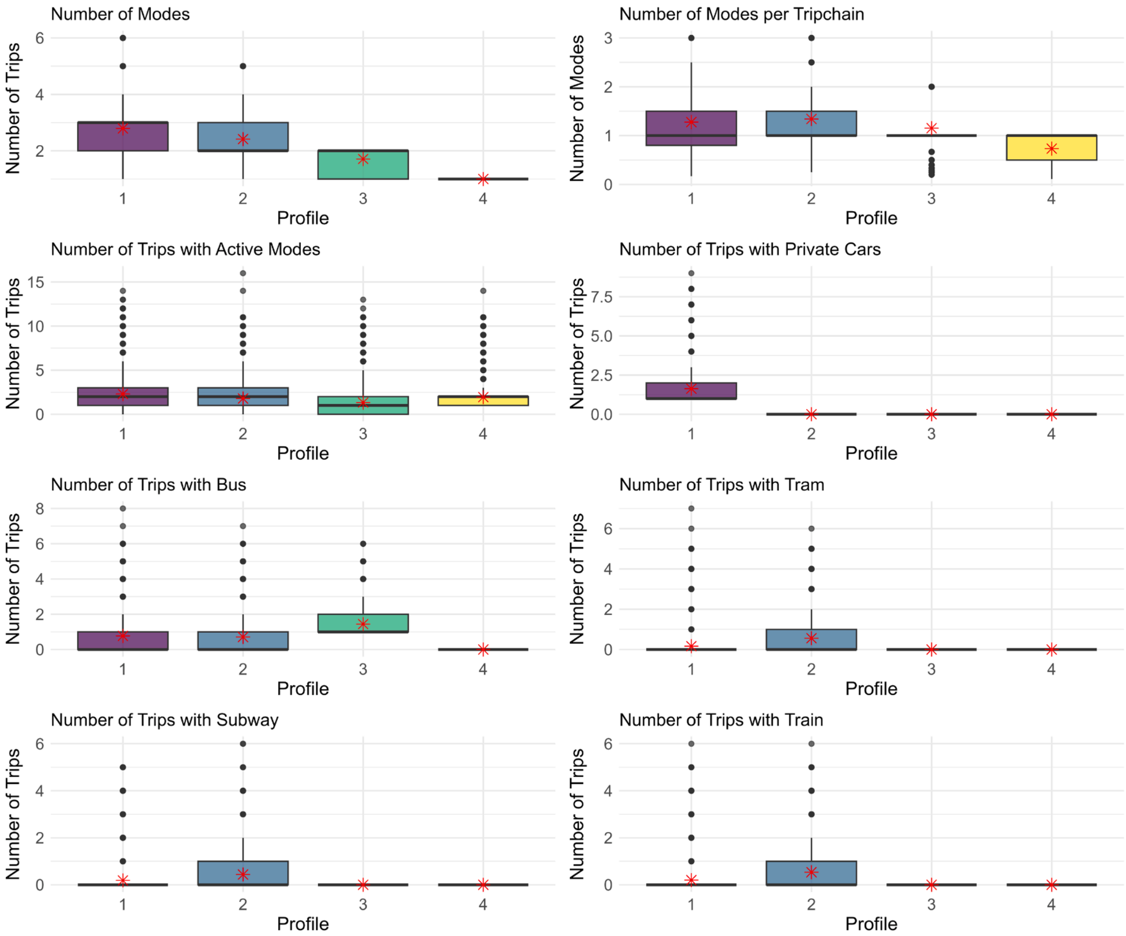

Fig. C.26. Boxplots of modal profiles for April 2024.

References

Agard, B., Morency, C., Trépanier, M., 2006. Mining public transport user behaviour from smart card data. IFAC Proc. Vol. 39 (3), 399–404. https://doi.org/10.3182/20060517-3-FR-2903.00211.

Akaike, H., 1987. Factor analysis and AIC. Psychometrika 52, 317–332. https://doi.org/10.1007/BF02294359.

An, Z., Heinen, E., Watling, D., 2022. Multimodal travel behaviour, attitudes, and cognitive dissonance. Transp. Res. Part F 91, 260–273. https://doi.org/10.1016/j.trf.2022.10.007.

Bachir, D., Khodabandelou, G., Gauthier, V., El Yacoubi, M., Puchinger, J., 2019. Inferring dynamic origin-destination flows by transport mode using mobile phone data. Transp. Res. Part C 101, 254–275. https://doi.org/10.1016/j.trc.2019.02.013.

Banister, D., 2008. The sustainable mobility paradigm. Transp. Policy 15 (2), 73–80. https://doi.org/10.1016/j.tranpol.2007.10.005.

Barry, J. J., Freimer, R., Slavin, H., 2009. Use of entry-only automatic fare collection data to estimate linked transit trips in New Nork City. Transp. Res. Rec. 2112 (1), 53–61. https://doi.org/10.3141/2112-07.

Briand, A.-S., Côme, E., El Mahrsi, M. K., Oukhellou, L., 2016. A mixture model clustering approach for temporal passenger pattern characterization in public transport. Int. J. Data Sci. Anal. 1 (1), 37–50. https://doi.org/10.1007/s41060-015-0002-x.

Briand, A.-S., Côme, E., Trépanier, M., Oukhellou, L., 2017. Analyzing year-to-year changes in public transport passenger behaviour using smart card data. Transp. Res. Part C 79, 274–289. https://doi.org/10.1016/j.trc.2017.03.021.

Buehler, R., Hamre, A., 2015. The multimodal majority? Driving, walking, cycling, and public transportation use among American adults. Transportation 42, 1081–1101. https://doi.org/10.1007/s11116-014-9556-z.

Cats, O., 2024. Identifying human mobility patterns using smart card data. Transp. Rev. 44 (1), 213–243. https://doi.org/10.1080/01441647.2023.2251688.

Cats, O., Ferranti, F., 2022. Unravelling individual mobility temporal patterns using longitudinal smart card data. Res. Transp. Bus. Manag. 43, 100816. https://doi.org/10.1016/j.rtbm.2022.100816.

Celeux, G., Soromenho, G., 1996. An entropy criterion for assessing the number of clusters in a mixture model. J. Classif. 13, 195–212. https://doi.org/10.1007/BF01246098.

Cheng, L., Cai, X., Liu, Z., Huang, Z., Chen, W., Witlox, F., 2024. Characterising travel behaviour patterns of transport hub station area users using mobile phone data. J. Transp. Geogr. 116, 103855. https://doi.org/10.1016/j.jtrangeo.2024.103855.

CIT UPC, 2023. Mobilitapp: deep learning to predict the transport mode used by citizens and protect their data securely. https://cit.upc.edu/en/portfolio-item/mobilitapp-deep-learning-to-predict-the-transport-mode-used-by-citizens-and-protect-their-data-securely/.Accessed: 2025-03-17.

Clifton, K., Muhs, C. D., 2012. Capturing and representing multimodal trips in travel surveys: review of the practice. Transp. Res. Rec. 2285 (1), 74–83. https://doi.org/10.3141/2285-09.

Diana, M., 2010. From mode choice to modal diversion: a new behavioural paradigm and an application to the study of the demand for innovative transport services. Technol. Forecast. Soc. Change 77 (3), 429–441. https://doi.org/10.1016/j.techfore.2009.10.005.

Dixit, M., Cats, O., van Oort, N., Brands, T., Hoogendoorn, S., 2024. Validation of a multi-modal transit route choice model using smartcard data. Transportation 51 (5), 1809–1829. https://doi.org/10.1007/s11116-023-10387-z.

Du, R., He, D., Song, Y., Ding, Z., Chan, S., Li, X., 2025. Privacy-preserving short-term travel demand forecasting based on federated learning. IEEE Trans. Veh. Technol. 74 (7), 11436–11449. https://doi.org/10.1109/TVT.2025.3547823.

El Mahrsi, M. K., Côme, E., Oukhellou, L., Verleysen, M., 2017. Clustering smart card data for urban mobility analysis. IEEE Trans. Intell. Transp. Syst. 18 (3), 712–728. https://doi.org/10.1109/TITS.2016.2600515.

Fu, X., van Lierop, D., Ettema, D., 2024. Is multimodality advantageous? Assessing the relationship between multimodality and perceived transport adequacy and accessibility in different travel contexts. Transp. Res. Part A 179, 103893. https://doi.org/10.1016/j.tra.2023.103893.

Gonzales, E. J., Keklikoglou, A., Sipetas, C., et al., 2018. Measuring Left-Behinds on Subway. Technical Report. Massachusetts Department of Transportation. https://rosap.ntl.bts.gov/view/dot/64851.

Gordon, J. B., Koutsopoulos, H. N., Wilson, N. H. M., Attanucci, J. P., 2013. Automated inference of linked transit journeys in london using fare-transaction and vehicle location data. Transp. Res. Rec. 2343 (1), 17–24. https://doi.org/10.3141/2343-03.

Graells-Garrido, E., Caro, D., Parra, D., 2018. Inferring modes of transportation using mobile phone data. EPJ Data Sci. 7 (1), 1–23. https://doi.org/10.1140/epjds/s13688-018-0177-1.

Graells-Garrido, E., Opitz, D., Rowe, F., Arriagada, J., 2023. A data fusion approach with mobile phone data for updating travel survey-based mode split estimates. Transp. Res. Part C 155, 104285. https://doi.org/10.1016/j.trc.2023.104285.

Gutiérrez, A., Domènech, A., Zaragozí, B., Miravet, D., 2020. Profiling tourists' use of public transport through smart travel card data. J. Transp. Geogr. 88, 102820. https://doi.org/10.1016/j.jtrangeo.2020.102820.

He, L., Trépanier, M., Agard, B., 2021. Space–time classification of public transit smart card users' activity locations from smart card data. Public Transp. 13 (3), 579–595. https://doi.org/10.1007/s12469-021-00274-0.

Heinen, E., Chatterjee, K., 2015. The same mode again? An exploration of mode choice variability in great britain using the national travel survey. Transp. Res. Part A 78, 266–282. https://doi.org/10.1016/j.tra.2015.05.015.

Heinen, E., Ogilvie, D., 2016. Variability in baseline travel behaviour as a predictor of changes in commuting by active travel, car and public transport: a natural experimental study. J. Transp. Health 3 (1), 77–85. https://doi.org/10.1016/j.jth.2015.11.002.

Houston, D., Luong, T. T., Boarnet, M. G., 2014. Tracking daily travel; assessing discrepancies between GPS-derived and self-reported travel patterns. Transp. Res. Part C 48, 97–108. https://doi.org/10.1016/j.trc.2014.08.013.

Huang, Y., Ma, L., De Vos, J., 2024. Travel behaviour and multimodality: a conceptual framework and critical review of research. Transp. Rev. 44 (3), 709–730. https://doi.org/10.1080/01441647.2023.2300088.

Huang, Z., Ling, X., Wang, P., Zhang, F., Mao, Y., Lin, T., Wang, F.-Y., 2018. Modeling real-time human mobility based on mobile phone and transportation data fusion. Transp. Res. Part C 96, 251–269. https://doi.org/10.1016/j.trc.2018.09.016.

Huang, Z., Sipetas, C., de Villafranca, A. E. M., Quach, T., 2023a. Identifying shifts in multi-modal travel patterns during special events using mobile data: celebrating Vappu in Helsinki. https://doi.org/10.48550/arXiv.2305.17925.

Huang, Z., de Villafranca, A. E. M., Sipetas, C., 2022. Sensing multi-modal mobility patterns: a case study of helsinki using bluetooth beacons and a mobile application. In: 2022 IEEE International Conference on Big Data (Big Data). IEEE, pp. 2007–2016. https://doi.org/10.1109/BigData55660.2022.10020578.

Huang, Z., de Villafranca, A. E. M., Sipetas, C., Quach, T., 2023b. Crowd-sensing commuting patterns using multi-source wireless data: a case of Helsinki commuter trains. https://doi.org/10.48550/arXiv.2302.02661.

Huang, Z., Wang, P., Liu, Y., 2020. Statistical characteristics and transportation mode identification of individual trajectories. Int. J. Mod. Phys. B 34 (10), 2050092. https://doi.org/10.1142/S0217979220500927.

Iogansen, X., Lee, Y., Young, M., Compostella, J., Circella, G., Jenn, A., 2025. Ridehailing use, travel patterns and multimodality: a latent-class cluster analysis of one-week GPS-based travel diaries in California. Travel Behav. Soc. 38, 100855. https://doi.org/10.1016/j.tbs.2024.100855.

Iqbal, M. S., Choudhury, C. F., Wang, P., González, M. C., 2014. Development of origin–destination matrices using mobile phone call data. Transp. Res. Part C 40, 63–74. https://doi.org/10.1016/j.trc.2014.01.002.

Jee, H., Sun, W., Schmöcker, J.-D., Nakamura, T., 2023. Demonstrating the feasibility of using Wi-Fi sensors for dynamic bus-stop queue length estimation. Public Transp., 1–18.https://doi.org/10.1007/s12469-023-00336-5.

Jiang, S., Ferreira, J., Gonzalez, M. C., 2017. Activity-based human mobility patterns inferred from mobile phone data: a case study of Singapore. IEEE Trans. Big Data 3 (2), 208–219. https://doi.org/10.1109/TBDATA.2016.2631141.

Kim, J., Kwan, M.-P., 2021. The impact of the COVID-19 pandemic on people's mobility: a longitudinal study of the US from March to September of 2020. J. Transp. Geogr. 93, 103039. https://doi.org/10.1016/j.jtrangeo.2021.103039.

Kostakos, V., Camacho, T., Mantero, C., 2013. Towards proximity-based passenger sensing on public transport buses. Pers. Ubiquitous Comput. 17, 1807–1816. https://doi.org/10.1007/s00779-013-0652-4.

Kroesen, M., van Cranenburgh, S., 2016. Revealing transition patterns between mono-and multimodal travel patterns over time: a mover-stayer model. Eur. J. Transp. Infrastruct. Res. 16 (4). https://doi.org/10.18757/ejtir.2016.16.4.3169.

Krueger, R., Bierlaire, M., Bansal, P., 2023. A data fusion approach for ride-sourcing demand estimation: a discrete choice model with sampling and endogeneity corrections. Transp. Res. Part C 152, 104180. https://doi.org/10.1016/j.trc.2023.104180.

Kuhnimhof, T., Buehler, R., Wirtz, M., Kalinowska, D., 2012. Travel trends among young adults in Germany: increasing multimodality and declining car use for men. J. Transp. Geogr. 24, 443–450. https://doi.org/10.1016/j.jtrangeo.2012.04.018.

Kuhnimhof, T., Chlond, B., Von Der Ruhren, S., 2006. Users of transport modes and multimodal travel behavior: steps toward understanding travelers' options and choices. Transp. Res. Rec. 1985 (1), 40–48. https://doi.org/10.1177/0361198106198500105.

Kusakabe, T., Asakura, Y., 2014. Behavioural data mining of transit smart card data: a data fusion approach. Transp. Res. Part C 46, 179–191. https://doi.org/10.1016/j.trc.2014.05.012.

Lee, I., Cho, S.-H., Kim, K., Kho, S.-Y., Kim, D.-K., 2022. Travel pattern-based bus trip origin-destination estimation using smart card data. PLoS One 17 (6), e0270346. https://doi.org/10.1371/journal.pone.0270346.

Lin, M., Huang, Z., Zhao, T., Zhang, Y., Wei, H., 2022. Spatiotemporal evolution of travel pattern using smart card data. Sustainability 14 (15), 9564. https://doi.org/10.3390/su14159564.

Liu, J., Ma, Z., Wen, L., Zhang, N., Qian, Z., 2025. Urban rail transit fare reconciliation method using multi-source data. Transp. Res. Rec., 03611981241310140.https://doi.org/10.1177/03611981241310140.

Manley, E., Zhong, C., Batty, M., 2018. Spatiotemporal variation in travel regularity through transit user profiling. Transportation 45, 703–732. https://doi.org/10.1007/s11116-016-9747-x.

McBride, E. C., Davis, A. W., Goulias, K. G., 2018. A spatial latent profile analysis to classify land uses for population synthesis methods in travel demand forecasting. Transp. Res. Rec. 2672 (49), 158–170. https://doi.org/10.1177/0361198118799168.

McLachlan, G. J., Lee, S. X., Rathnayake, S. I., 2019. Finite mixture models. Annu. Rev. Stat. Appl. 6, 355–378. https://doi.org/10.1146/annurev-statistics-031017-100325.

Molin, E., Mokhtarian, P., Kroesen, M., 2016. Multimodal travel groups and attitudes: a latent class cluster analysis of Dutch travelers. Transp. Res. Part A 83, 14–29. https://doi.org/10.1016/j.tra.2015.11.001.

Munizaga, M. A., Palma, C., 2012. Estimation of a disaggregate multimodal public transport origin–destination matrix from passive smartcard data from Santiago, Chile. Transp. Res. Part C 24, 9–18. https://doi.org/10.1016/j.trc.2012.01.007.

Nitti, M., Pinna, F., Pintor, L., Pilloni, V., Barabino, B., 2020. iABACUS: a Wi-Fi-based automatic bus passenger counting system. Energies 13 (6), 1446. https://doi.org/10.3390/en13061446.

Nobis, C., 2007. Multimodality: facets and causes of sustainable mobility behavior. Transp. Res. Rec. 2010 (1), 35–44. https://doi.org/10.3141/2010-05.

Okoye, K., Hosseini, S., 2024. Correlation tests in r: pearson cor, kendall's tau, and spearman's rho. In: R Programming: Statistical Data Analysis in Research. Springer, pp. 247–277. https://doi.org/10.1007/978-981-97-3385-9_12.

Olafsson, A. S., Nielsen, T. S., Carstensen, T. A., 2016. Cycling in multimodal transport behaviours: exploring modality styles in the danish population. J. Transp. Geogr. 52, 123–130. https://doi.org/10.1016/j.jtrangeo.2016.03.010.

Pang, Y., Tsubouchi, K., Yabe, T., Sekimoto, Y., 2018. Replicating urban dynamics by generating human-like agents from smartphone GPS data. In: Proceedings of the 26th ACM Sigspatial International Conference on Advances in Geographic Information Systems, pp. 440–443. https://doi.org/10.1145/3274895.3274935.

Papacharalampous, A. E., Cats, O., Lankhaar, J.-W., Daamen, W., Van Lint, H., 2016. Multimodal data fusion for big events. Transp. Res. Rec. 2594 (1), 118–126. https://doi.org/10.3141/2594-15.

Provost, F., Fawcett, T., 2013. Data science and its relationship to big data and data-driven decision making. Big Data 1 (1), 51–59. https://doi.org/10.1089/big.2013.1508.

Rogers, E. M., 1962. Diffusion of innovations. Simon and Schuster. https://doi.org/10.4324/9780203887011.

Roncoli, C., Chandakas, E., Kaparias, I., 2023. Estimating on-board passenger comfort in public transport vehicles using incomplete automatic passenger counting data. Transp. Res. Part C 146, 103963. https://doi.org/10.1016/j.trc.2022.103963.

Rosenberg, J. M., Schoemann, A. M., Spaeth, M., 2018. tidyLPA: Easily Carry Out Latent Profile Analysis. R package version 1.0.0. https://CRAN.R-project.org/package=tidyLPA.

Sánchez-Martínez, G. E., 2017. Inference of public transportation trip destinations by using fare transaction and vehicle location data: dynamic programming approach. Transp. Res. Rec. 2652 (1), 1–7. https://doi.org/10.3141/2652-01.

Schwarz, G., 1978. Estimating the dimension of a model. Ann. Stat., 461–464. https://www.jstor.org/stable/2958889.

Seaborn, C., Attanucci, J., Wilson, N. H. M., 2009. Analyzing multimodal public transport journeys in london with smart card fare payment data. Transp. Res. Rec. 2121 (1), 55–62. https://doi.org/10.3141/2121-06.

Sipetas, C., Huang, Z., Espinosa Mireles de Villafranca, A., 2024. Evaluation framework for multi-modal public transport systems based on connectivity and transfers at stop level. Transp. Res. Rec. 2678 (10), 147–162. https://doi.org/10.1177/03611981241231963.

Sipetas, C., Keklikoglou, A., Gonzales, E. J., 2020. Estimation of left behind subway passengers through archived data and video image processing. Transp. Res. Part C 118, 102727. https://doi.org/10.1016/j.trc.2020.102727.

Spurk, D., Hirschi, A., Wang, M., Valero, D., Kauffeld, S., 2020. Latent profile analysis: a review and "how to" guide of its application within vocational behavior research. J. Vocat. Behav. 120, 103445. https://doi.org/10.1016/j.jvb.2020.103445.

Stopher, P., Shen, L., 2011. In-depth comparison of global positioning system and diary records. Transp. Res. Rec. 2246 (1), 32–37. https://doi.org/10.3141/2246-05.

Stopher, P. R., Greaves, S. P., 2007. Household travel surveys: where are we going? Transp. Res. Part A 41 (5), 367–381. https://doi.org/10.1016/j.tra.2006.09.005.

Ton, D., Zomer, L.-B., Schneider, F., Hoogendoorn-Lanser, S., Duives, D., Cats, O., Hoogendoorn, S., 2020. Latent classes of daily mobility patterns: the relationship with attitudes towards modes. Transportation 47, 1843–1866. https://doi.org/10.1007/s11116-019-09975-9.

Toole, J. L., Colak, S., Sturt, B., Alexander, L. P., Evsukoff, A., González, M. C., 2015. The path most traveled: travel demand estimation using big data resources. Transp. Res. Part C 58, 162–177. https://doi.org/10.1016/j.trc.2015.04.022.

Trépanier, M., Tranchant, N., Chapleau, R., 2007. Individual trip destination estimation in a transit smart card automated fare collection system. J. Intell. Transp. Syst. 11 (1), 1–14. https://doi.org/10.1080/15472450601122256.

Tu, L., Wang, S., Zhang, D., Zhang, F., He, T., 2019. ViFi-MobiScanner: observe human mobility via vehicular internet service. IEEE Trans. Intell. Transp. Syst. 22 (1), 280–292. https://doi.org/10.1109/TITS.2019.2956744.

Varga, B., Ormándi, T., Tettamanti, T., Aba, A., Esztergár-Kiss, D., 2025. Strategic trip planning for commuting considering user preferences of employees in organizations. Cities 165, 106097. https://doi.org/10.1016/j.cities.2025.106097.

Espinosa Mireles de Villafranca, A., Huang, Z., Sipetas, C., 2024. Using temporal public transport demand profiles to reveal urban spatial patterns. In: Proceedings of the 12th Symposium of the European Association for Research in Transportation (hEART). Espoo, Finland.

Wang, F., Chen, C., 2018. On data processing required to derive mobility patterns from passively-generated mobile phone data. Transp. Res. Part C 87, 58–74. https://doi.org/10.1016/j.trc.2017.12.003.

Wang, P., Huang, Z., Lai, J., Zheng, Z., Liu, Y., Lin, T., 2021. Traffic speed estimation based on multi-source GPS data and mixture model. IEEE Trans. Intell. Transp. Syst. 23 (8), 10708–10720. https://doi.org/10.1109/TITS.2021.3095408.

Wang, X., Yan, X., Zhao, X., Cao, Z., 2022. Identifying latent shared mobility preference segments in low-income communities: ride-hailing, fixed-route bus, and mobility-on-demand transit. Travel Behav. Soc. 26, 134–142. https://doi.org/10.1016/j.tbs.2021.09.011.

Welch, T. F., Widita, A., 2019. Big data in public transportation: a review of sources and methods. Transp. Rev. 39 (6), 795–818. https://doi.org/10.1080/01441647.2019.1616849.

Wickham, H., 2014. Tidy data. J. Stat. Softw. 59 (10), 1–23. https://doi.org/10.18637/jss.v059.i10.

Zhao, J., Rahbee, A., Wilson, N. H. M., 2007. Estimating a rail passenger trip origin-destination matrix using automatic data collection systems. Computer-Aided Civ. Infrastruct. Eng. 22 (5), 376–387. https://doi.org/10.1111/j.1467-8667.2007.00494.x.

Zhao, J., Zhang, F., Tu, L., Xu, C., Shen, D., Tian, C., Li, X.-Y., Li, Z., 2016. Estimation of passenger route choice pattern using smart card data for complex metro systems. IEEE Trans. Intell. Transp. Syst. 18 (4), 790–801. https://doi.org/10.1109/TITS.2016.2587864.

Zheng, X., Chen, W., Wang, P., Shen, D., Chen, S., Wang, X., Zhang, Q., Yang, L., 2015. Big data for social transportation. IEEE Trans. Intell. Transp. Syst. 17 (3), 620–630. https://doi.org/10.1109/TITS.2015.2480157.

Zheng, Y., 2011. User Guide - GeoLife Version. Microsoft Research Asia. https://www.microsoft.com/en-us/research/publication/geolife-gps-trajectory-dataset-user-guide/.

Zheng, Y., Li, Q., Chen, Y., Xie, X., Ma, W.-Y., 2008. Understanding mobility based on GPS data. In: Proceedings of the 10th International Conference on Ubiquitous Computing, pp. 312–321. https://doi.org/10.1145/1409635.1409677.

Zheng, Z., Huang, Z., Zhang, F., Wang, P., 2018. Understanding coupling dynamics of public transportation networks. EPJ Data Sci. 7, 1–16. https://doi.org/10.1140/epjds/s13688-018-0148-6.