Transportation Research Part C 186 (2026) 105607

Contents lists available at ScienceDirect

Transportation Research Part C

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

Real-time on-board passenger comfort estimation in complex public transport networks with intersecting lines

Charalampos Sipetas [a,b,*], Claudio Roncoli [c,a], Ektoras Chandakas [d], Ioannis Kaparias [e]

[a] Department of Built Environment, Aalto University, Espoo, Finland
[b] Division of Engineering, New York University Abu Dhabi, Abu Dhabi, United Arab Emirates
[c] Mobility and Industrial Management (MIM), KU Leuven, Leuven, Belgium
[d] Laboratoire Ville Mobilité Transport, École des Ponts ParisTech, Paris, France
[e] Transportation Research Group, University of Southampton, Southampton, United Kingdom

ARTICLE INFO

Keywords:
Public transport
On-board comfort
Automatic passenger counting
Multi-line public transport network
Kalman filter

ABSTRACT

Comfort on-board public transport vehicles is a critical metric of user experience and service performance. The quantification of this metric requires knowledge of the number of passengers on-board every time a vehicle arrives at or departs from a stop or station. Automatic Passenger Counting (APC) systems allow obtaining such knowledge in real-time, but the information is often incomplete due to system malfunctions, or, more commonly, a lack of the relevant equipment in some vehicles. This study develops an advanced method for passenger estimation that fills gaps in incomplete APC datasets, with computational performance allowing real-time application, and calculates comfort levels on-board public transport vehicles in complex networks where stations are served by multiple lines. The proposed method is tested on a case study considering the Helsinki commuter train network, comprising 6 service lines and 20 stations. The results indicate that the proposed framework can achieve comfort level estimations with high precision across the different cases evaluated. Furthermore, the study provides insight into the key practical question of the number of vehicles that need to be equipped with APC devices in order to obtain sufficiently accurate on-board passenger comfort estimates, and it is shown that it is possible to obtain these estimates even when only a small subset of the runs of any single day are performed by equipped vehicles. Finally, the proposed estimation approach is a valuable tool for operators to obtain a better understanding of daily mobility patterns, evaluate their services through quantifying user experience, and enhance their operations.

1. Introduction

Monitoring the factors that affect user experience in public transport systems is of critical importance to ensure viable and sustainable services. There are a number of determinants of user experience, with comfort on-board vehicles being one of the most significant (Göransson and Andersson, 2023). Comfort is directly affected by the number of passengers on-board and influences the very choice of public transport as a mode for any trip, as other (less sustainable) modes are likely to be preferable if public transport involves travelling on crowded vehicles without seating availability. As a result, authorities and operators utilise on-board comfort as

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

https://doi.org/10.1016/j.trc.2026.105607
Received 28 May 2025; Received in revised form 19 December 2025; Accepted 25 February 2026
Available online 4 March 2026

a well-established measure of service performance in order to enhance user satisfaction and ensure public transport usage. Therefore, information on the number of passengers on-board public transport vehicles is essential.

Technological advances have led to greater availability of automatically collected data for quantifying transport demand. However, there is an increasing need for efficient models that are capable of supporting advanced real-time applications in the field of comfort estimation. This leads to the need to ensure the highest possible quality of available datasets. Therefore, the development of approaches to improve the quality of commonly used datasets is critical to ensure efficient service planning and management for services with high operational performance.

Automatic Passenger Counting (APC) is a well-established technology for collecting demand-related data in public transport networks (Boyle, 1998). The collection of APC data is based on technological equipment installed on the vehicles to count the number of passengers boarding and alighting and, consequently, the number of passengers on-board (e.g., through sensors installed above the vehicle doors) or directly the number of passengers on-board (e.g., through weight sensors) (Dib et al., 2023). Therefore, APC has been used for several studies related to crowding and comfort on-board public transport vehicles (Khomchuk et al., 2018; Hu et al., 2020; Roncoli et al., 2023a).

APC systems offer clear opportunities for demand analysis, but a commonly reported drawback is their high acquisition and maintenance costs. According to Dib et al. (2023), APC systems are optional for network operations and, since they are often regarded by operators as expendable costs, they are typically installed only on a subset of the fleet. While some networks achieve remarkably high APC coverage, e.g., 90% in the Regional Municipality of Waterloo in Canada (Jung and Casello, 2020), coverage can be much lower in many cases, e.g., about 20% on Stockholm's busiest high-frequency bus line in Sweden (Jenelius, 2019a), and only 5–10% in the tramway network of Nantes in France (Roncoli et al., 2023a). Moreover, APC technologies are not error-proof, with system malfunctions being an all-too-common occurrence (Grgurević et al., 2021). Hence, there is a critical need to develop models, methods, and tools that allow filling the gaps of the generated datasets, so that they can be utilised for advanced applications within the network.

This paper builds on the previous work by Roncoli et al. (2023a), which utilises APC in conjunction with Automatic Vehicle Locating (AVL) data in order to estimate passenger loads and comfort levels, with a focus on tramway networks and a case study in the French city of Nantes. Such a method achieves good results even in the case of a low penetration rate of APC devices (i.e., 5-10% as in the case study); however, its main limitation is that it is based on the assumption that stations serve a single line and passengers always board the first vehicle that arrives. However, this is not the case in more complex networks in which stations are served by multiple lines, where passengers are faced with a choice of which vehicle to board rather than board the first one that arrives. Therefore, the present study lifts this constraint, generalising the methodology by Roncoli et al. (2023a) and proposing an estimation method for comfort levels in complex public transport networks, based on non-exhaustive (incomplete) APC datasets. The novel estimation framework introduces updates in the main modelling components of the existing estimation framework, including also the estimation of passengers alighting and historical information.

The case study of the western branch of the Helsinki commuter train network is used here, in which stations are served by up to six lines. APC devices are installed on virtually all vehicles, but technical faults and malfunctions result in missing APC records, which leads to a variable APC coverage per day. The estimation of the number of passengers on-board trains between two consecutive stations and the resulting estimation of comfort levels on-board is achieved per vehicle run, line stop, and time interval of a typical working day. Hence, the proposed model and its outputs can be relevant both for passengers (who may choose between alternative departures at the same station) and for operators (who can monitor service quality and intervene when overcrowding occurs). Beyond real-time applications, such models can serve as tools for assisting decision-making for more long-term planning of the public transport network.

The rest of this paper is structured as follows. Section 2 presents relevant literature and highlights this paper's contribution. Section 3 describes the proposed methodology and the validation framework. Section 4 presents the Helsinki commuter train network and the APC data that are available, including also data analyses to demonstrate operational and demand-related characteristics. Section 5 presents and discusses the results from implementing the proposed methodology. Section 6 presents the results of sensitivity analyses. Finally, Section 7 concludes the study and identifies areas of future work.

2. Literature review

2.1. Crowding and comfort in public transport

Public transport vehicles with high volumes of passengers on-board are seen as overcrowded, leading to negative feelings and perceptions toward public transport. A recent systematic review on crowding valuation in public transport (Fedujwar and Agarwal, 2024) highlighted the need to understand the effect of crowding on public transport systems and their users. The findings indicate that the passenger perception towards crowding differs depending on mode of transport, data types, and different modelling frameworks, among others. The recent COVID-19 pandemic also motivated a number of studies related with crowding in public transport (Basnak et al., 2022; Yap et al., 2025; Karatsoli et al., 2024; Cho et al., 2024).

Shao et al. (2022) emphasise the role of crowding as an external cost for public transport users and focus on its influence on the travel choice behaviour. Their findings suggest that under certain conditions the perceived time on-board crowded vehicles can be twice as long as the actual time. In Tirachini et al. (2013), the authors analyse crowding effects in relation to various factors, such as operating speed, travel time reliability, and passenger well-being, among others. The authors conclude that demand is either over- or under-estimated, if crowding cost is discounted. It is also found that crowding affects not only the passengers who are already on-board, but also the passengers who are waiting to board. Several studies have focused on the phenomenon of passengers being

left behind due to overcrowded vehicles, and have explored what the occurrence of this phenomenon means in terms of mode and route choice (Trozzi et al., 2013, 2015; Zhu et al., 2017; Ma et al., 2019; Sipetas et al., 2020).

Crowding on-board public transport vehicles can be translated into comfort levels, which is an approach for better understanding the passenger experience as a result of the number of passengers with whom they share the same vehicle. Bouscasse and de Lapparent (2019) model comfort as a function of objective attributes and subjective individual perceptions, such as seating availability, perceived use of on-board travel time and ease of public transport usage, among others. Maternini and Cadei (2014) propose and validate a scale of comfort on-board accounting for standing passengers and road characteristics. The relationship between crowding levels, perceived comfort and security is investigated in Tirachini et al. (2017). Results highlight that crowding perception varies across the population, depending on gender, income and age.

Research has also focused on quantifying on-board comfort through composite metrics. For example, Chandakas (2009) assigns comfort levels based on the on-board density of standing passengers and the availability of occupied seats. Definitions of in-vehicle comfort levels are also included in the US Transit Capacity and Quality of Service Manual (TCQSM) (Board, 2013).

2.2. Public transport demand data sources

Estimating crowding and comfort levels on-board public transport vehicles requires information about demand flows within the public transport network, and numerous methods for collecting such information have been developed over the years. Surveys have been a traditional way of collecting demand-related information for various purposes, as for example, to explore rider characteristics (Buehler and Pucher, 2012), traveller attitudes during mode choice (Popuri et al., 2011), and comfort perceptions (da Silva and Mendes, 2020). However, surveys and other manual data collection methods are laborious and expensive (Mohammed and Oke, 2023), while advances in technology allow for more accurate and precise information nowadays. More recent literature includes studies that quantify demand using WiFi and Bluetooth mobile devices (Gonzales et al., 2018; Mehmood et al., 2019; Ryu et al., 2020; Pu et al., 2020). Nevertheless, such data collection technologies have drawbacks, like their tendency towards underestimation due to obstacles or overestimation due to noise from devices that are detected from beyond the area of interest (Bai et al., 2017). More recent research utilises traveller information deriving from public transport ticketing applications for smartphones (Sipetas et al., 2024). According to Huang et al. (2022), however, the quality of such data is variable, as it depends highly on the user acceptance towards being recorded while using the application.

In addition to traditional manual surveys and recent sensing technologies, there are a variety of data sources installed within public transport infrastructure and vehicles to automatically collect demand-related data. Such sources, often collectively referred to as Automatic Data Collection (ADC) systems (Hsu et al., 2020), have made a significant contribution to many studies related with public transport over the last few decades. Smartcard Automated Fare Collection (AFC) is one such technology (Pelletier et al., 2011) that has seen many applications in the last years, such as for example in understanding travel patterns and behaviours (Ma et al., 2013; Munizaga et al., 2014; Briand et al., 2017). AFC data often lack full tap-in/tap-out coverage across networks, which limits accuracy and requires inferences to estimate passenger flows.

APC complements all the above technologies, in what it is directly related with numbers of passengers on-board, and consequently also crowding and comfort. APC devices usually rely on infrared or pressure/weight sensors, which enable counting the numbers of passengers entering and exiting public transport vehicles, and hence, provide good estimates of the number of passengers on-board at any one time. However, APC technology is usually associated with high costs of acquisition and maintenance, which often leads to only a fraction of the vehicles in the public transport network being equipped. Moreover, APC systems, like any other technology, are not error-free and suffer malfunctions, which result in even lower network coverage, thus necessitating the development of advanced methods for filling the data gaps. Several studies have focused on evaluating APC data accuracy and precision (Kimpel et al., 2003; Olivo et al., 2019; Ellenberger and Siebert, 2025) aiming at supporting its continuous success.

2.3. Data-driven methods for estimating passengers on-board public transport

Although many studies have focused on estimating travel demand over the years, there is limited literature related with real-time estimation of the numbers of passengers on-board public transport vehicles, and the main reason is the lack of reliable data sources to date. A notable example is He et al. (2018), where manual passenger counts have been used to develop a framework that allows the control of air-conditioning on-board buses in Beijing based on the number of passengers on-board. The proposed framework includes methods, such as Monte Carlo simulation, neural networks and Markov chains. Other studies have also utilised WiFi-enabled technologies for estimating passengers on-board. For example, Mikkelsen et al. (2016) develop an algorithmic framework to achieve estimations of passengers on-board buses in Aalborg, Denmark. A low-cost automatic data collection system that utilises smartphone MAC address data and a mathematical framework to estimate public transport occupancy is proposed by Vieira et al. (2020). Therein, the contribution potential of Kalman filtering techniques in reducing the prediction errors is highlighted.

Several methods have used AFC data for estimating passengers on-board, showcasing their value for monitoring service reliability and passenger behaviour, even if inference is often required. For instance, Zhang et al. (2017) employ trip chaining analysis with Kalman filtering, while Heydenrijk-Ottens et al. (2018) apply random forests and gradient boosting to tram networks in the Hague. Other studies use AFC for destination inference in incomplete datasets, such as Sánchez-Martínez (2017) in Boston's tap-in-only system. AFC has also been applied to study impacts of incidents and disruptions, for example in London's network (Freemark, 2013; Liu et al., 2021). These studies highlight the diverse applications and limitations of AFC in passenger load estimation.

In recent years, several studies have used APC data within different methods for estimating the number of passengers on-board public transport vehicles. Bayesian estimation (Khomchuk et al., 2018), neural networks (Hu et al., 2020), time series and machine learning (Pasini et al., 2019), Kalman filtering and support vector regression (Wang et al., 2021) are among the relevant methods present in the literature. Some studies fuse APC data with other data sources, such as Jenelius (2019b), where various methods (stepwise regression, lasso regression, and boosted tree ensembles) are used to predict real-time vehicle-specific on-board crowding levels in metro networks. Dib et al. (2023) combine AFC with APC data of partial coverage for estimating unified occupancy on a public transport network, using data from several French networks. Acknowledging the limitation of APC datasets in providing full information for the number of passengers on-board public transport vehicles, Roncoli et al. (2023a,b) develop a Kalman filter-based framework for filling the gaps in APC datasets without using other demand-related data sources (e.g., AFC). Their case study refers to the tram network of Nantes, France.

2.4. Summary and research gaps

Existing literature highlights the importance of quantifying (dis-)comfort due to crowding in public transport vehicles, and this can be achieved by estimating the number of passengers on-board. Methods developed in this field depend heavily on different types of data sources. Manual data collection techniques are often considered time and labour-intensive, while some advanced sensing technologies are, to date, insufficiently mature and can, therefore, not showcase their full capacity. Traditional, yet automated, data collection through APC devices is promising for several real-time applications in the field of vehicle occupancy and comfort. Previous studies, including the one by Roncoli et al. (2023a), allowed to improve incomplete APC datasets, making them usable within real-time applications for comfort estimation. However, the developed models refer to stations that serve only one line, which is more common for lighter rail modes (e.g. trams) in mid-size cities, but less common for heavy rail systems (e.g. commuter trains) or in large cities. Therefore, there is a critical need for further expanding the methodology to be usable in any complex network, in which stations are served by multiple lines, which is the goal of the present study.

3. On-board comfort estimation methodology

Motivated by the incomplete APC datasets typically available to transport operators and the resulting challenge of monitoring crowding and comfort on-board vehicles, the goal of this paper is to develop a framework that accurately estimates the number of passengers on-board public transport vehicles and translates these estimates into comfort levels as experienced by travellers. The proposed method relies on two data sources:

1. **APC data** from devices installed on a subset of operating vehicles, which report boarding and alighting counts per run and station.
2. **AVL data** for all operating vehicles, which provide the departure times of each vehicle from every station.

It should be noted that, regardless of the specific formatting of APC and AVL data, the required inputs for the proposed approach can always be obtained. Although APC coverage may vary across networks, the framework remains applicable as long as at least limited APC data are available (e.g., as low as 5–10% in Roncoli et al., 2023a). Hence, the proposed modeling is suitable for networks with intersecting lines where partial APC and full AVL data are available. Moreover, the proposed approach is capable of dealing with incomplete information due to intermittent measurement and communications failures; nevertheless, it is assumed that a baseline of information is available for each line, either through historical records or real-time observations, to ensure the identifiability of line-specific dynamics.

3.1. Estimation framework

The adopted technique for passenger estimation and, consequently, passenger comfort level estimation relies on the Kalman Filter (KF) (Kalman and Bucy, 1961; Anderson and Moore, 1979), which is an established methodology employed for different estimation and prediction purposes within transportation studies (see, e.g., Wang and Papageorgiou, 2005; Bekiaris-Liberis et al., 2016, 2017; Saeedmanesh et al., 2021; Trinh et al., 2022; Abewickrema et al., 2023). In the frame of a KF, for a variable (vector) x with a-priori (i.e. predicted) estimation denoted as $\hat{x}^-$, a-posteriori (i.e. updated) estimation denoted as $\hat{x}$, and a state-transition model A, for any time interval k, the following are employed:

$$\hat{x}^-(k) = A(k-1)\hat{x}(k-1) \tag{1}$$

$$P^-(k) = A(k-1)P^+(k-1)A(k-1)^T + Q \tag{2}$$

$$K(k) = P^-(k)C(k)^T\left[R + C(k)P^-(k)C(k)^T\right]^{-1} \tag{3}$$

$$\hat{x}(k) = \hat{x}^-(k) + K(k)[\bar{z}(k) - C(k)\hat{x}(k)] \tag{4}$$

$$P^+(k) = \left[I_d - K(k)C(k)P^-(k)\right] \tag{5}$$

where $Q = Q^T > 0$ is a tuning parameter that represents the (ideally known) covariance matrices of the process, K is the optimal Kalman gain, C describes the observation model of x, $R = R^T > 0$ is a tuning parameter for the measurement noise, $\bar{z}$ is a (noisy) measurement of x, and I_d represents an identity matrix of proper size. The a-priori estimation for variable x state at time k is obtained considering solely the previous (estimated) state at $k-1$ and the system's dynamics (1). Accordingly, the uncertainty of the a-priori

state, i.e., when only the system dynamics is considered, is calculated according to (2). The Kalman gain K is then calculated as the gain that minimises the residual error in term of the minimum mean square error of the estimation (3). The a-posteriori (updated) state estimate of x, i.e., when available measurements are considered, is calculated via (4). Finally, the uncertainty of the system's state when measurements are considered is computed via (5). The algorithm is initialised considering $\hat{x}(k_0) = \mu$ and $P(k_0) = H$, where μ and $H = H^T > 0$ represent the mean and auto-covariance of $\hat{x}(k_0)$ and $P(k_0)$, respectively. For clarity, a complete glossary of the notation employed throughout the paper is provided in Appendix A.

In this work, the KF algorithm presented above is separately applied to estimate two critical components: 1) the number of passengers boarding a vehicle ($\hat{b}_j^{\mathrm{r}}$), and 2) the alighting rate of a vehicle ($\hat{\gamma}_j^{\mathrm{r}}$), defined as

$$\hat{\gamma}_j^{\mathrm{r}} = \frac{\hat{a}_j^{\mathrm{r}}}{\hat{p}_j}, \tag{6}$$

where $\hat{a}_j^{\mathrm{r}}$ is the number of passengers that alight and $\hat{p}_j$ is the number of passengers on-board run j. This is performed by implementing two parallel KFs to estimate these quantities, which are then combined under the consideration of the conservation of passengers, enabling the estimation of the number of passengers on-board ($\hat{p}_j$), according to:

$$\hat{p}_j(k+1) = \hat{p}_j(k) + \hat{b}_j^{\mathrm{r}}(k) - \hat{a}_j^{\mathrm{r}}(k). \tag{7}$$

Replacing (6) into (7), the latter is rewritten as:

$$\hat{p}_j(k+1) = [1 - \hat{\gamma}_j^{\mathrm{r}}(k)] \cdot \hat{p}_j(k) + \hat{b}_j^{r}(k). \tag{8}$$

Values $\hat{p}_j$ are then used for estimating comfort levels on-board, as elaborated in greater detail in Section 3.5.

It should be noted that the current study applies KF to estimate the same two critical components as in Roncoli et al. (2023a); however, the dynamical process and measurement models are substantially modified, enabling additional estimation capabilities. This section proceeds first by presenting an overview of the previous study's methodology, then by describing in detail the proposed critical modelling changes. More specifically, Section 3.2 summarises the estimation models employed for the case of simple networks with one line per station. Each main modelling component is presented separately, with the dynamic modelling approach for the number of passengers boarding presented in Section 3.2.1 and the dynamic modelling approach for the number of passengers alighting in Section 3.2.2. The novel models for estimating the number of passengers boarding as proposed in this study in order to account for multiple lines per station are presented in Section 3.3. Section 3.4 discusses the observability properties of the proposed models. Finally, Section 3.5 presents the validation framework utilised in this study.

3.2. Overview of existing modeling framework for one line per station

In the presented estimation framework, a station is denoted by index $i \in I$, with I being the set of available stations within the modelled public transport network. A set of runs denoted by $j \in J$ operate across the network. It is assumed that the estimation algorithm runs on a daily basis, considering an "operational" day, which typically starts in the morning of a calendar day (usually at 4 or 5 AM) and finishes in the early hours of the next calendar day (usually at 1 or 2 AM). Considering the dynamic modelling proposed here and the definition of the discrete-time domain, a time step of size T in seconds (e.g., 30–60 s) is employed here, with the actual time defined as $t = kT$ where k is the time step index. For notational simplicity, all variables in the following sections are expressed as functions of the time-step index k only. In the case study implementation, we set $T = 1$, which effectively makes k the unit of time. However, the formulation is general and remains valid for other values of T. In addition, for every variable ω, $\bar{\omega}$ denotes its measured value, $\tilde{\omega}$ denotes its historical value, and $\hat{\omega}$ denotes its estimated value using the proposed methodology.

The proposed modelling framework is based on the assumption that the following information is available at any time k:

- Real-time AVL information (i.e. actual or scheduled departure times from the current and following stations) for all runs and at all stations, providing $\eta_{i,j}^{\mathrm{r}}(k), \forall i \in I, j \in J$.
- Real-time APC data for a limited number of runs $\bar{J} \subset J$, providing $\bar{b}_j^{\mathrm{r}}(k)$, $\bar{a}_j^{\mathrm{r}}(k)$, and $\bar{p}_j(k)$, for $j \in \bar{J}$. This allows to assign $\beta_j^{\mathrm{r}} = 1$ if $j \in \bar{J}$ and $\beta_j^{\mathrm{r}} = 0$ otherwise.
- Historical information obtained by processing AVL and APC data available for a set of previous days $d \in D$, providing boarding, lightning, and passengers on board.

The models employed in this paper for estimation are station-based, i.e., we estimate the boarding and alighting at every station, every time a vehicle arrives or departs; however, it is necessary to introduce relevant assumptions referring to vehicle-based quantities. Considering a reasonable time step T (e.g. above 30 s and below 2 min), for all public transport networks it is safe to assume that:

1. There is only one run j that departs from station i during time $(k-1, k]$
2. A run j can depart from only one station during a time interval $(k-1, k]$
3. At the time that a vehicle departs from station i, all passengers waiting to travel at time k will board the vehicle during $(k-1, k]$

It is noted that, in crowded public transport networks, and especially during peak hours, passengers might be left behind from the first arriving vehicle due to overcrowding phenomena on-board the vehicle (Sipetas et al., 2020). Due to the complexity entailed, the consideration of such phenomena is deemed beyond the scope of the present study, and will be explored as part of future work. It is, therefore, assumed here that everyone waiting at the station will be able to board their desired vehicle.

3.2.1. Dynamic model for boarding passengers

The dynamics for boarding passengers is modelled as follows:

$$\begin{bmatrix} w_i(k+1) \\ e_i(k+1) \end{bmatrix} = \begin{bmatrix} 1-\eta_i(k) & 1 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} w_i(k) \\ e_i(k) \end{bmatrix} + \xi_i(k) \tag{9}$$

where $w_i(k)$ represents the number of passengers waiting to board at station i at time k, $e_i(k)$ represents the number of passengers entering station i at time k, and $\xi_i(k)$ is an unknown modelling error (e.g. zero-mean Gaussian noise). Since $\eta_i(k)$ is a known (i.e. measured) variable, (9) represents a linear-parameter-varying system.

The measurement model is:

$$\begin{bmatrix} \bar{z}_i(k) \\ \tilde{z}_i(k) \end{bmatrix} = \begin{bmatrix} \beta_i(k) & 0 \\ 0 & \phi(1-\beta_i(k)) \end{bmatrix} \cdot \begin{bmatrix} w_i(k) \\ e_i(k) \end{bmatrix} + \psi_i(k) \tag{10}$$

where $\psi_i(k)$ is an unknown measurement error (e.g. zero-mean Gaussian noise). The measurement $\bar{z}_i(k)$ is obtained from quantities available at any time step k as follows:

$$\bar{z}_i(k) = \sum_{j \in J} \eta^{\mathrm{r}}_{i,j}(k) \cdot \bar{b}^{\mathrm{r}}_j(k). \tag{11}$$

The measurement $\tilde{z}(k)$ is obtained from historical data. Historical values consider the number of passengers having entered a station during a number of d historical days. The historical value of the number of passengers entering station i in interval T to board during $(\tilde{k}\tilde{T}, (\tilde{k}+1)\tilde{T}-1]$ is:

$$\tilde{z}_i(\tilde{k}) = \frac{\sum_{d \in D} \sum_{k=\tilde{k}\tilde{T}}^{(\tilde{k}+1)\tilde{T}-1} \frac{\bar{b}^{\mathrm{d}}_{i,d}(k)}{t^{\mathrm{d}}_{i,d}(k)}}{\sum_{d \in D} \sum_{k=\tilde{k}\tilde{T}}^{(\tilde{k}+1)\tilde{T}-1} \beta^{\mathrm{d}}_{i,d}(k)}, \tag{12}$$

where $\tilde{k}$ is the bin index, $\tilde{T}$ is the bin size, D is the set of days considered, $b^{\mathrm{d}}_{i,d}(k)$ is the number of vehicles boarding at station i during $(k-1, k]$ on day d, $t^{\mathrm{d}}_{i,d}(k)$ is the time since the last vehicle stopped at station i on day d considering the current time k, and $\beta^{\mathrm{d}}_{i,d}(k)$ equals 1 if a vehicle departing from station i during $(k-1, k]$ on day d provides passenger loads information, and 0 otherwise.

3.2.2. Dynamic model for alighting passengers

The alighting rate per station, $\gamma_i(k)$, is defined as:

$$\gamma_i(k) = \sum_{j \in J} \eta^{\mathrm{r}}_{i,j}(k) \cdot \gamma^{\mathrm{r}}_j(k) \tag{13}$$

where $\eta^{\mathrm{r}}_{i,j}(k)$ equals 1 if run j departs from station i and 0 otherwise.

Due to a lack of physical dynamics, the alighting rate is modelled via a purely data-driven approach, considering random walk dynamics as follows:

$$\gamma_i(k+1) = \gamma_i(k) + \xi^{\gamma}_i(k), \tag{14}$$

where $\xi^{\gamma}_i(k)$ is zero-mean Gaussian noise. The measurement model is:

$$\begin{bmatrix} \bar{z}^{\gamma}_i(k) \\ \tilde{z}^{\gamma}_i(k) \end{bmatrix} = \begin{bmatrix} \beta_i(k) \\ \phi(1-\beta_i(k)) \end{bmatrix} \gamma_i(k) + \psi^{\gamma}_i(k), \tag{15}$$

where $\psi^{\gamma}_i(k)$ is Gaussian noise.

The measurement $\bar{z}^{\gamma}_i(k)$ is obtained from quantities measured in real-time at any time step k, as

$$\bar{z}^{\gamma}_i(k) = \sum_{j \in J} \eta^{\mathrm{r}}_{i,j}(k) \cdot \frac{\bar{a}^{\mathrm{r}}_j(k)}{\bar{p}_j(k)}. \tag{16}$$

Similarly to the arriving passengers, $\tilde{z}^{\gamma}_i(k)$ is calculated employing historical alighting rates as:

$$\tilde{z}^{\gamma}_i(\tilde{k}) = \frac{\sum_{d \in D} \sum_{k=\tilde{k}\tilde{T}}^{(\tilde{k}+1)\tilde{T}-1} \frac{\bar{\gamma}^{\mathrm{d}}_{i,d}(k)}{t^{\mathrm{d}}_{i,d}(k)}}{\sum_{d \in D} \sum_{k=\tilde{k}\tilde{T}}^{(\tilde{k}+1)\tilde{T}-1} \beta^{\mathrm{d}}_{i,d}(k)}, \tag{17}$$

where $\bar{\gamma}^{\mathrm{d}}_{i,d}(k)$ is the alighting rate at station i during $(k-1, k]$ in day d.

3.3. Modelling stations serving multiple lines

The case of a station serving multiple lines, which partly overlap their routes, introduces a major challenge for estimating the number of boarding passengers. The models described in Section 3.2 are based on the assumption that every traveller who is waiting to board a vehicle will board the first arriving vehicle, irrespectively of the line that it operates in, thus do not allow to capture the interactions among different lines that might operate at the same station.

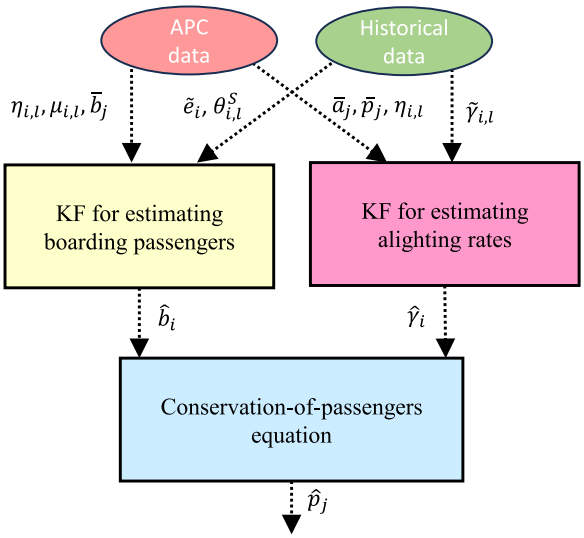

Fig. 1. The proposed estimation scheme.

In fact, in complex networks, travellers may arrive at a station i that serves multiple lines $l \in L$ and multiple runs per line $j \in J_l$ and are willing to board a vehicle only if that serves their destination station. In addition, they might choose to board the vehicle of the first line that arrives and serves their destination or, if they have prior knowledge of the time schedules of all lines, they might board the vehicle of a line that departs later from the current station but is the first to arrive at their destination station. Indeed, when a station serves multiple lines, certain lines often connect to more distant stations with fewer intermediate stops, while others maintain service to all minor stations along the route. Furthermore, some of these lines might operate with small time differences, hence, the service frequency of the station is no longer equivalent to the service frequency associated with only one line. Therefore, the waiting times for later lines' arrivals are not necessarily high. For these reasons, passenger demand might be distributed among lines per station in ways that need to be properly modelled.

Considering the above, the dynamics of the state vector of the overall system, previously described by Eq. (9), are redefined as follows (note that a variable with superscript ω^{ml} denotes a new version of a previously introduced variable ω redefined for estimation in the case of multiple lanes):

$$\begin{bmatrix} w_i^{\text{ml}}(k+1) \\ e_i^{\text{ml}}(k+1) \end{bmatrix} = \begin{bmatrix} 1 - \sum_{l \in L} \eta_{i,l}^{\text{ml}}(k) \cdot \theta_{i,l}(k) & 1 \\ 0 & 1 \end{bmatrix}(k) \cdot \begin{bmatrix} w_i^{\text{ml}}(k) \\ e_i^{\text{ml}}(k) \end{bmatrix} + \xi_i^{\text{ml}}(k), \tag{18}$$

where $\eta_{i,l}^{\text{ml}}(k)$ is a binary variable indicating if a vehicle of line l is at (or departs from) station i during $(k-1, k]$, which can be calculated as $\eta_{i,l}^{\text{ml}}(k) = \sum_{j \in J_l} \eta_{i,j}^{\text{r}}$. Eq. (18) introduces a novel parameter, $\theta_{i,l}(k)$, which represents the ratio of all passengers entering station i that will board a vehicle of line l. This value should be calibrated considering that the passengers who enter station $i \in S$ during $(k-1, k]$, namely $e_i^{\text{ml}}(k)$, will face a choice among all runs departing within time horizon h and will board the next available run of line $l \in L$ that will arrive at their destination station first. In order to calibrate $\theta_{i,l}(k)$, we propose the following algorithm:

1. The runs operating on lines l departing from station i within horizon h are identified.
2. The first departing run of each line l, identified in step 1, is considered, disregarding the remaining ones.
3. The stations, $S_{i,l}$, that each alternative line serves after the current station i, are identified.
4. For every station $m \in S_{i,l}$ after the current station i, the run of line l that arrives first is identified, which enables determining the values $\mu_{i,m,l}(k)$. This is a binary variable equal to 1 if the first run of line l departing from the current station i within time interval $[kT, kT + h)$ serves station $m \in S_{i,l}$ before the runs of every other lines that depart from station i within a time horizon h and serve station m, and equal to 0 otherwise. According to this definition, it always holds that $\sum_{l \in L} \mu_{i,m,l} \leq 1$.
5. For each station $m \in S_{i,l}$ after station i, the attractiveness, $\tilde{a}_m^{\text{q}}(k)$, is calculated based on the expected number of passengers that alight at this station during time period $[\lambda q\ , (\lambda + 1)q)$; this could be estimated, e.g., on the basis of historical APC alighting records of N previous days.
6. Finally, the ratio of all passengers entering station i that will board line l during time $t = kT$ is calculated as follows:

$$\theta_{i,l}(k) = \frac{\sum_{m \in S_{i,l}} \tilde{a}_m^{\text{q}}(k) \cdot \mu_{i,m,l}(k)}{\sum_{l \in L} \sum_{m \in S_{i,l}} \tilde{a}_m^{\text{q}}(k) \cdot \mu_{i,m,l}(k)}. \tag{19}$$

The time horizon h represents the forward-looking window at each station, within which runs of all lines departing are considered as alternatives for newly arriving passengers. Passengers entering station i during $(k-1, k]$ are therefore assumed to distribute only

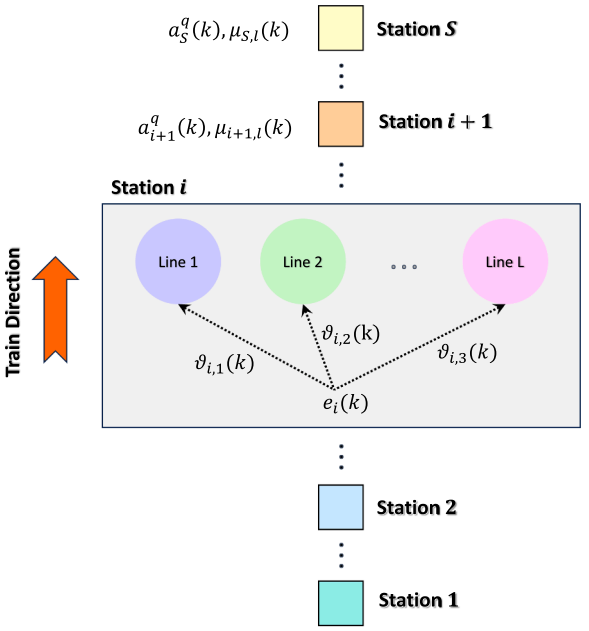

Fig. 2. The proposed calibration scheme for the ratio of passengers per line.

among runs departing within the next h time steps. Thus, h can be defined as

$$h = \min\left\{\bar{H},\ h_{\max}\right\}, \tag{20}$$

where $\bar{H}$ is the maximum headway of all lines serving the station, and $h_{\max}$ is an upper bound related to passenger arrival behavior, i.e., $h_{\max}$ indicates how long before the train's departure passengers are assumed to arrive at the station. In practice, the exact value of $h_{\max}$ is case-specific and should be calibrated empirically.

Since APC devices only provide information on passenger loads for some runs, station-level arrival rates are not directly measured. To address this, the framework first estimates the total number of passengers arriving at each station within a given time interval based on historical patterns and then distributes these passengers across departures of different lines using the splitting algorithm described above. This aggregation bridges the gap between limited train-level observations and station-level demand, enabling accurate estimation of on-board loads and comfort levels. Moreover, broad applicability is maintained using only APC records and without requiring additional data sources (e.g., station camera feeds).

The process for calibrating $\theta_{i,l}(k)$ is illustrated in Fig. 2. It should be noted that $\sum_{l\in L}\theta_{i,l}(k) = 1$ must always hold. It is also noted that both $\theta_{i,l}(k)$ and $\eta_{i,l}(k)$ are known (i.e. measured) variables, hence, Eq. (18) also represents a linear-parameter-varying system. A special case must be addressed in the calibration framework when there are no historical alightings after station i on line l during $[kT, kT+h)$, which is common during late-night or early-morning runs. Yet, some passengers are waiting to board on the current day. This may occur due to: 1) a shift in demand patterns, or 2) missing historical APC data caused by technical issues. In such cases, the proposed framework would incorrectly estimate $\theta_{i,l}(k) = 0$, preventing from capturing any boardings; this is addressed by replacing it with $\theta_{i,l}(k) = 1$.

Considering the above modifications, the measurement model, previously described by Eq. (10), becomes:

$$\begin{bmatrix} \bar{z}_{i,1}^{\mathrm{ml}}(k) \\ \vdots \\ \bar{z}_{i,l}^{\mathrm{ml}}(k) \\ \vdots \\ \bar{z}_{i,L}^{\mathrm{ml}}(k) \\ \bar{z}_{i}^{\mathrm{ml}}(k) \end{bmatrix} = \begin{bmatrix} \beta_{i,1}(k) & 0 \\ \vdots & \vdots \\ \beta_{i,l}(k) & 0 \\ \vdots & \vdots \\ \beta_{i,L}(k) & 0 \\ 0 & (1-\sum_{l\in L}\beta_{i,l}(k)) \end{bmatrix} \cdot \begin{bmatrix} w_i^{\mathrm{ml}}(k) \\ e_i^{\mathrm{ml}}(k) \end{bmatrix} + \psi_i^{\mathrm{ml}}(k), \tag{21}$$

where $\psi_i^{\mathrm{ml}}(k)$ is an unknown measurement error (e.g. zero-mean Gaussian noise). It is noted that $\beta_{i,l}(k)$ is a binary variable that equals 1 if the vehicle of line l that departs from station i is equipped with APC and zero if not. Furthermore, historical data is always included in estimation, considering its beneficial effects on results as demonstrated by Roncoli et al. (2023a). The measurement $\bar{z}_{i,l}^{\mathrm{ml}}(k)$, defined for $l \in 1, \dots, L$, is obtained from measured quantities available at any time step k as follows:

$$\bar{z}_{i,l}^{\mathrm{ml}}(k) = \sum_{j\in J_l} \eta_{i,j}^{\mathrm{r}}(k)\cdot \bar{b}_j^{\mathrm{r}}(k). \tag{22}$$

The historical value of the number of passengers entering station i in interval T during $(\tilde{k}\tilde{T}, (\tilde{k}+1)\tilde{T}-1]$, previously described by Eq. (12), is updated as follows:

$$\bar{z}_i^{\mathrm{ml}}(\tilde{k}) = \frac{\sum_{d\in D}\sum_{l\in L}\sum_{k=\tilde{k}\tilde{T}}^{(\tilde{k}+1)\tilde{T}-1} \frac{\bar{b}^{\mathrm{d}}_{i,l,d}(k)}{f^{\mathrm{d}}_{i,l,d}(k)}}{\sum_{d\in D}\sum_{l\in L}\sum_{k=\tilde{k}\tilde{T}}^{(\tilde{k}+1)\tilde{T}-1} \beta^{\mathrm{d}}_{i,l,d}(k)}, \tag{23}$$

where $b^{\mathrm{d}}_{i,l,d}(k)$ is the number of passengers boarding a vehicle of line l at station i in day d. One choice for the value $f^{\mathrm{d}}_{i,l,d}(k)$ can be the time difference between consecutive vehicle departures from the station, considering relatively uniform departures with reasonable time differences. Alternatively, a simpler approach would be to assume a constant value ($f^{\mathrm{d}}_{i,l,d}(k) \equiv \bar{f}$) that reflects how much earlier than the train departure the passengers that board a vehicle of line l arrive at the station i (e.g. 15 min, or any other reasonable value). Our experiments revealed that the latter option yields superior estimation performance. This can be explained to the frequent occurrence, particularly during peak hours, of stations serving multiple lines with only a few of time difference between consecutive departures. Consequently, assuming that passengers boarding line l arrived at the station within the brief period since the previous departure results in overestimation of arrival rates. Applying the KF (1)–(5) for the system (18), (21), one may extract estimates $\hat{\iota w}_i^{\mathrm{ml}}(k)$, which are then employed to calculate $\hat{b}_j^{\mathrm{r}}(k)$, according to

$$\hat{b}_j^{\mathrm{r}}(k) = \sum_{i\in I} \eta^{\mathrm{r}}_{i,j}(k) \cdot \hat{\iota w}_i^{\mathrm{ml}}(k) \tag{24}$$

which can then be used in (8).

Regarding the alighting rates modelling, unlike the modelling of the passenger boarding process, we express the estimations per line l. Therefore, the alighting rate, previously described by Eq. (13), is now defined by $\gamma^{\mathrm{ml}}_{i,l}(k)$, i.e. expressed also per line l and it is defined as follows:

$$\gamma^{\mathrm{ml}}_{i,l}(k) = \sum_{j\in J_l} \eta^{\mathrm{r}}_{i,j}(k) \cdot \gamma^{\mathrm{r}}_j(k) \tag{25}$$

where $\eta^{\mathrm{r}}_{i,j}(k)$ equals 1 if run $j \in J_l$ departs from station i and 0 otherwise.

The alighting rate dynamics, previously described by Eq. (14), is now expressed per line l using again random walk dynamics as follows:

$$\gamma^{\mathrm{ml}}_{i,l}(k+1) = \gamma^{\mathrm{ml}}_{i,l}(k) + \xi^{\gamma,\mathrm{ml}}_{i,l}(k), \tag{26}$$

where $\xi^{\gamma,\mathrm{ml}}_{i,l}(k)$ is zero-mean Gaussian noise. The measurement model, previously described by Eq. (15), now becomes:

$$\begin{bmatrix} \bar{z}^{\gamma,\mathrm{ml}}_{i,l}(k) \\ \bar{z}^{\gamma,\mathrm{ml}}_{i,l}(k) \end{bmatrix} = \begin{bmatrix} \beta_{i,l}(k) \\ 1-\beta_{i,l}(k) \end{bmatrix} \gamma^{\mathrm{ml}}_{i,l}(k) + \psi^{\gamma,\mathrm{ml}}_{i,l}(k), \tag{27}$$

where $\psi^{\gamma,\mathrm{ml}}_{i,l}(k)$ is Gaussian noise.

The measurement $\bar{z}^{\gamma,\mathrm{ml}}_{i,l}(k)$ is obtained from measured quantities at any time step k as:

$$\bar{z}^{\gamma,\mathrm{ml}}_{i,l}(k) = \sum_{j\in J_l} \eta^{\mathrm{r}}_{i,j}(k) \cdot \frac{\bar{a}^{\mathrm{r}}_j(k)}{\bar{p}_j(k)}. \tag{28}$$

The historical alighting rates, previously described by Eq. (17), are now calculated as:

$$\bar{z}^{\gamma,\mathrm{ml}}_{i,l}(\tilde{k}) = \frac{\sum_{d\in D}\sum_{k=\tilde{k}\tilde{T}}^{(\tilde{k}+1)\tilde{T}-1} \bar{\gamma}^{\mathrm{d}}_{i,l,d}(k)}{\sum_{d\in D}\sum_{k=\tilde{k}\tilde{T}}^{(\tilde{k}+1)\tilde{T}-1} \beta^{\mathrm{d}}_{i,l,d}(k)}, \tag{29}$$

where $\bar{\gamma}^{d}_{i,l,d}(k)$ is the alighting rate of line l at station i during $(k-1, k]$ in day d.

Regarding the historical alighting rates, it is noted here that low APC availability and/or few historical days incorporated might lead to zero alighting rates in the relevant time bins when this occurs, simply because there would not have been enough values captured for the estimation. These occurrences might lead to underestimations of alighting passengers. In such cases, a suggested technique to address the issues caused by the zero values is to perform interpolation between consecutive non-zero values to replace the non-accurate zeros. Applying the KF (1)–(5) for the system (26), (27), one may extract estimates $\hat{\gamma}^{\mathrm{ml}}_{i,l}(k)$, which are then employed to calculate $\hat{\gamma}^{\mathrm{r}}_j(k)$, according to

$$\hat{\gamma}^{\mathrm{r}}_j(k) = \sum_{\substack{i\in I \\ l: j\in J_l}} \eta^{\mathrm{r}}_{i,j}(k) \cdot \hat{\gamma}^{\mathrm{ml}}_{i,l}(k) \tag{30}$$

which can then be used in (8).

The updated estimation framework that accounts for stations with multiple lines is summarised in Fig. 1 for a vehicle j. It is further noted that each station might serve more than one direction of movement. In such a case, each direction is treated as a separate station and the proposed methodology is implemented separately for each one.

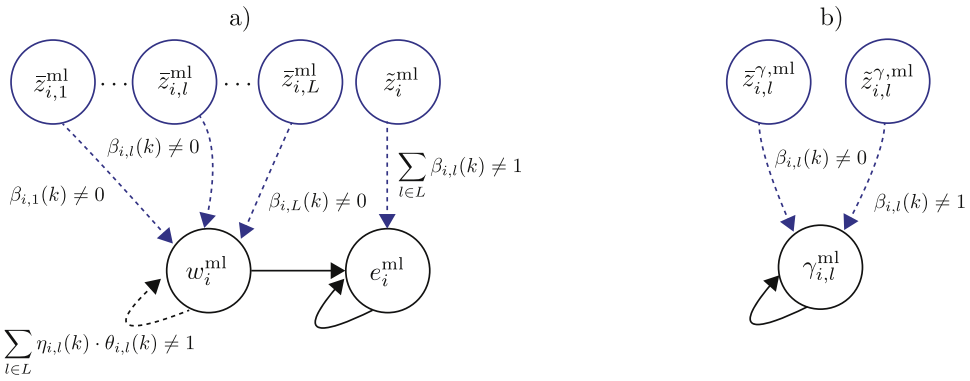

Fig. 3. The structural graphs $\mathcal{G}(\mathcal{A}^T, C^T)$ for patterns $\mathcal{A}$ and C that include matrices A and C, respectively, of a) the boarding process model (18) and boarding measurement model (21), and b) the alighting rate process model (26) and alighting rate measurement model (27). State variables are represented in black, while measurement variables are in blue. Dashed lines indicate that the edge may exist, depending on the conditions written next to them.(For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3.4. System observability

When dealing with real-time state estimation, the observability property guarantees that the dynamic evolution of internal states of a system (those not always directly measured), such as, in our case, boarding passengers and alighting rates, can be reproduced in real-time without bias from available partial measurements using an estimator such as the KF (Antsaklis and Michel, 2006). The observability of the systems formulated in the previous section is investigated following a graph-theoretic approach, denoted *structural observability*, by examining the system's structure, defined by zero and non-zero elements of the A and C matrices (Liu et al., 2013). This approach is preferable when dealing with time-varying or parameter-varying systems (as with our models), since the parameters affect the system's matrices in real-time. It should be noted that structural observability is a *necessary* condition for observability; however, a structurally observable system may temporarily lose observability when specific parameter combinations satisfy particular conditions. Conversely, if no parameter combinations exist that guarantee structural observability, no estimator would be able to reconstruct the system state from measured outputs.

Structure matrices $\mathcal{A}$ and C are introduced to characterise patterns of zero and non-zero elements in system matrices A and C, respectively. These patterns can be represented via graphs $\mathcal{G}(\mathcal{A}^T, C^T)$, shown in Fig. 3 for both the boarding passenger model, considering (18) and (21), and the alighting rate model, considering (26) and (27). We consider the following condition for structural observability (Liu et al., 2013). A linear system (A, C) is structurally observable if and only if: i) the graph $\mathcal{G}(\mathcal{A}^T, C^T)$ contains no non-accessible vertex; and ii) the graph $\mathcal{G}(\mathcal{A}^T, C^T)$ contains no dilation. Based on this definition, both systems satisfy the conditions for structural observability, meaning that there exist parameter combinations that guarantee observability. This is evident by assuming $\beta_{i,l} = 1$ (or more generally, non-zero) and observing that for both systems, all vertices can be accessed, while no dilation exists in the graphs. Additionally, specific parameter value combinations render the system observable or cause a temporary loss of observability, which is investigated by examining the resulting graphs when some dashed edges are removed. The following claims can be made:

- the boarding passenger system for a given station i is fully observable at time k when $\exists l \in L : \beta_{i,l}(k) \neq 0$ (i.e., when a vehicle operating on any line $l \in L$ equipped with APC stops at station i) and partly observable otherwise, i.e., when $\sum_{l\in L} \beta_{i,l}(k) \neq 1$ due to the non-accessibility of vertex w_i^{ml}; while vertex e_i^{ml} remains always accessible.
- the alighting rate system for each station i and line l is always observable, since, regardless of the value $\beta_{i,l}(k)$, one of the two dashed edges is present.

To summarise, the structural observability property holds and implies the proper operation of the proposed estimation scheme. Since the loss of observability occurs only temporarily (either for some states or for all states) in the cases described above, when observability is restored, the estimation capabilities of the proposed scheme are again guaranteed.

It is noted that APC data incompleteness may also arise under different patterns. Even for APC-equipped vehicles, measurements may be missing due, e.g., to technical errors including missing sensor data or communication disruptions. In such cases, there is no change in the observability results discussed earlier, i.e., the system remains structurally observable, implying a correct functioning of the proposed methodology. On the contrary, if APC data are entirely unavailable for all vehicles operating on a given line, both in real time and historically, key quantities such as the passenger split ratios $\theta_{i,l}$ and the historical alighting rates $\tilde{z}_{i,l}^{\gamma,\text{ml}}$ cannot be calculated. This results in a) the boarding passenger system becoming not well-posed because the state transition matrix depends on the uncomputable quantity $\theta_{i,l}$ and b) the alighting rate system becoming unobservable for all lines $l^{\text{miss}} \in L^{\text{miss}} \subset L$ since there is no historical data and $\beta_{i,l^{\text{miss}}} = 0$, meaning the vertex $\gamma_{i,l^{\text{miss}}}^{\text{ml}}$ is non-accessible.

3.5. Evaluation and validation framework

The proposed method aims at estimating the number of passengers on-board vehicles through the numbers of passengers boarding and alighting at stations for the runs that do not have APC records. However, since information is not available for runs without APC records, an alternative validation approach is developed, which employs only data from APC-equipped runs. Specifically, when the estimation algorithm processes an APC-equipped run, estimation is first performed assuming that no APC measurement is available, denoting such estimate $\check{\omega}$ (for a generic variable ω) and storing the respective estimated values; then, the estimates are re-calculated considering the available APC measurements, and these latter values are then used for the subsequent estimation steps. In this way, the estimation method can be evaluated in a fair manner through a comparison of the estimates obtained without considering APC measurements ($\check{\omega}$) with their corresponding actual measurements ($\bar{\omega}$).

The number of passengers on-board can be further translated into comfort by utilising existing methods. In the present study, the Level of Service (LOS) framework of Chandakas (2009) is utilised. This framework is used to measure in-vehicle comfort on-board PT vehicles by assigning a comfort level on the basis of the on-board density of standing passengers (D) and of the proportion of occupied seats (R). The LOS and the respective conditions are as follows:

- LOS 1: $R \leq 0.25$ (free choice of a seat)
- LOS 2: $0.25 < R \leq 0.5$ (possibility to choose a seat, constrained to availability)
- LOS 3: $0.5 < R \leq 1$ (possibility to travel seated, limited choice available)
- LOS 4: $0 < D \leq 1.8$ (standing conditions tolerated)
- LOS 5: $1.8 < D \leq 4$ (no tolerance from standees, reliability issues)
- LOS 6: $D > 4$ (density greater than 4 standees per m^2, exceptional situation).

It is noted that LOS 1-3 are defined based on seat occupancy (R), while LOS 4-6 are defined based on standing passenger density (D > 0). In practice, LOS 4-6 typically occur when all seats are occupied (R > 1). The Level of Service (LOS) framework by Chandakas (2009) is adopted here due to its alignment with the proposed modeling framework and for consistency with earlier studies (i.e., Roncoli et al., 2023a), but any alternative comfort measure could equally be applied. Note that, since the transformation from the estimated boardings and alighting rates to LOS is nonlinear, there is no general guarantee of unbiasedness of LOS estimation. However, our numerical results show that, although there is certainly an impact on the variance of the resulting comfort estimates, the empirical bias looks minimal.

In this paper, the validation framework is composed of two levels. First, the performance of the models for estimating the number of passengers boarding, the estimated alighting rate and ultimately the estimated number of passengers on-board, are evaluated. Then, the model outputs are fed to the comfort level estimation framework of Chandakas (2009).

For the first level of validation, the estimations performed at the station level throughout the duration of an operational day regarding the number of passengers boarding, the alighting rate and the number of passengers on-board, are considered. For each one of these measures the following aggregated measures are calculated: the mean absolute error (MAE) for quantifying the average magnitude of the estimation differences from real values (31); the root mean squared error (RMSE) for capturing how well the estimated values align with the observed values, hence, targeting the evaluation of estimation accuracy (32); and the mean absolute percent error (MAPE), as another metric for targeting the quantification of accuracy (33). The metrics are calculated as follows:

$$\text{MAE} = \frac{\sum_{k=0}^{K} \sum_{i \in I} \sum_{l \in L} |\check{\omega}^{s}_{i,l}(k) - \bar{\omega}^{s}_{i,l}(k)|}{\sum_{k=0}^{K} \sum_{i \in I} \sum_{l \in L} \beta_{i,l}(k)} \tag{31}$$

$$\text{RMSE} = \sqrt{\frac{\sum_{k=0}^{K} \sum_{i \in I} \sum_{l \in L} |\check{\omega}^{s}_{i,l}(k) - \bar{\omega}^{s}_{i,l}(k)|^2}{\sum_{k=0}^{K} \sum_{i \in I} \sum_{l \in L} \beta_{i,l}(k)}} \tag{32}$$

$$\text{MAPE} = \frac{\sum_{k=0}^{K} \sum_{i \in I} \sum_{l \in L} |\check{\omega}^{s}_{i,l}(k) - \bar{\omega}^{s}_{i,l}(k)|}{\sum_{k=0}^{K} \sum_{i \in I} \sum_{l \in L} |\check{\omega}^{s}_{i,l}(k)|} \tag{33}$$

These equations consider that at each station $i \in I$ and for each line $l \in L$ the actual value (i.e. number of passengers boarding, the alighting rate, or the number of passengers on-board) $\bar{\omega}^{s}_{i,l}$ is available (through true measurements), and the estimation $\check{\omega}^{s}_{i,l}$ is for an estimation horizon K (i.e. one operational day in this study).

Although the performance of each one of the proposed method's components is analysed separately (i.e. boardings, alightings and passengers on-board), the main focus and concern here is on how well the comfort of passengers on-board can be estimated. In fact, this is critical information with a high value for real-life applications. In this case, the sequence of estimated comfort levels is compared with the sequence of comfort levels calculated from actual measured data. The proposed models enable the quantification of comfort levels per run and station. Therefore, for a run $r \in R_j$ that serves $j \in N_r$ stations, estimations $\bar{\omega}_{r,j}$ of the true comfort levels $\check{\omega}_{r,j}$ can be obtained. The performance metrics considered for this level of validation are defined as follows:

- Percent precision $= \dfrac{\sum_{r=1}^{R} \sum_{j=1}^{N_r} \mathbf{1}\left(\bar{\omega}_{r,j} = \check{\omega}_{r,j}\right)}{\sum_{r=1}^{R} N_r} \times 100,$ (34)

  which reflects the percentage of estimations that are precise, therefore, the estimated comfort level equals the true comfort level.

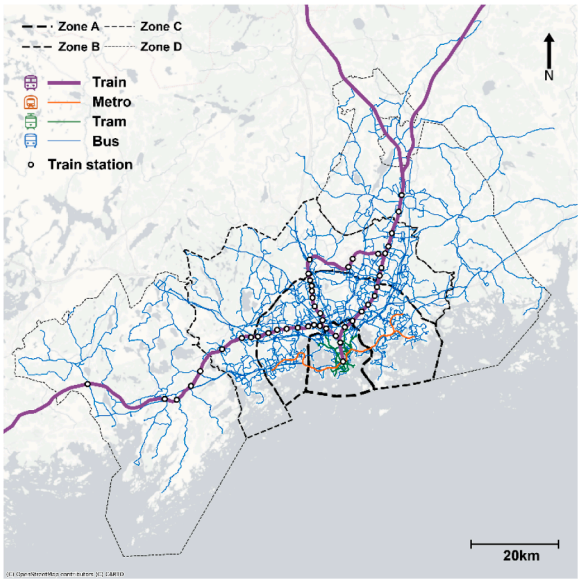

Fig. 4. The Helsinki PT network. (Source: Huang et al., 2023) .

- Absolute maximum error $= \max_{r=1,\ldots,R}\left(\max_{j=1,\ldots,N_r}\left|\bar{\omega}_{r,j} - \breve{\omega}_{r,j}\right|\right)$, (35)
  which expresses the maximum absolute difference between real and estimated comfort level met across all runs and stations.
- Average Absolute Error $= \frac{1}{\sum_{r=1}^{R} N_r}\sum_{r=1}^{R}\sum_{j=1}^{N_r}\left|\breve{\omega}_{r,j} - \bar{\omega}_{r,j}\right|$, (36)
  which represents the average absolute difference between real and estimated comfort level met across all runs and stations, considering R runs in total.

4. Study site and available dataset

4.1. Study site

This paper considers the commuter train system of the Helsinki metropolitan region (Finland) as a case study. The area covers approximately 770 km^2 and has a population of 1.2 million. There are a variety of mobility options within the area, from fixed public transport routes (metro, tram, train, bus, and ferry) to flexible micro-mobility (shared e-scooters and shared bicycles) and ride-hailing services (e.g. UBER). The local public transport operator is the Helsinki Regional Transport Authority (or HSL from "Helsingin Seudun Liikenne"). A geographical map of the overall public transport network is presented in Fig. 4, with the commuter train network shown in purple colour. The operational area of HSL is split in four fare zones, indicated as Zone A, B, C and D in Fig. 4. Stops outside of these zones are operated by another transport provider and are not part of this study. The train lines start from the centre of Helsinki (i.e. the centre of Zone A) and extend to the North, North-East, West, and South-West, covering a major part of the area. Some train stations offer transfer opportunities to and from the other available modes (i.e. metro, tram, and bus), and typically multiple lines operate from these stations.

The HSL commuter train network's stations and lines are shown in Fig. 5. The green-coloured parts of the illustrated lines represent stations and operations outside the HSL area. The train network comprises 12 lines, two of which are circular (serving Helsinki's airport). The major stations are Helsinki and Pasila, while other important stations include Huopalahti, Leppävaara, Tikkurila, Kerava, Espoo and Kirkkonummi. Excluding the circular lines, the network can be divided into two wings: west, including lines A, E, L, U, X, and Y, and east, including lines K, R, T, D, Z. The current study focuses on the west wing and considers both directions of movement. We denote the two directions of movement as the "Helsinki-outbound" and the "Helsinki-inbound". As discussed above, in the frame of the proposed modelling, since each physical station serves two directions, each physical station is split into two "stations" with two distinct indices.

The train vehicles utilised are Stadler FLIRT EMU, procured by Junakalusto Oy. In total, 81 vehicles of the type Sm5 operate to serve the daily commuting needs of travellers within the metropolitan area and in Southern Finland. The vehicles are composed of a low and a high floor. The low floor facilitates access by travellers with heavy luggage, prams, bicycles and wheelchairs. The length of each Sm5 train is 75 metres. A train in operation can be composed of maximum three train units. The seating capacity of a train is 232 seats, including 28 fold-up seats. The standing capacity is estimated as 323 passengers, considering approximately 4 passengers per m^2. An overview of the interior space of an Sm5 train is provided in Fig. 6.

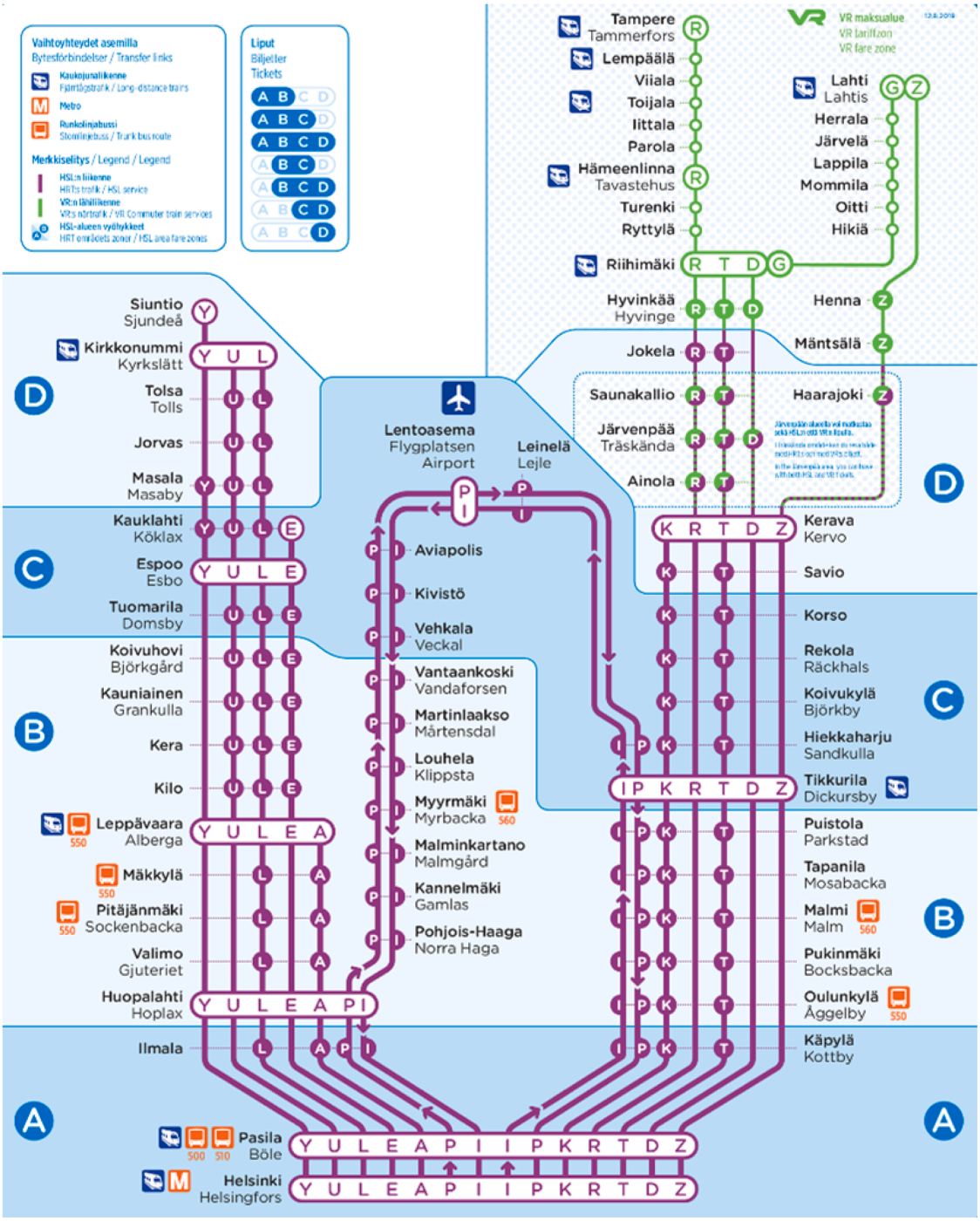

Fig. 5. A graphical representation of the Helsinki PT network (Source: HSL website (https://www.hsl.fi)) .

4.2. Data

The APC dataset utilised in this study has been provided by HSL and refers to all 30 days (i.e. comprising both weekdays and weekends) of September 2021. In the present study, APC data provided by HSL have AVL information already incorporated; therefore, no further fusion is needed. It should be noted that mobility patterns in September 2021 were likely influenced by residual COVID-19 effects, which may have led to lower overall passenger demand compared to pre- and post-pandemic levels. However, since the proposed modeling framework is designed to operate across any demand level, these conditions do not limit its applicability.

Fig. 6. The interior space of Helsinki trains a) the seating area, and b) the standing area including fold-up seats and special space for wheelchair (Source: Junakalusto Oy website (www.junakalusto.fi)) .

The weekdays of 13-17.09.2021 are used here for calibrating the parameters (i.e., $\theta_{i,l}(k)$) and producing the necessary historical data (i.e., $\tilde{z}_i^{\mathrm{ml}}(\tilde{k})$ and $\tilde{z}_{i,l}^{\gamma,\mathrm{ml}}(\tilde{k})$). The week starting on 20.09.2021 is used for testing and validation (weekdays only). HSL has equipped all line vehicles of the west wing with APC devices; however, technical problems often lead to missing records that need to be filled. For correct entries, the dataset provides full information about boardings, alightings, and on-board passengers, as well as info about the train locations and times of planned and actual departures per station.

Fig. 7 shows the distribution of the number of runs per day per line, considering the operation of a regular weekday (i.e. Wednesday 22.09.2021). It is apparent that the lines have very distinct operational characteristics, with some of them operating throughout most of the day (e.g. line U) and others operating only during specific time periods (e.g. line L). Line A seems to be the dominant line in the west wing, with not only the most runs per hour for most of the day (i.e. lowest headways and hence more frequent service), but also with adjustments to the operation during peak hours (indicated by the peaks in the respective bar charts for both directions). The "competition" among lines regarding frequency and speed of service is important in the present study, since it affects the ratio of passengers per line, $\theta_{i,l}(k)$, which is an essential part of the proposed modelling approach. It is noted that the average headway of vehicles per station per day is approximately 18 min for Helsinki outbound and Helsinki inbound, excluding Siuntio station that is scarcely served (i.e. approximately every two hours on average).

A critical aspect of this dataset is the percentage of available APC records. As shown in Fig. 8, the record availability for the week of 20.09.2021 is very good, with a weekly average availability reaching 86% for the Helsinki-outbound direction and 82% for the Helsinki-inbound direction, considering all lines together. The technical errors that lead to missing records here are random, however, in some cases, they might be persistent. For example, for line X in the Helsinki-inbound direction, only one run per day is available, and the respective records are missing for the entire week studied here. In both directions, the average weekly availability per line ranges from 64% (Line L in the Helsinki-inbound direction) to 91% (Line U in the Helsinki-inbound direction), excluding the 0% availability of Line X in the Helsinki-inbound direction.

Figs. 9 and 10 present the average boardings and alightings for the two directions considering five weekdays of available historical APC data. The historical days refer to the five days before Wednesday 22.09.2021, to demonstrate how the historical demand is shaped during one of the five weekdays that are used in this analysis. As shown in Fig. 9a, Helsinki station exhibits a peak of boardings in the Helsinki-outbound direction around 5 pm, which refers to the evening peak. Similar observations apply to Pasila, which is the next station after Helsinki. This station represents a major transportation hub with proximity to key locations of the Helsinki metropolitan area, referring to business districts, commercial centres, and residential areas. In the Helsinki-inbound direction, large numbers of passengers board at Leppävaara with two distinct peaks: one around 8 am and one around 5 pm (Fig. 9b). Considering the alightings in Fig. 10a, high passenger volumes moving from Helsinki alight at Leppävaara during the evening hours. This station is also a major hub offering several transfer options to other modes. In the Helsinki-inbound direction (Fig. 10b), most passengers alight at Pasila and Helsinki during two distinct peaks. The alighting patterns are significant for the purposes of the present study, as the ratio of passengers per line, $\theta_{i,l}(k)$, is calculated based on them.

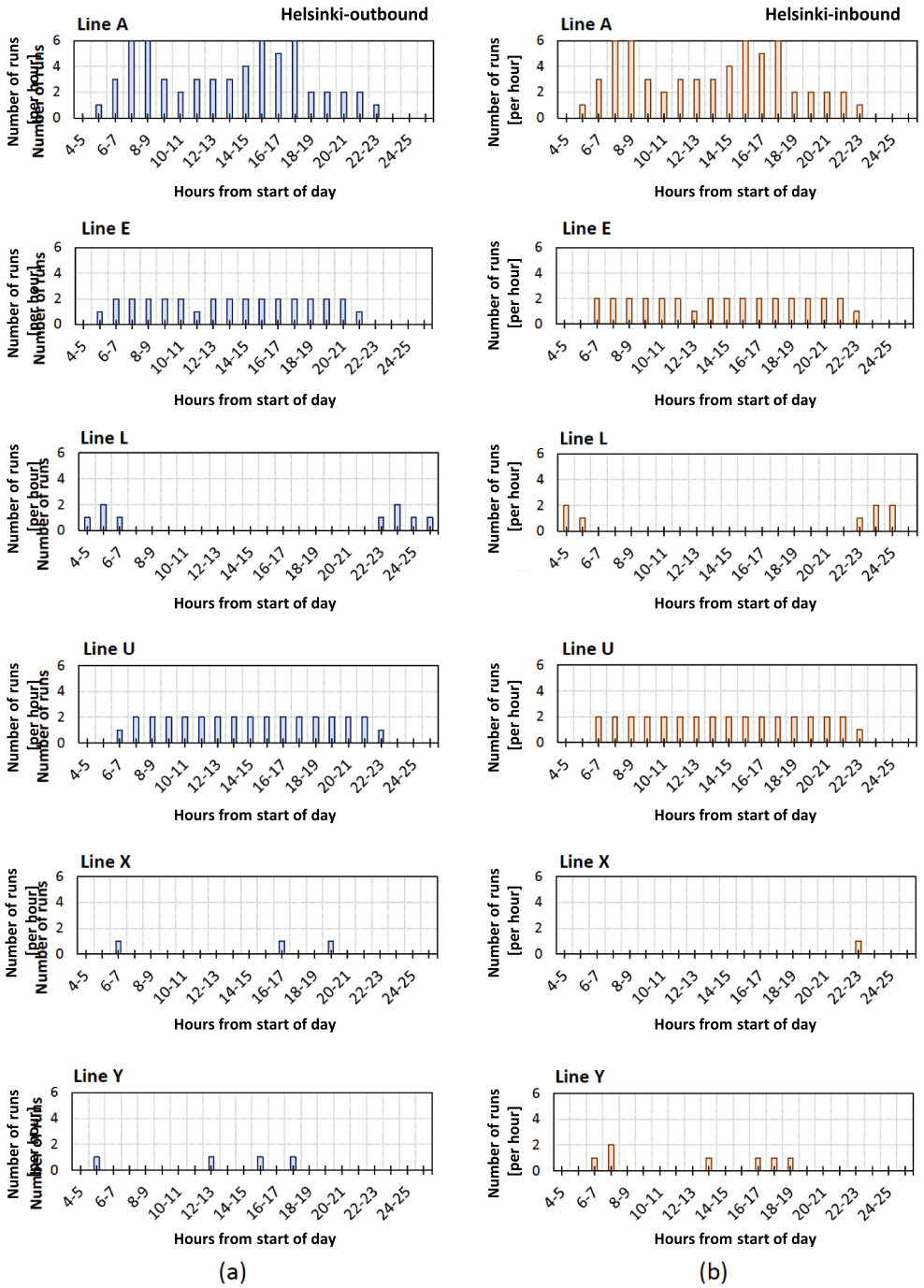

Fig. 7. Distribution of the number of runs per hour for Wednesday 22.09.2021 for a) Helsinki-outbound direction, and b) Helsinki-inbound direction.

5. Results and discussion

5.1. Passengers boarding, alighting rates and passengers on-board

The validation results from implementing the proposed methodology on the Helsinki commuter train network are presented in this section for five regular weekdays in September 2021. These results are obtained considering a time horizon of $h = 15$ min and five historical days of alightings per station in order to estimate the splitting factor, θ. The choice of h is based on the operational characteristics of the network, with frequent vehicle dispatches (Fig. 7). The frequent service per station implies that a 15-minute time

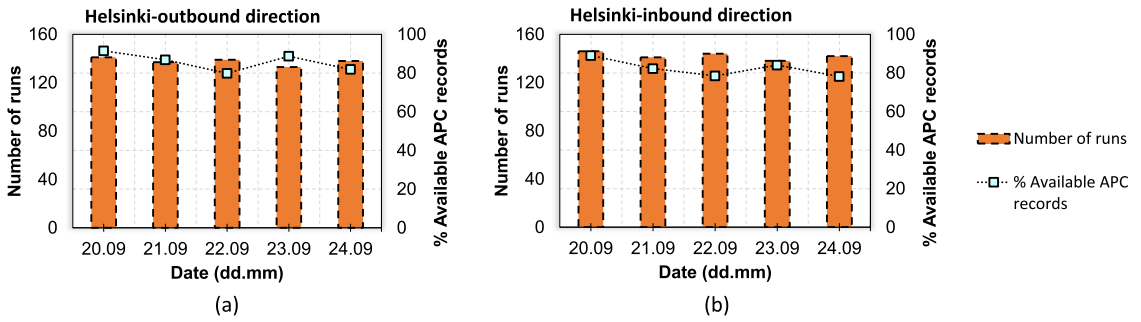

Fig. 8. Total number of runs per day and percent record availability for a) Helsinki-outbound direction, and b) Helsinki-inbound direction.

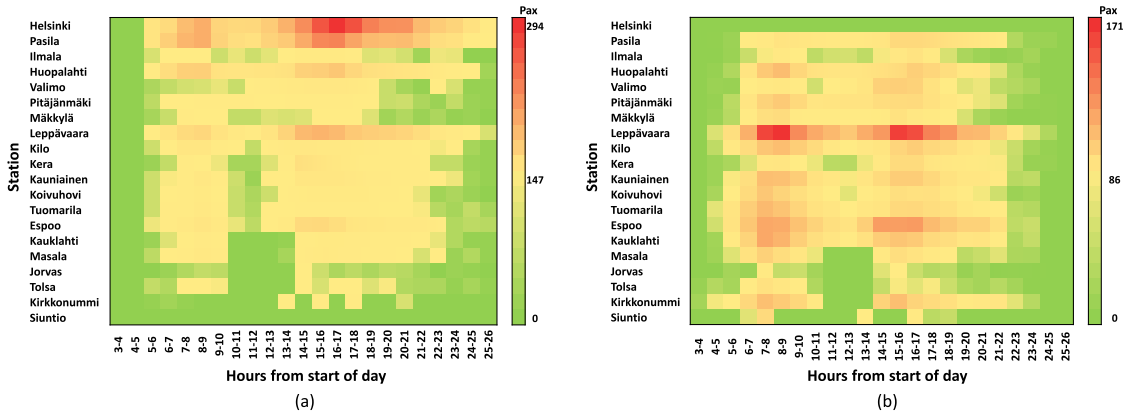

Fig. 9. Average number of passengers boarding considering 5 historical days for a) Helsinki-outbound direction, and b) Helsinki-inbound direction (before Wednesday 22.09.2021).

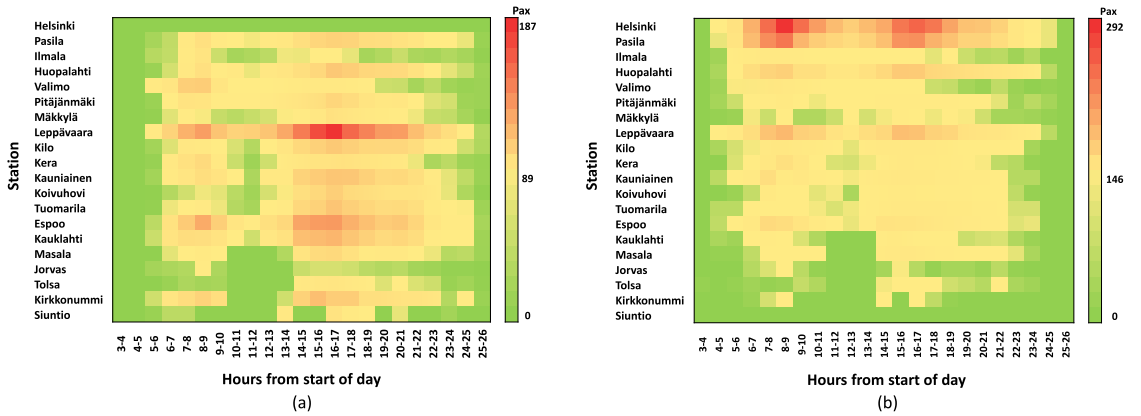

Fig. 10. Average number of passengers alighting considering 5 historical days for a) Helsinki-outbound direction, and b) Helsinki-inbound direction (before Wednesday 22.09.2021).

interval is sufficient for capturing the candidate runs and lines that depart from the station after the traveller arrives, enhanced also by the reasonable assumption that travelers do not arrive at the stations earlier than 15 min before their train's departure when the service is frequent. The choice of five historical days for determining the attractiveness of stations in terms of historical alightings is based on the assumption that five days will be enough to capture variations of demand patterns and also address challenges associated with potential lower APC data availability on some days. For similar reasons, historical information of five days about arrival rates per station and alighting rates per run and station is incorporated in the modelling.

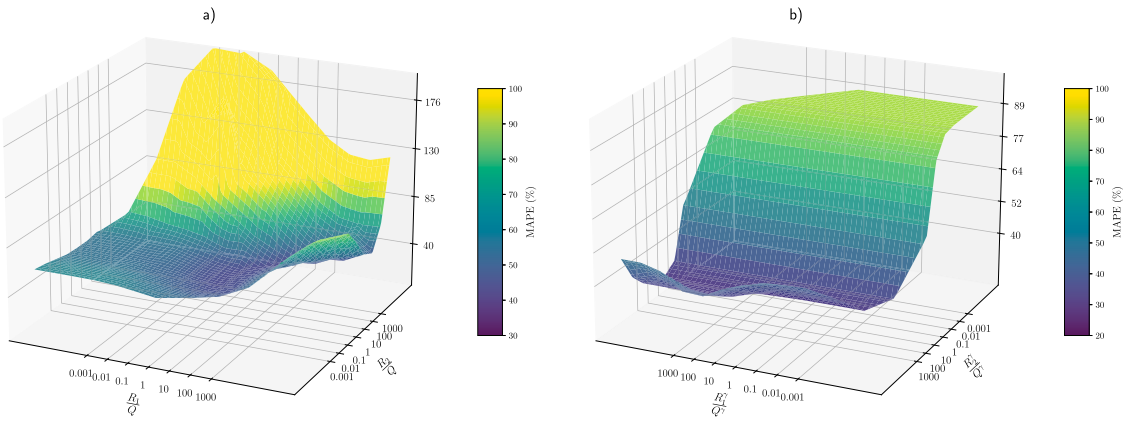

Fig. 11. Estimation performance (expressed as MAPE) as a function of covariance matrix parameters for process and measurement noise, for a) passenger boarding rate estimation and b) passenger alighting rate estimation.

Table 1
Estimation errors for boarding passengers.

Day	Helsinki-outbound			Helsinki-inbound		
	MAE (pax)	RMSE (pax)	MAPE (%)	MAE (pax)	RMSE (pax)	MAPE (%)
20.09.2021	4.41	6.29	46.48%	5.61	7.84	51.19%
21.09.2021	4.63	6.52	43.12%	5.85	8.15	56.11%
22.09.2021	4.66	6.52	44.27%	5.83	8.20	48.12%
23.09.2021	4.68	6.51	43.35%	5.99	7.95	54.41%
24.09.2021	4.63	6.22	47.49%	5.51	7.37	51.00%

Before analysing the estimation results, we investigate the choice of parameters to be employed in the Kalman filter, namely Q and R appearing in (2) and (3), respectively. We employ diagonal matrices, defined for the boarding estimator as

$$Q := \begin{bmatrix} q & 0 \\ 0 & q \end{bmatrix}, \qquad R := \begin{bmatrix} \bar{r} & & 0 & \\ & \ddots & & \\ 0 & & \bar{r} & \\ & & & \hat{r} \end{bmatrix} \tag{37}$$

and for the alighting rate estimator as

$$Q := q^{\gamma}, \qquad R := \begin{bmatrix} \bar{r}^{\gamma} & 0 \\ 0 & \hat{r}^{\gamma} \end{bmatrix}. \tag{38}$$

We perform a full factorial experiment, testing all combinations of parameters in the set $[0.001, 0.01, 0.1, 1, 10, 100, 1000]$. Fig. 11 presents the estimation performance for a representative sample day (Thursday 23.09.2021), in terms of MAPE, which, for visualisation purposes, is mapped against dimensionless ratios formed by the parameter combinations. It is evident that a large range of parameters leads to good estimation results, e.g., MAPE $\approx 43\%$ for boarding estimation and MAPE $\approx 31\%$ for alighting rate estimation, implying that the proposed estimation is robust to a wide set of proposed parameters. In the subsequent experiments, we employ $q = 0.01, \bar{r} = 1, \hat{r} = 0.1, q^{\gamma} = 0.001, \bar{r}^{\gamma} = 0.1, \hat{r}^{\gamma} = 1$.

Tables 1–4 present the validation results for the process of estimating numbers of passengers boarding per run (Table 1), alighting rates per run (Table 2), numbers of passengers alighting per run (Table 3) and numbers of passengers on-board between stations (Table 4). The results refer to each direction, separately.

According to these tables, all five regular weekdays perform very similarly in terms of magnitudes of errors, showing that the differences they have in APC availability do not play a critical role in the performance of the proposed modelling. In addition, it is observed that the Helsinki-outbound direction is associated with lower errors than the Helsinki-inbound direction in terms of numbers of passengers boarding. The reverse occurs in the validation of the estimated alighting rates and numbers of passengers alighting, hence the Helsinki-outbound direction is associated with higher errors than the Helsinki-inbound direction. This is an intuitively expected behaviour of the model. The Helsinki-inbound direction is associated with greater certainty in demand alighting than the Helsinki-outbound direction, since a significant proportion of the passengers alight at the two last stations, Pasila and Helsinki terminal, in this direction. (Fig. 10). Similarly, the Helsinki-outbound direction is associated with greater certainty regarding boarding patterns than the Helsinki-inbound direction, since a significant proportion of the passengers board at the same two stations of Pasila and Helsinki in this direction (Fig. 9). It is noteworthy that the estimated numbers of passengers alighting (Table 3) is the component of the estimation process that is associated with the greatest MAPE errors, which are notably higher than the other components (i.e.

Table 2
Estimation errors for alighting rates.

Day	Helsinki-outbound			Helsinki-inbound		
	MAE (unitless)	RMSE (unitless)	MAPE (%)	MAE (unitless)	RMSE (unitless)	MAPE (%)
20.09.2021	0.07	0.10	32.86%	0.04	0.06	34.24%
21.09.2021	0.07	0.10	35.35%	0.04	0.06	35.35%
22.09.2021	0.06	0.09	30.73%	0.04	0.07	34.17%
23.09.2021	0.06	0.09	32.01%	0.04	0.06	32.63%
24.09.2021	0.06	0.09	33.94%	0.04	0.07	35.04%

Table 3
Estimation errors for alighting passengers.

Day	Helsinki-outbound			Helsinki-inbound		
	MAE (pax)	RMSE (pax)	MAPE (%)	MAE (pax)	RMSE (pax)	MAPE (%)
20.09.2021	11.88	15.00	89.53%	8.13	10.66	66.68%
21.09.2021	12.35	15.24	90.25%	8.69	11.26	68.25%
22.09.2021	12.63	15.55	87.76%	8.59	11.09	67.57%
23.09.2021	13.06	16.09	87.14%	8.62	11.12	64.97%
24.09.2021	12.83	15.50	85.78%	9.05	11.51	64.62%

Table 4
Estimation errors for on-board passengers.

Day	Helsinki-outbound			Helsinki-inbound		
	MAE (pax)	RMSE (pax)	MAPE (%)	MAE (pax)	RMSE (pax)	MAPE (%)
20.09.2021	24.65	29.45	44.22%	15.81	18.64	30.34%
21.09.2021	26.96	32.39	44.10%	17.11	20.15	33.42%
22.09.2021	25.09	30.04	41.71%	16.27	19.18	28.95%
23.09.2021	26.17	31.27	40.63%	17.22	20.11	32.80%
24.09.2021	25.24	30.37	37.94%	17.58	20.92	29.95%

a MAPE from approximately 65% to approximately 90% while MAE from about 8 to about 13 passengers considering both directions and all five weekdays). Considering that the numbers of passengers alighting depend directly on both the estimated alighting rate and the estimated number of passengers on-board, an accumulation of estimation inefficiencies is expected, deriving from the two separate estimation mechanisms.

The main output of the proposed framework refers to the estimated numbers of passengers on-board the vehicles, since they are later used for estimating comfort levels. According to Table 4, this component is estimated with a relatively high accuracy that ranges approximately 29-44%, considering the MAPE of both directions and all five weekdays. It is noteworthy that the Helsinki-outbound direction is associated with greater errors than the Helsinki-inbound direction, with the maximum MAPE value of the Helsinki-outbound direction being approximately 44% and of the Helsinki-inbound direction being approximately 33%.

5.2. Comfort levels

The estimated numbers of passengers on-board are used for estimating comfort levels using the framework proposed by Chandakas (2009) introduced in Section 3.5. Tables 5 and 6 present the comfort level estimation errors for each direction, respectively. The five weekdays of 20-24.09.2021 are analysed. It is noted that the number of observations per day that are reported in these tables depend on the number of runs and stations served per run, as well as on the APC availability per day. Results indicate that, similarly to the analysis for estimating passengers on-board, all days studied here present similar performance in each direction. Good estimation precision is noted in both directions, with the Helsinki-inbound direction being associated with particularly precise estimations in approximately 80% of observations. The maximum absolute error equals 2 in all cases studied here, while the average absolute error has a maximum value of 0.31.

In order to investigate the distribution of errors across different comfort levels, Fig. 12a is constructed. This figure demonstrates combined errors of the five weekdays from 20.09.2021 to 24.09.2021, in both directions. According to this figure, the real comfort levels are mostly low (i.e. there is overall good passenger comfort), ranging from 1 to 4 for the Helsinki-outbound direction and from 1 to 3 for the Helsinki-inbound direction. In the Helsinki-outbound direction, there are only three observations associated with comfort level 4, with the estimation being associated with an error of -1. This is intuitively expected, since the overall Helsinki public transport network is known for the good quality of service and the mild crowding phenomena on-board. In addition, during September 2021, COVID-19 effects would have been present in the demand patterns and the crowding on-board. In order to account for the

Table 5
Estimation errors for comfort levels in Helsinki-outbound direction.

Day	Precision (%)	Max Absolute Error	Average Absolute Error	No. observations
20.09.2021	72.08%	2	0.28	1440
21.09.2021	70.88%	2	0.29	1346
22.09.2021	69.46%	2	0.30	1231
23.09.2021	68.50%	2	0.31	1289
24.09.2021	71.80%	2	0.28	1305

Table 6
Estimation errors for comfort levels in Helsinki-inbound direction.

Day	Precision (%)	Max Absolute Error	Average Absolute Error	No. observations
20.09.2021	81.32%	2	0.19	1419
21.09.2021	79.85%	2	0.22	1305
22.09.2021	77.80%	2	0.24	1234
23.09.2021	76.60%	2	0.25	1269
24.09.2021	76.92%	2	0.25	1274

low demand levels, the same plots have been constructed assuming that vehicles operated with half their actual capacity (Fig. 12b). This change allowed the developed models to demonstrate their performance at higher levels of comfort. More specifically, in both directions, there are real comfort levels from 1 to 5.

It is noted that in both figures, the lower the comfort level, the greater the estimation precision in both directions. The Helsinki-outbound direction is associated with errors that refer mostly to underestimating the real comfort level. In contrast, the Helsinki-inbound direction presents a better balance between overestimating and underestimating the real comfort level. Depending on the purpose of requiring the information about comfort levels, each type of error in estimations can lead to inconveniences. Namely, informing passengers about better crowding conditions than in reality will likely create false hopes and lead to customer dissatisfaction. On the other hand, predicting more crowded conditions than the actual ones may not be as critical for travellers (who are likely to welcome an unexpected benefit), but could still lead operators to re-allocating resources from other critical activities (e.g. improving infrastructure) to dealing with a non-existing issue. Hence, a balanced level of overestimating and underestimating comfort errors is considered a safe approach for both operators and users. This balance is observed in the Helsinki-inbound direction, but not in the Helsinki-outbound direction in this study. Still, when the model is insufficiently precise, the estimation error is mostly equal to 1 in absolute value, indicating a good performance.

6. Sensitivity analysis

In this section, a series of sensitivity analysis experiments is performed in order to investigate how the performance of the proposed modelling framework changes as input values vary from those selected in the previous section. In all analyses, the data from the regular weekday of 23.09.2021 and the Helsinki-outbound direction are considered. The choice of this day is arbitrary, since all days were found to perform equivalently. The choice of this direction is made based on the fact that this direction was found to be more challenging in terms of estimating passengers on-board and therefore comfort levels.

6.1. Passengers split per line

The first sensitivity analysis for the proposed framework refers to investigating the effects of the time horizon h (i.e., the forward-looking time window for identifying alternative choices of lines for passengers at station i) and the number of historical days for station attractiveness in the ratio of passengers per line, $\theta^{s}_{i,l}(k)$. This ratio affects the estimation of the number of passengers boarding and, ultimately, passengers on-board and comfort levels. In this analysis the estimation framework considers five historical days for station arrival rates and passenger alighting rates.

Fig. 13 presents the errors in passengers boarding and on-board estimations. Scenarios considering four time horizons (i.e. 5, 10, 15, and 20 min) and up to five historical weekdays are investigated for estimating passenger volumes. These figures reveal that the time horizon of 5 min combined with one historical day presents the greatest errors, which improve as more historical days are considered. The best performance is met for time horizons of 15 and 20 min, with the addition of more historical days leading to insignificant improvements. Such findings are perceived to be case-specific, since they depend on aspects such as line frequency during a certain time period, the stations that are served after the current station, as well as the attractiveness of these stations, among others. Although the average service headway per station is found to be 18 min, headways as low as 1 minute are met in busy stations such as Helsinki, Pasila, and Leppävaara during peak hours. Therefore, the identification of the most appropriate time horizon size h should be investigated thoroughly for every new case study. Methods to pre-determine the optimal value of h for a case study could be a future extension of the present paper.

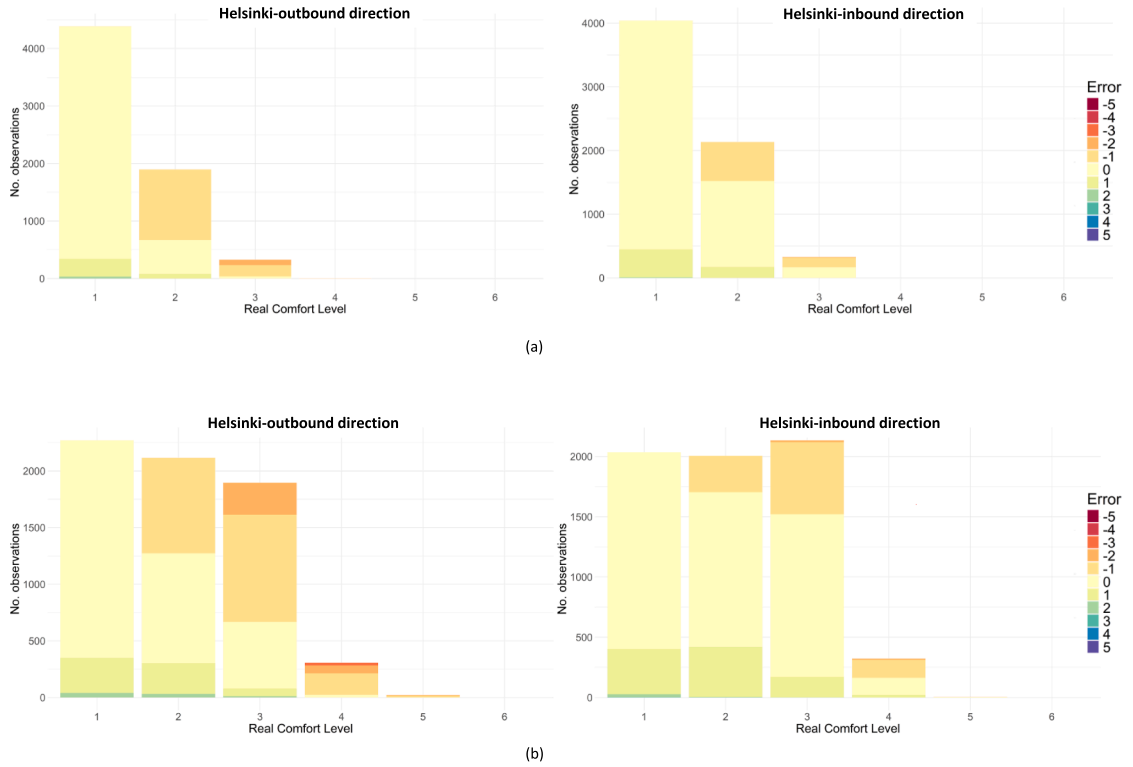

Fig. 12. Errors per comfort level in each direction for the case of a) full capacity, and b) half capacity (Negative values indicate underestimation of the actual comfort level, while positive values indicate overestimation).

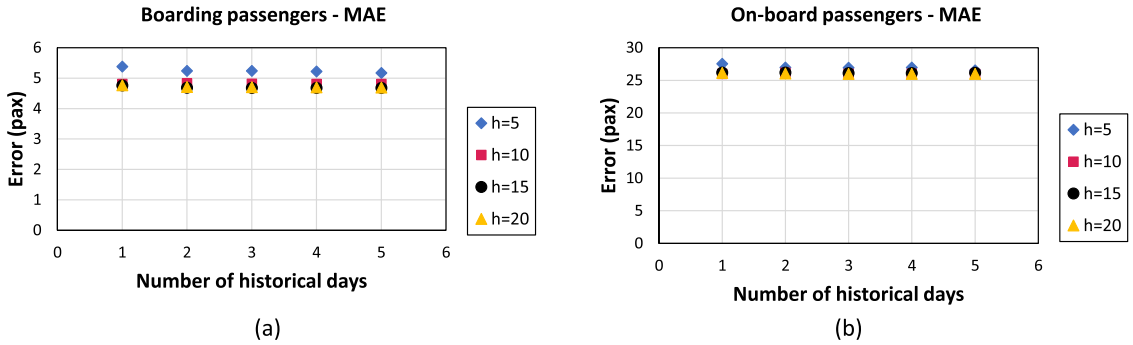

Fig. 13. Errors of estimated number of a) passengers boarding and b) passengers on-board resulting from different input values for the passenger split.

6.2. Number of historical days for arrival and alighting rates

The numbers of historical days for arrival and alighting rates are investigated in this Section, considering all other inputs as in Section 5. According to Fig. 14, not considering historical information for the arrival and alighting rates is a strategy associated with the greatest errors for on-board passengers' estimation. It is evident that historical information is needed for revealing the underlying demand patterns that are beneficial for estimating the number of passengers on-board. However, the addition of more than one historical days is not improving the performance of the proposed modelling framework significantly.

It is important to also analyse how the errors of the estimated numbers of passengers on-board are shaped per run and across the day. This is shown in Fig. 15, which includes the MAE results. In this analysis four time periods within the day are considered, namely: a) start of service - 10.00; b) 10.00 - 14.00; c) 14.00 - 18.00; and d) 18.00 - end of service. These time periods are selected based on the authors' personal experience on the network's demand levels during the day, as well as insights from the demand analysis presented in Figs. 9 and 10. The first and third time periods include the morning and evening peaks, respectively. In most cases, the consideration of historical days improves the performance of estimations per run. The afternoon peak exhibits greater errors

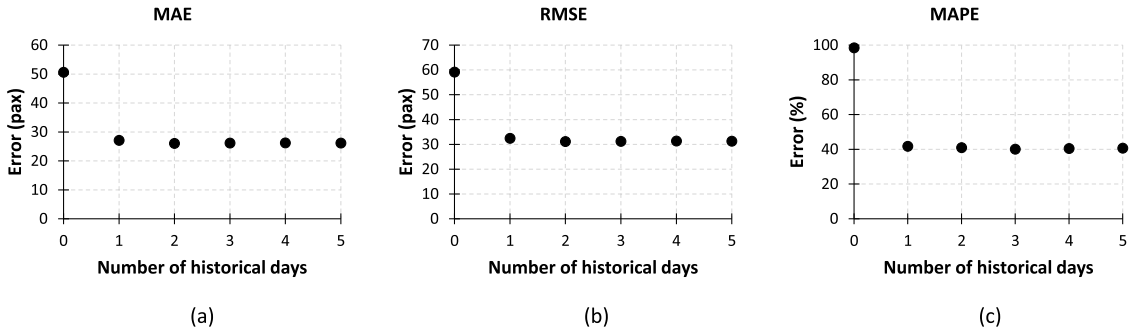

Fig. 14. Estimated number of passengers on-board errors of a) MAE, b) RMSE and c) MAPE.

Table 7
Estimation errors for comfort levels under different time intervals of passengers' arrival before train departure ($\bar{f}$).

Time interval (minutes)	Precision (%)	Max Absolute Error	Average Absolute Error
5	33.67	4	0.98
10	63.46	3	0.42
15	70.91	2	0.30
20	68.50	2	0.31
25	66.80	2	0.33
30	65.79	2	0.34

than other time periods. Lower errors are observed during the first period. Regarding the six lines that operate in the network (i.e., A,E,L,U,X, and Y), no consistent error pattern is observed. Line A, the most frequent line in the network (i.e., approximately 47% of daily runs), is the one associated with the greatest error in each time period when no historical days are considered. However, there is a significant improvement as long as at least one historical day is considered in all cases. Therefore, higher or lower levels of errors cannot be associated with a specific line when different numbers of historical days are considered, suggesting that there is no line-specific bias in the estimations.

Considering the great errors observed when no historical information is considered for arrival and alighting rates, this analysis is extended to observe the effect of historical information for arrival and alighting rates on the comfort level estimation. According to Fig. 16, historical information for passenger arrival and alighting rates improves the performance of the proposed models in terms of precision (i.e. zero errors). Not considering historical days, however, presents a better balance between over-estimations and underestimations, similar to how the Helsinki-inbound direction was evaluated in the previous section. This implies that the models with historical information are associated with lower levels of estimated demand compared to models with historical information. This is intuitively expected, since the historical information is smoothening the results and therefore some outliers of higher demand might not be properly captured by the modelling framework.

This observation motivated a further sensitivity analysis of the proposed models regarding the constant arrival time interval utilised to calculate historical arrival rates ($\bar{f}$). This constant indicates how long before the train's departure passengers of line l are assumed to arrive at station i. Table 7 presents the evaluation of the estimated comfort levels when arrival time intervals of 5, 10, 15, 20, 25 and 30 min are considered. This table shows that a time interval as low as 5 min significantly reduces the performance of the models to unacceptable levels. The time interval of 15 min is associated with the best performance, which is very close to the performance of the time interval selected in this study, i.e. 20 min. It is highlighted that the average service headway in this case study is 18 min, which is a measure that could be associated with average expected time of arrival at a station. These findings highlight the importance of carefully selecting the time interval of passengers' arrival at the stations. Methods to optimise this input value in advance could be an extension of the proposed modelling framework.

6.3. APC availability

In this study, the availability of APC data is notably high. As mentioned, the missing data are related to cases where technical errors did not allow the correct recording of passenger counts. The goal of this section is to investigate the performance of the models proposed here in the case of lower APC availability. The value of this insight is twofold, as it can both support the models' robustness, and offer some preliminary understanding on the role of APC availability on the performance of advanced estimation methods. For this purpose, experiments were conducted, during which run ids with available APC measurements were randomly removed from the dataset, with the respective runs treated as runs without APC availability. Nine cases of 10% to 90% APC availability with a step of 10% were considered. These levels refer to the percentage of the available APC information that is considered in the respective analysis. Therefore, if for a day 80% of all APC data are available due to technical errors, then reducing the available APC records

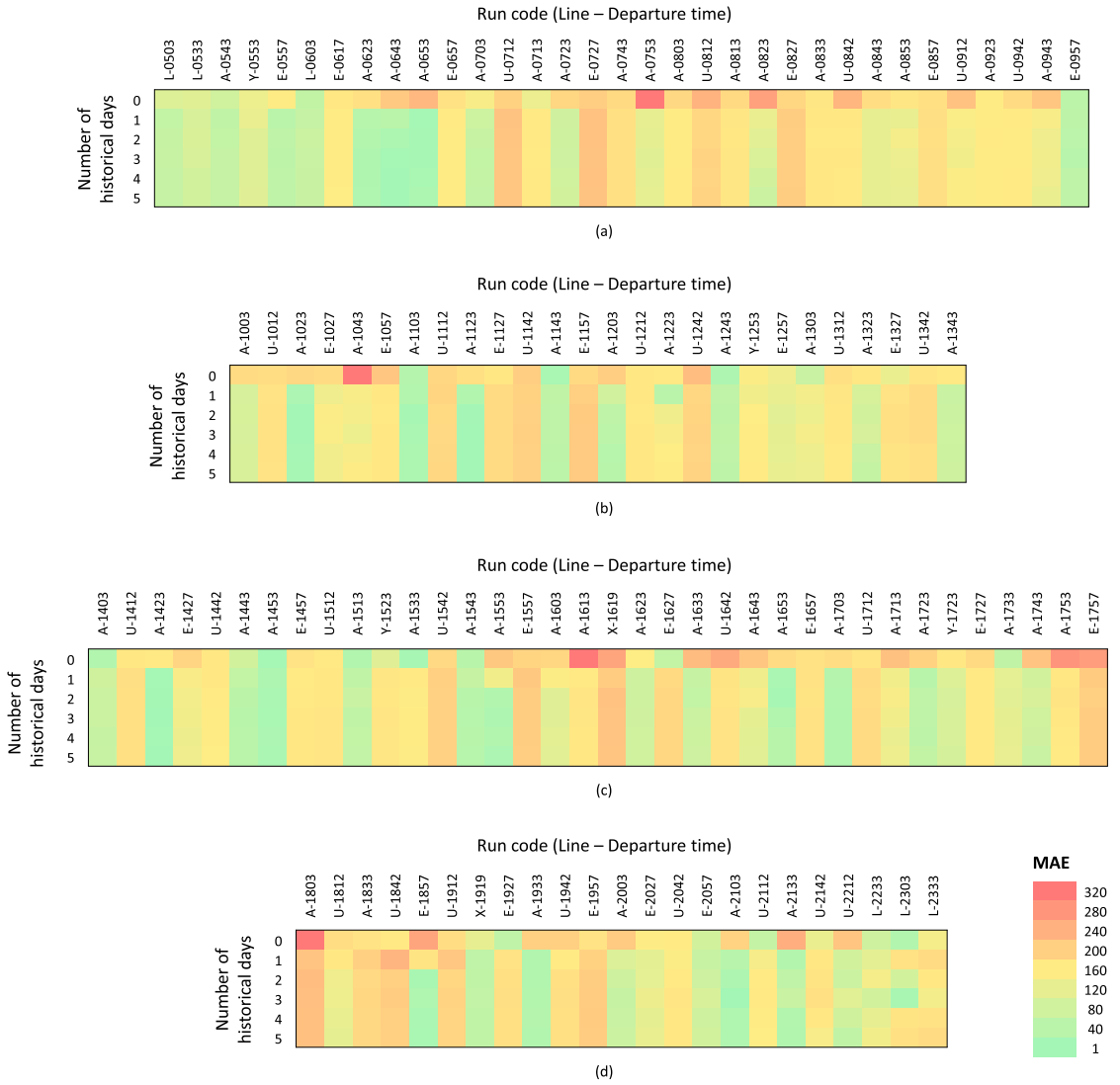

Fig. 15. MAE for estimated number of passengers on-board per run for the time periods of a) start of service - 10.00, b) 10.00 - 14.00, c) 14.00 - 18.00, and d) 18.00 - end of service.

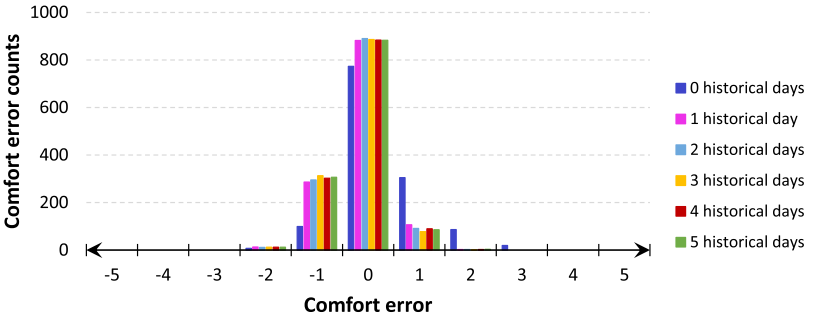

Fig. 16. Errors of estimated comfort levels (Negative values indicate underestimation of the actual comfort level, while positive values indicate overestimation).

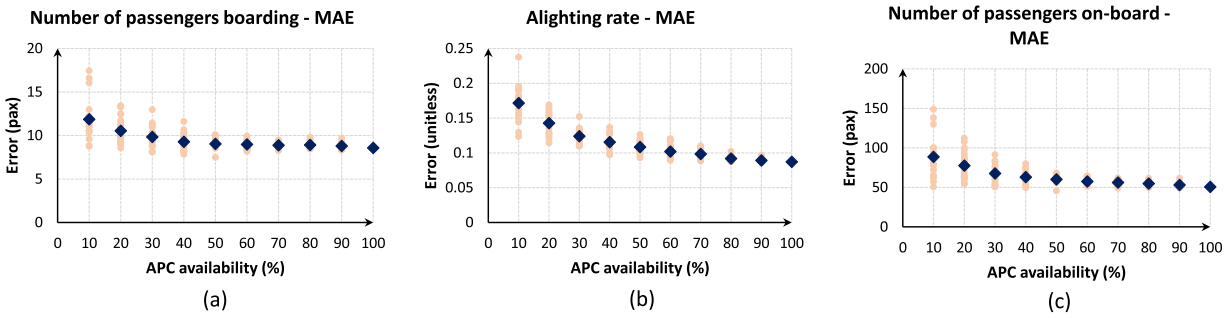

Fig. 17. MAE for various rates of APC availability for 23.09.2021 in direction 1 without considering historical days.

Table 8
MAE results for 10% APC availability in the cases of considering zero and five historical days for arrival and alighting rates.

	0 historical days		5 historical days	
	Mean	Sd	Mean	Sd
Number of passengers boarding	11.87	2.35	6.70	0.31
Alighting rates	0.17	0.03	0.12	0.01
Number of passengers on-board	88.63	25.48	38.61	2.82
Comfort level precision (%)	49.29	4.39	63.10	2.06
Max absolute comfort error	4.90	0.31	2.75	0.72
Average absolute comfort error	0.82	0.17	0.41	0.03

to 80% means that an APC availability of 64% will be obtained. For each level of APC availability, 20 different sets of runs were considered to be removed, therefore, leading to 20 different sets of runs with available APC records. For each set the proposed methodology was implemented. All other inputs were as in Section 5. Historical information for arrival and alighting rates was not considered, since this was found to be the worst case in terms of modelling performance of passengers on-board in previous section.

The results for MAE are shown in Fig. 17. In the day studied here, the APC availability was 89%, hence, 89% of runs were associated with measured APC records. The red dots represent the MAE for each set of runs with APC availability. The blue diamond-shaped dots represent the mean value of MAE for each level of APC availability. As shown in Fig. 17, the lower the level of APC availability, the greater the range of errors associated with the performance of the proposed models. The range of errors is noteworthy for 10% APC availability (i.e. 8.9% of all runs), with MAE ranging from 9 to 18 passengers. This highlights the significance of selecting very carefully which runs to equip with APC devices, since different sets of runs with available APC lead to different model performances when the APC availability is low. The range of errors is almost zero when APC availability is close to 90% of the available APC data (i.e. 80.10% of all runs). It is noted that the errors of alighting rates follow the shape of a curve, with the errors improving continuously as the APC availability increases. Considering the number of passengers boarding and on-board, it is observed that after 60% of APC availability (i.e. 53.4% of all runs) the mean values of errors (blue diamond-shaped dots) follow an almost flat line, hence implying an equivalent average performance. This observation implies that, at least for accurately predicting comfort, there is no need for equipping all vehicles with APC devices, since advanced prediction methods like the one proposed here can perform equally well with lower levels. Low levels of APC availability are associated with greater estimation uncertainties when no historical days are considered for arrival and alighting rates.

According to Table 8, when five historical days are considered, the uncertainty, as expressed through standard deviation (sd), is significantly reduced for the worst case of 10% APC availability considered here. Focusing on the comfort level estimation, it is noteworthy that on average more than half of comfort levels are precisely estimated with the proposed framework at a low APC availability of 10% in a complex network of multiple lines operating per station. However, the maximum comfort error is rather high. The consideration of five historical days increases the precision to more than 60%, while also reducing the maximum comfort error.

7. Conclusions

7.1. Summary of findings

Accurate information of comfort levels on-board public transport vehicles is of critical importance for public transport operators and users. Kalman filtering techniques are found valuable for providing demand predictions, but existing literature is limited in filling gaps in APC datasets in the case of simple networks where only one line serves each station. The current study developed a modelling framework to estimate demand information missing from APC datasets in the case of complex public transport networks, in which stations are served by multiple lines.

The proposed framework was validated considering five regular weekdays and the two directions of movement within the case study network of Helsinki. The results highlighted the models' robust performance by showing that it is equivalent among the analysed days, which had similar characteristics in terms of operation and demand. The two analysed directions are associated with different demand patterns, which are also reflected through different performance in estimating passengers on-board and the comfort levels they experience due to crowding. The proposed models offer mostly accurate estimations of comfort levels, whether full or half capacity is considered in the validation framework. The Helsinki-inbound direction was found to offer balanced estimation errors between over-estimation and underestimation, while the Helsinki-outbound direction was associated with greater underestimation.

Sensitivity analyses for different components of the modelling framework were also performed. The respective results led to invaluable insights on the underlying mechanisms of the proposed models. The input values chosen for estimating the ratio parameter θ, were found to require time horizons greater than 10 min, in order to capture "competitive" runs and lines at each station. The consideration of more than one day of historical information regarding the attractiveness of stations was not found significant in improving the performance of the proposed models in terms of estimating numbers of passengers on-board in the current case study. These results are considered case-specific and depend on the local operational characteristics of the network, among others.

The second sensitivity analysis performed here showed that incorporating historical information of passengers' arrival rates at the stations and alighting rates per vehicle for at least one day was found to improve the models' performance. The addition of more than one historical days led to a steady performance, almost equivalent to having one historical day. The models were found to be unbiased in terms of quality of estimations per line, while the afternoon peak runs were associated with greater errors. It was interesting to observe that in the Helsinki-outbound direction, including less historical days was associated with a better balance in terms of underestimations and over-estimations, hence, addressing the concerns that were raised when the results of models with five historical days revealed this imbalance in the first place. These findings suggest that the underlying mechanism of including historical days, which leads to more smoothed estimations, is likely not sufficiently able to capture peaks of high demand if they appear as outliers. A sensitivity analysis of the constant time interval for calculating historical arrival rates revealed that they can significantly affect the models' performance.

The last sensitivity analysis focused on challenging the proposed models by considering lower levels of APC availability. The performed experiments showed that there is no need of having vehicles fully equipped with costly APC devices, since lower levels of APC availability might lead to estimations equally efficient as with higher levels of APC availability. However, the uncertainty in estimations quality is greater for lower APC availability. The value of adding historical information for passengers' arrival and alighting rates can reduce this uncertainty significantly. A significantly low percentage of APC data availability can accurately estimate more than 50% of comfort levels without historical information, while this performance can increase by 10% when historical days are considered.

To summarize, the proposed framework's performance is influenced more by low APC penetration rates and rapidly deviating outliers than by overall demand levels. Experiments with reduced vehicle capacities confirm its robustness under higher crowding. The Kalman filter's smoothing effect improves stability for slowly varying processes but reduces responsiveness to unexpected extremes. Low APC penetration rates can be mitigated effectively by incorporating historical information. Regarding the role of historical data in the proposed framework, including just one historical day of information was sufficient to improve performance substantially. It is noteworthy that additional days yielded minimal benefit, indicating that transport operators do not need to maintain long historical datasets for robust estimations.

7.2. Implications for practice

Information about the number of passengers on-board public transport vehicles is crucial for transport operators and managers that focus on achieving the highest possible performance of their networks. Network complexity varies according to several factors, such as for example, number of stations, number of lines serving each station, scheduled headways per station, operational delays in schedules, irregular demand patterns, and many others. Therefore, models developed for estimating on-board passengers numbers and comfort levels need to account for all these factors, while ensuring high performance in reasonable computational times. In addition, transport managers and operators require flexible models that allow them to adjust critical input values in order not only to calibrate the models based on their network, but also serve better the specific purposes of the models' implementation.

The modelling framework proposed in this study could be applied for a variety of purposes within complex public transport networks in which each station serves multiple lines. In this context, it is important to ensure at least intermittent APC coverage across all lines, as the complete absence of information for a line would limit the ability to infer passenger dynamics. It is noted that such case corresponds to a structural absence of information for an interconnected subset of stations and runs rather than a specific limitation of the proposed methodology. An obvious application relates to offering real-time information to passengers regarding crowding levels on-board (e.g. through smartphones). Although the framework is associated with good levels of precision, it is important to account for the estimation errors that were presented and analysed thoroughly. More specifically, transport authorities should adjust the input values in order to achieve an estimation precision that will be considered satisfactory, also checking the maximum errors and balancing underestimations and over-estimations according to their needs and goals. Offering information about expected crowding levels lower than in reality can lead to user inconvenience and dissatisfaction. Information about travel conditions that are worse than in reality can lead travellers to choose other modes, leading to a series of negative outputs for the operators at a multimodal level (e.g. unexpected overcrowding on other modes and reputation damage).

The proposed framework could also be utilised by transport authorities for self-evaluation purposes, in order to proceed with improvements in their services. For example, high crowding levels at certain stations during certain time periods could imply the

need for allocating more resources to these stations during these time periods, e.g. by increasing the service frequency or the size of service vehicles. Similarly to the real-time information application, transport stakeholders should adjust the models' inputs according to their goals and needs by also considering the implications of their choices. In the area of self-evaluation, estimating lower crowding levels than in reality can lead to omitting important changes that need to be made at the network, while estimating higher crowding levels than in reality can lead to unnecessary allocation of resources to parts of the network that may not require improvement.

The proposed modelling framework can serve as a tool also in the case of making important decisions regarding acquiring and installing APC devices. As already discussed, APC technology is invaluable for revealing demand-related information but can often be associated with high costs, which commonly leads to low penetration rate of such devices within public transport networks. The fact that this is not the case in Helsinki allowed this research to perform an investigation of how much APC availability is needed when implementing in practice such estimation frameworks. It was interesting to find that with relatively low APC availability it is possible to achieve relatively good performance of estimations. As the analysis revealed, not all runs are equivalent in terms of how they contribute to the models' performance, if equipped with APC. This is a significant finding that highlights the importance of strategically selecting which runs to equip with such devices, at least in the case of complex networks with varying demand patterns and non-uniformly distributed vehicle dispatches across stations during the day. Therefore, practitioners that are interested in investing in APC technology can utilise this framework for investigating how many and which runs should be equipped with APC devices in their network.

7.3. Future directions

The current study will be extended in the future in different directions. The factor that allows to split passengers per line after they enter a station depends on local operational and demand-related characteristics. Hence, a future direction will focus on developing a method that will allow a generalised selection of this important factor based on the specific network that is studied and its unique characteristics (e.g. headways, number of lines, APC availability). A similar direction refers to optimising the number of historical days that need to be considered, by also optimising the time interval that should be considered when calculating historical arrival rates. Such approaches might imply the need for accounting for different runs, lines, and stations separately, rather than implementing a uniform time interval. Another direction refers to investigating the development of enhanced models with varying assumptions per time period, as the sensitivity analysis here showed lower performance during the afternoon peak demand. Such direction could include calibrating input values separately for each time period, accounting for different schedules and demand patterns. This study can also be extended to account for networks experiencing left-behind passenger phenomena due to on-board overcrowding. Finally, a future extension of this study refers to exploring methods that will incorporate the framework developed in order to determine strategies that will allow the selection of runs that should be equipped with APC devices to achieve high quality predictions at the lowest possible operational cost.

CRediT authorship contribution statement

Charalampos Sipetas: Writing – original draft, Methodology, Formal analysis, Data curation, Conceptualization; **Claudio Roncoli:** Writing – original draft, Validation, Supervision, Methodology, Formal analysis, Conceptualization; **Ektoras Chandakas:** Writing – review & editing, Supervision, Conceptualization; **Ioannis Kaparias:** Writing – review & editing, Supervision, Conceptualization.

Data availability

The authors do not have permission to share data.

Declaration of competing interest

The authors do not have any conflicts of interest to declare.

Acknowledgements

The authors thank HSL for access to the APC dataset, and for their time and discussions about this study. The work of C. Sipetas and C. Roncoli was supported by the EU Horizon Europe project ACUMEN (no. 101103808).

Appendix A. Notation glossary

This appendix provides a comprehensive glossary of the notation used throughout the paper, summarized in Table A.1.

Table A.1
Glossary of notation.

Symbol	Description
x	State vector of the system
$\hat{x}^{-}(k)$	A-priori (predicted) estimate of the state at time step k
$\hat{x}(k)$	A-posteriori (updated) estimate of the state at time step k
$A(k)$	State-transition matrix at time step k
$P^{-}(k)$	A-priori error covariance at time step k
$P^{+}(k)$	A-posteriori error covariance at time step k
Q	Process noise covariance matrix
$K(k)$	Kalman gain at time step k
$C(k)$	Observation (measurement) matrix at time step k
R	Measurement noise covariance matrix
$\bar{z}(k)$	Measurement of the system at time step k
I_d	Identity matrix of proper size
μ	Initial mean of state estimate
H	Initial error covariance matrix
$\hat{b}_j^{\mathrm{r}}(k)$	Estimated number of boarding passengers for run j at time step k
$\hat{a}_j^{\mathrm{r}}(k)$	Estimated number of alighting passengers for run j at time step k
$\hat{p}_j(k)$	Estimated number of passengers on-board run j at time step k
$\hat{\gamma}_j^{\mathrm{r}}(k)$	Estimated alighting rate for run j at time step k
$i \in I$	Index of stations, with I the set of all stations
$j \in J$	Index of runs, with J the set of all runs
$l \in L$	Index of lines, with L the set of all lines
T	Discrete time step size (in seconds)
$t = kT$	Continuous time from the beginning of the "operational" day (in seconds)
$\bar{\omega}$	Measured value of variable ω
$\tilde{\omega}$	Historical value of variable ω
$\hat{\omega}$	Estimated value of variable ω
$\eta_{i,j}^{\mathrm{r}}(k)$	Binary: 1 if run j departs from station i during $(k-1,k]$, else 0
$\bar{b}_j^{\mathrm{r}}(k)$	Measured boardings for run j at time step k
$\bar{a}_j^{\mathrm{r}}(k)$	Measured alightings for run j at time step k
$\bar{p}_j(k)$	Measured passengers on-board run j at time step k
β_j^{r}	Binary: 1 if run j has APC data, else 0
$w_i(k)$	Number of waiting passengers at station i at time step k
$e_i(k)$	Number of entering passengers at station i at time step k
$\xi_i(k)$	Process noise at station i (e.g., Gaussian)
$\psi_i(k)$	Measurement noise at station i (e.g., Gaussian)
$\gamma_i(k)$	Alighting rate at station i at time step k
$\xi_i^{\gamma}(k)$	Process noise for alighting rate at station i
$\psi_i^{\gamma}(k)$	Measurement noise for alighting rate at station i
$w_i^{\mathrm{ml}}(k)$	Waiting passengers at station i (multi-line case)
$e_i^{\mathrm{ml}}(k)$	Entering passengers at station i (multi-line case)
$\eta_{i,l}^{\mathrm{ml}}(k)$	Binary variable: 1 if a vehicle of line l departs from station i during $(k-1,k]$
$\theta_{i,l}(k)$	Fraction of passengers at station i boarding line l at time step k
$\mu_{i,m,l}(k)$	Binary variable: 1 if first run of line l from station i reaches station m before other lines
$\hat{a}_m^{\mathrm{q}}(k)$	Attractiveness of station m, based on expected alightings
h	Forward-looking time horizon for passenger choices
$\gamma_{i,l}^{\mathrm{ml}}(k)$	Alighting rate at station i for line l at time step k (multi-line case)
$\xi_{i,l}^{\gamma,\mathrm{ml}}(k)$	Process noise for alighting rate in multi-line case
$\psi_{i,l}^{\gamma,\mathrm{ml}}(k)$	Measurement noise for alighting rate in multi-line case
$t_{i,d}^{\mathrm{d}}(k)$	Time since the last vehicle stopped at station i on day d (used in historical calculations)
$f_{i,l,d}^{\mathrm{d}}(k)$	Normalizing factor used in the historical boarding aggregation for station i, line l, day d and time step k. One choice is the time difference between consecutive departures; alternatively it may be set constant as $\bar{f}$.
$\bar{f}$	A constant choice for $f_{i,l,d}^{\mathrm{d}}(k)$, representing the assumed arrival window prior to departure for boarding passengers.

References

Abewickrema, W., Yildirimoglu, M., Kim, J., 2023. Multivariate time-varying kalman filter approach for cycle-based maximum queue length estimation. Transp. Res. Part C Emerging Technol. 154, 104238.

Anderson, B. D.O., Moore, J.B., 1979. Optimal Filtering. Prentice-Hall.

Antsaklis, P.J., Michel, A.N., 2006. Linear Systems. Birkhäuser Boston, Boston, MA.

Bai, L., Ireson, N., Mazumdar, S., Ciravegna, F., 2017. Lessons learned using Wi-Fi and bluetooth as means to monitor public service usage. In: Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers, pp. 432–440.

Basnak, P., Giesen, R., Muñoz, J.C., 2022. Estimation of crowding factors for public transport during the COVID-19 pandemic in Santiago, Chile. Transp. Res. Part A Policy Pract. 159, 140–156.

Bekiaris-Liberis, N., Roncoli, C., Papageorgiou, M., 2016. Highway traffic state estimation with mixed connected and conventional vehicles. IEEE Trans. Intell. Transp. Syst. 17 (12), 3484–3497.

Bekiaris-Liberis, N., Roncoli, C., Papageorgiou, M., 2017. Highway traffic state estimation per lane in the presence of connected vehicles. Transp. Res. Part B Methodol. 106, 1–28.

Board, U.S. T.R., 2013. Transit Capacity and Quality of Service Manual, Third Edition. The National Academies Press, Washington, D.C.

Bouscasse, H., de Lapparent, M., 2019. Perceived comfort and values of travel time savings in the Rhône-Alpes region. Transp. Res. Part A Policy Pract. 124, 370–387.

Boyle, D., 1998. TCRP Synthesis of Transit Practice 29: Passenger Counting Technologies and Procedures. Transportation Research Board of the National Academies, Washington, D.C.

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 Emerging Technol. 79, 274–289.

Buehler, R., Pucher, J., 2012. Demand for public transport in germany and the USA: an analysis of rider characteristics. Transp. Rev. 32 (5), 541–567.

Chandakas, E., 2009. La capacité des transports ferroviaires d'Île-de-France face à la hausse du trafic à long-terme. Master thesis, Ecole Nationale des Ponts et Chaussées, France.

Cho, S.-H., Park, H.-C., Choo, S., Park, S.H., 2024. How crowding impedance affected travellers on public transport in the COVID-19 pandemic. Transp. Res. Part F Psychol. Behav. 100, 69–83.

da Silva, P. F.P., Mendes, J., 2020. Passengers comfort perception and demands on railway vehicles: a review. KnE Eng. 2020, 257–270.

Dib, A., Cherrier, N., Graive, M., Rérolle, B., Schmitt, E., 2023. Unified occupancy on a public transport network through combination of AFC and APC data. In: 2023 IEEE 26th International Conference on Intelligent Transportation Systems (ITSC), pp. 1963–1970.

Ellenberger, D., Siebert, M., 2025. Artificial intelligence in automatic passenger counting: cost-efficient validation using the partitioned equivalence test. Transportmetrica A: Transp. Sci. 21 (3), 2267702.

Fedujwar, R., Agarwal, A., 2024. A systematic review on crowding valuation in public transport. Public Transp. 16 (3), 743–773.

Freemark, Y. Y.S., 2013. Assessing Journey Time Impacts of Disruptions on London's Piccadilly Line. Ph.D. thesis. Massachusetts Institute of Technology.

Gonzales, E.J., Keklikoglou, A., Sipetas, C., et al., 2018. Measuring left-behinds on subway. Massachusetts Department of Transportation.

Göransson, J., Andersson, H., 2023. Factors that make public transport systems attractive: a review of travel preferences and travel mode choices. Eur. Transp. Res. Rev. 15 (1), 32.

Grgurević, I., Juršić, K., Rajič, V., 2021. Review of automatic passenger counting systems in public urban transport. In: 5th EAI International Conference on Management of Manufacturing Systems. Springer, pp. 1–15.

He, H., Yan, M., Sun, C., Peng, J., Li, M., Jia, H., 2018. Predictive air-conditioner control for electric buses with passenger amount variation forecast. Appl. Energy 227, 249–261.

Heydenrijk-Ottens, L., Degeler, V., Luo, D., van Oort, N., van Lint, J., 2018. Supervised learning: predicting passenger load in public transport. In: CASPT Conference on Advanced Systems in Public Transport and TransitData, pp. 30–32.

Hsu, Y.-W., Chen, Y.-W., Perng, J.-W., 2020. Estimation of the number of passengers in a bus using deep learning. Sensors 20 (8), 2178.

Hu, R., Chiu, Y.-C., Hsieh, C.-W., 2020. Crowding prediction on mass rapid transit systems using a weighted bidirectional recurrent neural network. IET Intell. Transp. Syst. 14 (3), 196–203.

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), pp. 2007–2016.

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

Jenelius, E., 2019a. Data-driven bus crowding prediction based on real-time passenger counts and vehicle locations. In: 6th International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS2019), 5-7 June 2019, Kraków, Poland.

Jenelius, E., 2019b. Data-driven metro train crowding prediction based on real-time load data. IEEE Trans. Intell. Transp. Syst. 21 (6), 2254–2265.

Jung, Y.-J., Casello, J.M., 2020. Assessment of the transit ridership prediction errors using AVL/APC data. Transportation 47 (6), 2731–2755.

Kalman, R.E., Bucy, R.S., 1961. New results in linear filtering and prediction theory. Trans. ASME Ser. D, J. Basic Eng. 83, 95–108

Karatsoli, M., Nathanail, E., Basbas, S., Cats, O., 2024. Crowdedness information and travel decisions of pedestrians and public transport users in the COVID-19 era: a stated preference analysis. Cities 149, 104973.

Khomchuk, P., Tuladhar, S.R., Sivananthan, S., 2018. Predicting passenger loading level on a train car: a Bayesian approach. arXiv preprint arXiv:1808.06962.

Kimpel, T.J., Strathman, J.G., Griffin, D., Callas, S., Gerhart, R.L., 2003. Automatic passenger counter evaluation: implications for national transit database reporting. Transp. Res. Rec. 1835 (1), 93–100.

Liu, T., Ma, Z., Koutsopoulos, H.N., 2021. Unplanned disruption analysis in urban railway systems using smart card data. Urban Rail Transit 7 (3), 177–190.

Liu, Y.-Y., Slotine, J.-J., Barabási, A.-L., 2013. Observability of complex systems. Proc. Natl. Acad. Sci. 110 (7), 2460–2465.

Ma, X., Wu, Y.-J., Wang, Y., Chen, F., Liu, J., 2013. Mining smart card data for transit riders' travel patterns. Transp. Res. Part C Emerging Technol. 36, 1–12.

Ma, Z., Koutsopoulos, H.N., Chen, Y., Wilson, N. H.M., 2019. Estimation of denied boarding in urban rail systems: alternative formulations and comparative analysis. Transp. Res. Rec. 2673 (11), 771–778.

Maternini, G., Cadei, M., 2014. A comfort scale for standing bus passengers in relation to certain road characteristics. Transp. Lett. 6 (3), 136–141.

Mehmood, U., Moser, I., Jayaraman, P.P., Banerjee, A., 2019. Occupancy estimation using WiFi: a case study for counting passengers on busses. In: 2019 IEEE 5th World Forum on Internet of Things (WF-IoT), pp. 165–170.

Mikkelsen, L., Buchakchiev, R., Madsen, T., Schwefel, H.P., 2016. Public transport occupancy estimation using WLAN probing. In: 2016 8th IEEE International Workshop on Resilient Networks Design and Modeling (RNDM), pp. 302–308.

Mohammed, M., Oke, J., 2023. Origin-destination inference in public transportation systems: a comprehensive review. Int. J. Transp. Sci. Technol. 12 (1), 315–328.

Munizaga, M., Devillaine, F., Navarrete, C., Silva, D., 2014. Validating travel behavior estimated from smartcard data. Transp. Res. Part C Emerging Technol. 44, 70–79.

Olivo, A., Maternini, G., Barabino, B., 2019. Empirical study on the accuracy and precision of automatic passenger counting in European bus services. Open Transp. J. 13 (1), 250–260.

Pasini, K., Khouadjia, M., Ganansia, F., Oukhellou, L., 2019. Forecasting passenger load in a transit network using data driven models. In: WCRR 2019, 12th World Congress on Railway Research.

Pelletier, M.-P., Trépanier, M., Morency, C., 2011. Smart card data use in public transit: a literature review. Transp. Res. Part C Emerging Technol. 19 (4), 557–568.

Popuri, Y., Proussaloglou, K., Ayvalik, C., Koppelman, F., Lee, A., 2011. Importance of traveler attitudes in the choice of public transportation to work: findings from the regional transportation authority attitudinal survey. Transportation 38 (4), 643–661.

Pu, Z., Zhu, M., Li, W., Cui, Z., Guo, X., Wang, Y., 2020. Monitoring public transit ridership flow by passively sensing wi-fi and bluetooth mobile devices. IEEE Internet Things J. 8 (1), 474–486.
Roncoli, C., Chandakas, E., Kaparias, I., 2023a. Estimating on-board passenger comfort in public transport vehicles using incomplete automatic passenger counting data. Transp. Res. Part C Emerging Technol. 146, 103963.
Roncoli, C., Chandakas, E., Kaparias, I., 2023b. A kalman filter based estimation method for on-board public transport passenger comfort using incomplete automatic passenger counting data. In: Proceedings of the 102nd Annual Meeting of the Transportation Research Board. Washington, D.C.
Ryu, S., Park, B.B., El-Tawab, S., 2020. Wifi sensing system for monitoring public transportation ridership: a case study. KSCE J. Civil Eng. 24 (10), 3092–3104.
Saeedmanesh, M., Kouvelas, A., Geroliminis, N., 2021. An extended kalman filter approach for real-time state estimation in multi-region MFD urban networks. Transp. Res. Part C Emerging Technol. 132, 103384.
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.
Shao, M., Xie, C., Li, T., Sun, L., 2022. Influence of in-vehicle crowding on passenger travel time value: insights from bus transit in Shanghai, China. Int. J. Transp. Sci. Technol. 11 (4), 665–677.
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.
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 Emerging Technol. 118, 102727.
Tirachini, A., Hensher, D.A., Rose, J.M., 2013. Crowding in public transport systems: effects on users, operation and implications for the estimation of demand. Transp. Res. Part A Policy Pract. 53, 36–52.
Tirachini, A., Hurtubia, R., Dekker, T., Daziano, R.A., 2017. Estimation of crowding discomfort in public transport: results from Santiago de Chile. Transp. Res. Part A Policy Pract. 103, 311–326.
Trinh, X.-S., Ngoduy, D., Keyvan-Ekbatani, M., Robertson, B., 2022. Incremental unscented kalman filter for real-time traffic estimation on motorways using multi-source data. Transportmetrica A Transp. Sci. 18 (3), 1127–1153.
Trozzi, V., Gentile, G., Bell, M., Kaparias, I., 2013. Dynamic user equilibrium in public transport networks with passenger congestion and hyperpaths. Procedia Soc. Behav. Sci. 80, 427–454.
Trozzi, V., Gentile, G., Kaparias, I., Bell, M. G.H., 2015. Effects of countdown displays in public transport route choice under severe overcrowding. Netw. Spatial Econ. 15, 823–842.
Vieira, T., Almeida, P., Meireles, M., Ribeiro, R., 2020. Public transport occupancy estimation using WLAN probing and mathematical modeling. Transp. Res. Procedia 48, 3299–3309.
Wang, P., Chen, X., Chen, J., Hua, M., Pu, Z., 2021. A two-stage method for bus passenger load prediction using automatic passenger counting data. IET Intell. Transp. Syst. 15 (2), 248–260.
Wang, Y., Papageorgiou, M., 2005. Real-time freeway traffic state estimation based on extended kalman filter: a general approach. Transp. Res. Part B Methodol. 39 (2), 141–167.
Yap, M., Wong, H., Cats, O., 2025. Public transport crowding valuation in a post-pandemic era. Transportation 52 (1), 287–306.
Zhang, J., Shen, D., Tu, L., Zhang, F., Xu, C., Wang, Y., Tian, C., Li, X., Huang, B., Li, Z., 2017. A real-time passenger flow estimation and prediction method for urban bus transit systems. IEEE Trans. Intell. Transp. Syst. 18 (11), 3168–3178.
Zhu, Y., Koutsopoulos, H.N., Wilson, N. H.M., 2017. Inferring left behind passengers in congested metro systems from automated data. Transp. Res. Procedia 23, 362–379.