Estimating on-board crowding in complex public transport networks from incomplete automatic passenger counts

C. Sipetas[a,*], C. Roncoli[a], E. Chandakas[b] and I. Kaparias[c]

[a] Department of Built Environment, Aalto University, Espoo, Finland
charalampos.sipetas@aalto.fi, claudio.roncoli@aalto.fi
[b] Laboratoire Ville Mobilité Transport, École des Ponts ParisTech, Paris, France
ektoras.chandakas@enpc.fr
[c] Transportation Research Group, University of Southampton, Southampton, United Kingdom
i.kaparias@southampton.ac.uk
[*] Corresponding author

Extended abstract submitted for presentation at the Conference in Emerging Technologies in Transportation Systems (TRC-30)
September 02-03, 2024, Crete, Greece

April 24, 2024

Keywords: public transport; passengers comfort; Automatic Passenger Counting; Kalman filter

1 INTRODUCTION

On-board comfort is a strong determinant of whether travellers will choose a certain transport mode or not for their commute. Moreover, knowing the number of passengers on-board public transport vehicles is a piece of invaluable information for the efficient management and operation of the network. A data source commonly used is Automatic Passenger Counting (APC) data, but high costs of installation and maintenance lead to low numbers of vehicles being equipped with APC devices. Technical issues might also lead to gaps in the provided datasets, which makes the development of accurate and precise real-time predictions a particularly challenging task. This study extends the work of Roncoli *et al.* (2023a,b), which proposes a method based on Kalman filtering for estimating vehicle crowding levels, by making it applicable for complex networks in which a station serves multiple lines. It also extends the work of Sipetas *et al.* (2024) by incorporating historical information in the estimation and investigating their role in the estimation performance. The effects of having different levels of APC availability in a network are also investigated here. The case study is the west wing of the Helsinki commuter train network.

2 METHODOLOGY

We design an estimation framework capable of fusing information on the vehicle schedule with APC measurements (i.e., boarding and alighting passenger counts) from a subset of vehicles, to derive an estimate of the passengers on-board. Dynamic models are proposed for estimating, separately, boarding and alighting passengers, which are then employed to estimate the number of passengers on-board and the comfort level (Chandakas, 2009).

We first present dynamical models that capture the dynamics of boarding and alighting passengers. We assume that all passengers waiting on a platform, $w_i(k)$, will board any vehicle that departs from the current station i during $(k-1,\ k]$ if it arrives first at their destination station, as follows:

$$b_i^s(k) = \sum_{l \in L} [\eta_{i,l}^s(k) \cdot \theta_{i,l}^s(k)] \cdot w_i(k) \tag{1}$$

where $b_i^s(k)$ is the number of passengers boarding any train leaving station i at time $t = kT$, $\eta_{i,l}^s(k)$ equals 1 if line l departs from station i and 0 otherwise, and $\theta_{i,l}^s(k)$ represents the ratio of all passengers entering station i that will board line l during time $t = kT$, which is calculated as follows:

$$\theta_i^s(k) = \frac{\sum_{m \in S_{i,l}} \tilde{a}_m^q(k) \cdot \mu_{m,l}^H(k)}{\sum_{l \in L} \sum_{m \in S_{i,l}} \tilde{a}_m^q(k) \cdot \mu_{m,l}^H(k)} \tag{2}$$

where $S_{i,l}$ is the set of stations that each line serves after station i, H is a pre-determined time window in which we check which lines depart from i (competitive lines), $\mu_{m,l}^H(k)$ equals 1 if line l is the line arriving first at station m during $(kT, kT+H]$ among the competitive lines, $\tilde{a}_m^q(k)$ is the historical numbers of passengers alighting at station m during the time period $[\lambda q\ ,\ (\lambda+1)q)$ for which $\lambda q \leq kT < (\lambda+1)q$, and λ equals the number of time periods of length q in which a daily operational period is divided.

The dynamics of the state vector of the overall system evolve according to:

$$\begin{bmatrix} w_i \\ e_i \end{bmatrix}(k+1) = \begin{bmatrix} 1 - \eta_{i,l}^s(k) \cdot \theta_{i,l}^s(k) & 1 \\ 0 & 1 \end{bmatrix}(k) \cdot \begin{bmatrix} w_i \\ e_i \end{bmatrix}(k) + \xi_i^b(k) \tag{3}$$

where $e_i(k)$ is the number of passengers entering station i and $\xi_i^b(k)$ is zero-mean Gaussian noise that accounts for modelling errors. Real-time APC measurements, $z_i^w(k)$, and historical measurements, $z_i^{\tilde{e}}(k)$, can be treated as noisy measurements for $x_i(k)$, as follows:

$$\begin{bmatrix} z_i^w \\ z_i^{\tilde{e}} \end{bmatrix}(k) = \begin{bmatrix} \beta_{i,l}^s(k) \cdot \theta_{i,l}^s(k) & 0 \\ 0 & \varphi\left(1 - \beta_{i,l}^s(k)\right) \end{bmatrix} x_i(k) + \psi_i^w(k) \tag{4}$$

where $\beta_{i,l}^s(k)$ equals 1 if a run of line l that leaves station i is equipped with APC and 0 otherwise, ϕ equals 1 if historical data are available and 0 otherwise, and $\psi_i^w(k)$ is zero-mean Gaussian noise accounting for measurement errors.

To estimate alighting passengers, we formulate a dynamical model for the alighting rate $\gamma_i^s(k)$, which is defined as a station-based variable, as follows:

$$\gamma_i^s(k) = \sum_{j \in J_l} \eta_{i,j}^r(k) \frac{a_j^r(k)}{p_j(k)} \tag{5}$$

where $\eta_{i,j}^r(k)$ equals 1 if run j departs from station i and 0 otherwise, $a_j^r(k)$ is the number of passengers alighting from run j and $p_j(k)$ is the number of passengers on-board run j. The alighting rate is modelled via a random walk dynamic as follows:

$$\gamma_i^s(k+1) = \gamma_i^s(k) + \xi_i^\gamma(k) \tag{6}$$

where $\xi_i^\gamma(k)$ is zero-mean Gaussian noise accounting for modelling errors. Real-time APC measurements, $z_i^{\bar{\gamma}}(k)$, and historical measurements, $z_i^{\tilde{\gamma}}(k)$, are considered as follows:

$$\begin{bmatrix} z_i^\gamma \\ z_i^{\tilde{\gamma}} \end{bmatrix}(k) = \begin{bmatrix} \beta_{i,l}^s(k) \\ \varphi\left(1 - \beta_{i,l}^s(k)\right) \end{bmatrix} \gamma_i^s(k) + \psi_i^\gamma(k) \tag{7}$$

where $\psi_i^{\bar{\gamma}}(k)$ is zero-mean Gaussian noise accounting for measurement errors.

The process model (3) and measurement model (4) are used in a Kalman filter (Kalman & Bucy, 1961) to estimate the number of passengers boarding run j, $\hat{b}_j^r(k)$, and the process model (6) and measurement model (7) for estimating the alighting rate of run j, $\hat{\gamma}_j(k)$. Finally, the number of passengers on-board, $\hat{p}_j(k+1)$, is estimated as:

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

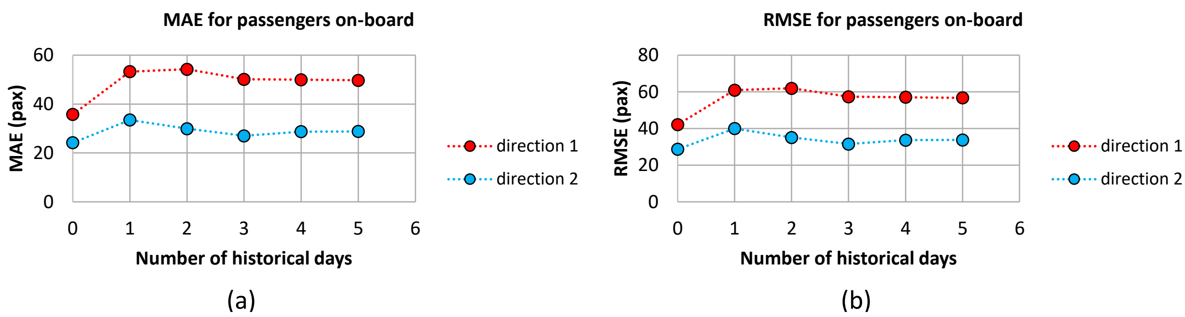

Figure 1 – *Estimation errors for the number of passengers on-board*

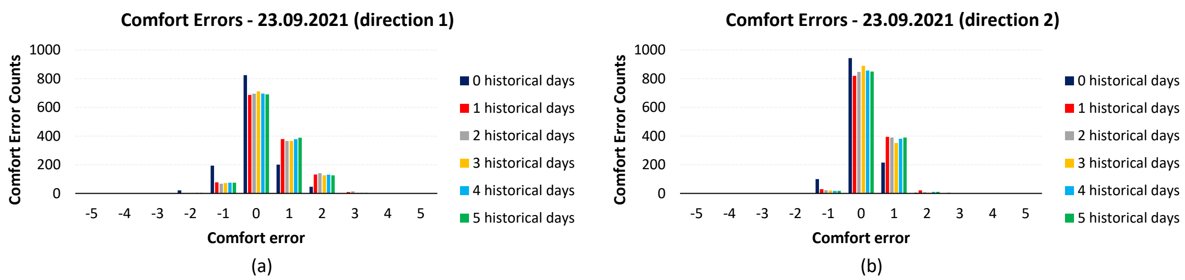

Figure 2 – *Comfort level errors*

3 RESULTS

The case study refers to a part of the Helsinki commuter train network, in Finland. The main terminal is the Helsinki station and trains move in two directions (Helsinki outbound and Helsinki inbound). Data from regular weekdays of September 2021 are considered. The APC data availability is quite high (i.e., more than 70%) for these days.

By implementing the proposed methodology, the estimation errors for passengers on-board are shown in Figure 1. Direction 2 (Helsinki inbound) is associated with the lowest errors both in terms of mean average error (MAE) and root mean squared error (RMSE) when compared with direction 1. Historical days do not improve the performance of the estimations, which can be partly attributed to the added demand uncertainty and the different APC availability per day. Helsinki terminal is the main demand attractor, hence it is intuitively expected and confirmed that direction 2 has greater uncertainty regarding which line travellers will board before Helsinki station (i.e., they will board any convenient line since all lines end in Helsinki terminal), but lower uncertainty regarding alighting rates. The reverse holds for direction 2.

The comfort level errors for the weekday of Thursday, 23.09.2021, and both directions of movement are shown in Figure 2. There are six comfort levels in the considered framework (i.e., level 1 is the least and level 6 is the most crowded), hence the comfort estimation errors range from -5 to 5. The errors are mostly equal to 0, followed by errors equal to 1, and a very small amount of higher errors, implying a good performance of the proposed methodology.

In order to investigate how the different levels of APC availability affect the performance of the dynamic passenger modelling, we selected random runs with observed counts and we treated them as non-observed. We considered the cases of 25%, 50%, and 75% APC availability and different numbers of historical days. We repeated ten times per case. The random selection of runs to be considered as non-observed led to different levels of errors in each case. Figure 3 shows the results and highlights that the effects of random selection are greater for lower APC availability. Historical days again do not improve the performance of the dynamic modeling in the investigated cases. Note that this figure is part of ongoing research.

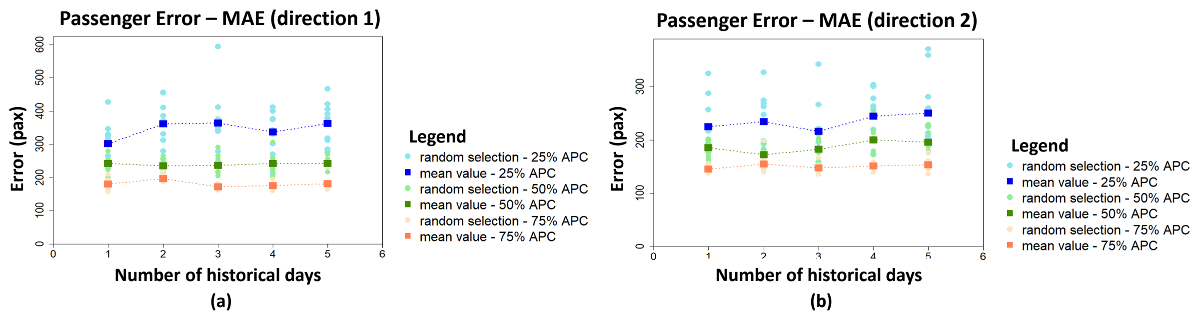

Figure 3 – *Estimation errors for the number of passengers on-board in different cases*

4 DISCUSSION

This paper presents a Kalman filtering-based framework for estimating on-board comfort levels in the case of real-time non-exhaustive APC data. The method has been successfully validated through application to the commuter train network of Helsinki. The results highlight the importance of accounting for demand uncertainties while developing and evaluating such methods, as indicated by the different performance in each direction and the role of historical days. The % APC availability is critical for the estimation performance, since each run's information availability contributes differently to the quality of estimations. Therefore, it is of major importance for public transport operators to strategically select which runs to equip with APC devices. Future work aims at extending the capabilities and accuracy of the proposed method.

5 ACKNOWLEDGEMENTS

The work of Charalampos Sipetas and Claudio Roncoli was partly funded by the Research Council of Finland project ALCOSTO (no. 349327) and the Horizon Europe project ACUMEN (no. 101103808).

References

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

Kalman, Rudolph E, & Bucy, Richard S. 1961. New results in linear filtering and prediction theory.

Roncoli, Claudio, Chandakas, Ektoras, & Kaparias, Ioannis. 2023a. Estimating on-board passenger comfort in public transport vehicles using incomplete automatic passenger counting data. *Transportation Research Part C: Emerging Technologies*, **146**, 103963.

Roncoli, Claudio, Chandakas, Ektoras, & Kaparias, Ioannis. 2023b. A Kalman filter based estimation method for on-board public transport passenger comfort using incomplete Automatic Passenger Counting data, *Presented at 102nd Annual Meeting of the Transportation Research Board, Washington, D.C.*

Sipetas, Charalampos, Roncoli, Claudio, Chandakas, Ektoras, & Kaparias, Ioannis. 2024. Real-time on-board crowding estimation in public transport networks with multiple lines using non-exhaustive Automatic Passenger Counting data, *Presented at 103rd Annual Meeting of the Transportation Research Board, Washington, D.C.*