Path-based crowding information in public transport based on the automated passenger counting data

Arkadiusz Drabicki*[1], Oded Cats[2], Constantinos Antoniou[3]

[1] Dr., Group of Machine Learning Research, Jagiellonian University in Krakow, Poland; Chair of Transportation Systems Engineering, Technical University of Munich, Germany

[2] Prof. Dr., Department of Transport and Planning, Delft University of Technology, The Netherlands

[3] Prof. Dr., Chair of Transportation Systems Engineering, Technical University of Munich, Germany

SHORT SUMMARY

Travel experience in public transport networks is negatively affected by passenger overcrowding and incomplete travel information. In this study, we introduce a concept of path-based crowding information (CI), i.e. a personalized CI on cumulative travel comfort conditions, generated from the APC passenger flow data. It extends the classical CI concept by accounting for both passenger loads (between stops) as well as the dwelling exchanges (at stops) and emerging seat probability along the O-D travel path. The path-based CI framework is first applied on the toy network, and then on the APC data of Rzeszów case study (Poland), revealing additional seating possibility in up to 25% of all travel instances, and in over 40% of such cases – along the majority of their O-D travel route. Our findings underscore the potential of personalised, path-based CI systems to improve travel experience, especially in congested PT networks.

Keywords: public transport; crowding; real-time information; automated passenger counting

1. INTRODUCTION

Passenger crowding can substantially decrease the perceived quality and reliability of public transport (PT) travel (Tirachini et al., 2013; Jenelius, 2018; Darsena et al., 2023). Crowding disutility is strongly correlated with lack of seat availability and rising standing discomfort, as well as lower travel satisfaction and higher safety and security concerns (Kim et al., 2015; Yap et al., 2020). Its importance has grown further in the aftermath of the COVID-19 pandemic crisis, which amplified the travellers' aversion to in-vehicle overcrowding (Fluegel and Hulleberg, 2022; Drabicki et al., 2025; Yap et al., 2025). Hence, a major challenge pertains to designing the effective measures for recognizing and mitigating the PT overcrowding risks, especially in high-density, congested PT networks.

Real-time crowding information (CI) is one of the prospective and cost-effective solutions that could help reduce crowding externalities, improve journey experience and capacity utilization (Zhang et al., 2017; Jenelius, 2020; Drabicki et al., 2021). Its feasibility is nowadays enabled by developments in intelligent transportation systems (ITS) and increasing presence of data collection systems on passenger movements, such as automated passenger counters (APC), which can be processed and conveyed in real time to PT users.

State-of-the-practice solutions typically report the *segment-based* CI, which is basically based on the reported passenger load (volume-capacity ratio) on a given PT vehicle segment between 2 consecutive stops. While informative, this approach fails to capture crucial dynamics in PT travel – i.e. the seat capacity rotation at stops. As passengers alight at intermediate stops, seats become available for standing passengers already on board (as well as newly boarding passengers). This implies that a certain, cumulative seating probability can be eventually aggregated along the passenger's O-D travel path. However, this cannot be captured and communicated to PT users under the conventional CI framework. The only prior framework for personalised, *path-based* CI is provided by Jenelius (2020), who defines probabilistic seat allocation metrics for O-D trips using the APC and AVL data. Still though, important gaps can be traced in state-of-the-art, especially: (1.) no universal framework exists for descriptive (categorical) path-based CI, which can be more comprehensible for users and conveyable for operators; (2.) existing seat allocation models lack an explicit treatment of alighting flow impacts and "competing" passenger flows at stops; (3.) little is known about the path-based CI effectiveness with respect to network congestion and magnitude of dwelling exchanges, and its prospective added value versus the conventional segment-based CI.

We aim to address these gaps in this study and introduce the concept of path-based CI, i.e. travel advice on cumulative comfort conditions for a given origin-destination (O-D) path in PT network. Path-based CI incorporates not only the data on volume-capacity ratio *between the stops*, but also on passenger out- and inflows *at the stops*, to account for the emerging, cumulative seat probability. We formulate the mathematical framework for path-based CI generation from passenger flow data from the APC systems and O-D trip distribution data. We define the so-called seat competition model to explicitly capture the impact of rotating seat capacity along a given PT trip. Output path-based CI is then a function of segment-based CI and accumulated seat probability at a given travel instance. Potential effects and consequences of this concept are described in further sections of this paper.

2. METHODOLOGY

The path-based CI algorithm consists of the following stages: (1.) computing the baseline segment-based CI from APC data, (2.) applying the seat competition model at each stop, (3.) accumulating the seat probability along the O-D path to derive the final path-based CI value.

Baseline segment-based CI

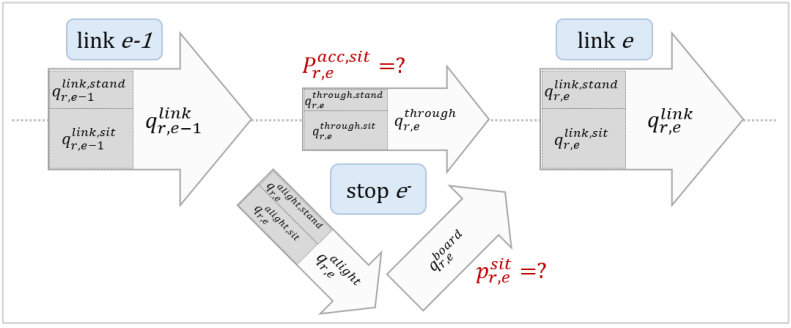
First, the segment-based CI $\beta_{r,e}^{CI_seg}$ is calculated at the disaggregate level of each individual travel segment. It is a function of the passenger load $q_{r,e}^{link}$, seated capacity $\kappa_{r,e}^{sit}$ and total capacity $\kappa_{r,e}^{total}$ of line run (departure) r along the line segment from stop e to stop $e+1$. Segment-based CI is mapped on the 4-level categorical scale (tab. 1):

Table 1: Segment-based CI mapping framework.

Segment-based CI $\beta_{r,e}^{CI_seg}$	Assumptions	Interpretation
level 1	$\frac{q_{r,e}^{link}}{\kappa_{r,e}^{sit}} \leq 0.5$	50%+ seats available
level 2	$0.5 < \frac{q_{r,e}^{link}}{\kappa_{r,e}^{sit}} \leq 1.0$	individual seats available
level 3	$1.0 < \frac{q_{r,e}^{link}}{\kappa_{r,e}^{sit}}$ and $\frac{q_{r,e}^{link}}{\kappa_{r,e}^{total}} \leq 0.75$	standing in moderate crowding
level 4	$0.75 < \frac{q_{r,e}^{link}}{\kappa_{r,e}^{total}}$	standing in (high) overcrowding

Seat competition model

Figure 1: Seat competition model at stop *e*.



Second, to explicitly account for the seat capacity rotation dynamics at stops (fig. 1), the seat competition model computes the unit seat probability $p_{r,e}^{sit}$ at stop *e* as the ratio of *rotating* seat capacity to *competing* passenger volume:

$$p_{r,e}^{sit} = \frac{c_{r,e}^{sit,avail}}{q_{r,e}^{sit,compete}} \tag{1}$$

Rotating seat capacity $c_{r,e}^{sit,avail}$ is equal to the residual seat capacity of vehicle run *r*, reduced by passengers seated on-board who pass through the stop *e*:

$$c_{r,e}^{sit,avail} = \kappa_{r,e}^{sit} - q_{r,e}^{through,sit} \tag{2}$$

Competing passenger volume $q_{r,e}^{sit,compete}$ is the sum of standing passengers passing through and newly boarding passengers at stop *e*:

$$q_{r,e}^{sit,compete} = q_{r,e}^{board} + q_{r,e}^{through,stand} \tag{3}$$

Crucially, our model assumes both these passenger groups as competing equally for freed seats. This departs from the classical seating priority rules and should thus better reflect the urban PT travel characteristics, where standing passengers may not always claim the vacated seats first.

The key control variable for explicit computation of seat capacity rotation is the share of alighting passengers $\alpha_{r,e}^{alight,sit}$ who were sitting on the link segment upstream of stop *e*. (E.g. the value of $\alpha_{r,e}^{alight,sit} = 1.0$ implies that none of those alighting at stop *e* were previously standing.). It is used to distinguish the seated from standing passengers in the alighting flows:

$$q_{r,e}^{alight,stand} = min \begin{cases} q_{r,e-1}^{link,stand} \\ q_{r,e}^{alight} \cdot \left(1 - \alpha_{r,e}^{alight,sit}\right) \end{cases} \quad (4)$$

$$q_{r,e}^{alight,sit} = q_{r,e}^{alight} - q_{r,e}^{alight,stand} \quad (5)$$

Accumulated seat probability and output path-based CI

Passengers traversing multiple line segments can observe repeated instances of seat capacity becoming (promptly) available at each stop where $c_{r,e}^{sit,avail} > 0$ and $q_{r,e}^{sit,compete} \neq 0$. Hence, the accumulated seat probability $P_{i,r,e}^{acc,sit}$ upon departing from stop *e* is a function of unit seat probabilities aggregated at all the upstream stops *{e₀, ..., e-2, e-1}* along travel path *i*, assumed to be equal to the joint probability of independently distributed events:

$$P_{i,r,e}^{acc,sit} = 1 - \prod_{e \in i}\left(1 - p_{r,e}^{sit}\right) \quad (6)$$

In the final step, the **path-based CI** $\boldsymbol{\beta_{i,r,e}^{CI_path}}$ is determined on a categorical, 4-level scale, as a function of 2 input attributes: segment-based CI $\beta_{r,e}^{CI_seg}$ and accumulated seat probability $P_{i,r,e}^{acc,sit}$ valid at stop *e* (tab. 2). Hence, the path-based CI can indicate the possibility of higher on-board comfort than segment-based CI if $P_{i,r,e}^{acc,sit}$ accumulated along the path *i* exceeds a given threshold.

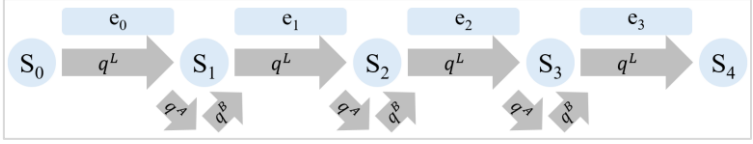
Table 2: Path-based CI mapping framework.

Path-based CI $\beta_{i,r,e}^{CI_path}$	Assumptions	Interpretation
level 1	$\beta_{r,e}^{CI_seg} = \{*\} \parallel 0.67 < P_{i,r,e}^{acc,sit}$	high seat probability
level 2	$\beta_{r,e}^{CI_seg} = \{**\} \parallel 0.33 \leq P_{i,r,e}^{acc,sit} < 0.67$	low or moderate seat probability
level 3	$\beta_{r,e}^{CI_seg} = \{***\}$ && $P_{i,r,e}^{acc,sit} \leq 0.33$	standing in moderate crowding
level 4	$\beta_{r,e}^{CI_seg} = \{****\}$ && $P_{i,r,e}^{acc,sit} \leq 0.33$	standing in (high) overcrowding

3. RESULTS AND DISCUSSION

Toy-network demonstration

Figure 2: Toy network topology.



We use a schematic, unidirectional bus corridor model (fig. 2) to understand the overall working principles of path-based CI and systematically explore its sensitivity with respect to: link passenger loads q_e^L (100 to 200% of seat capacity), dwelling exchange factor τ^{dwell} (i.e. boarding and alighting volumes equal to 10 – 90% of link volumes) and seated alighting share factor $\alpha_{r,e}^{alight,sit}$ (equal to 0, 0.25 or 0.5). Under such settings, the segment-based CI implies always standing crowding conditions ($\beta_{r,e}^{CI_seg}$ equal to level 3 or 4). However, the path-based CI can reveal a lower crowding level ($\beta_{r,e}^{CI_path}$ equal to level 1 or 2) after traversing a single stop already if $\tau^{dwell} =$ 0.3 – 0.5 of passenger volumes are exchanged there, or if the share of alighting standees $\alpha_{r,e}^{alight,sit}$ rises to 0.25 (or more). The influence of the 2 latter parameters is especially conspicuous in case of high link volumes equivalent to 150 – 200% of seat capacity, when they can enable the $\beta_{r,e}^{CI_path}$ of level 2. Aggregation of seating probability after 2 – 3 stops can trigger the path-based CI at level 1 or 2 even under significant network (over)saturation, provided that τ^{dwell} and $\alpha_{r,e}^{alight,sit}$ factors are sustained at specific threshold rates. Detailed results of segment- and path-based CI for all the scenarios, with respect to passenger flow assumptions at downstream stops, and underlying accumulated seat probabilities (fig. 3), are shown in fig. (4).

Figure 3: Toy-network results: accumulated seat probability.

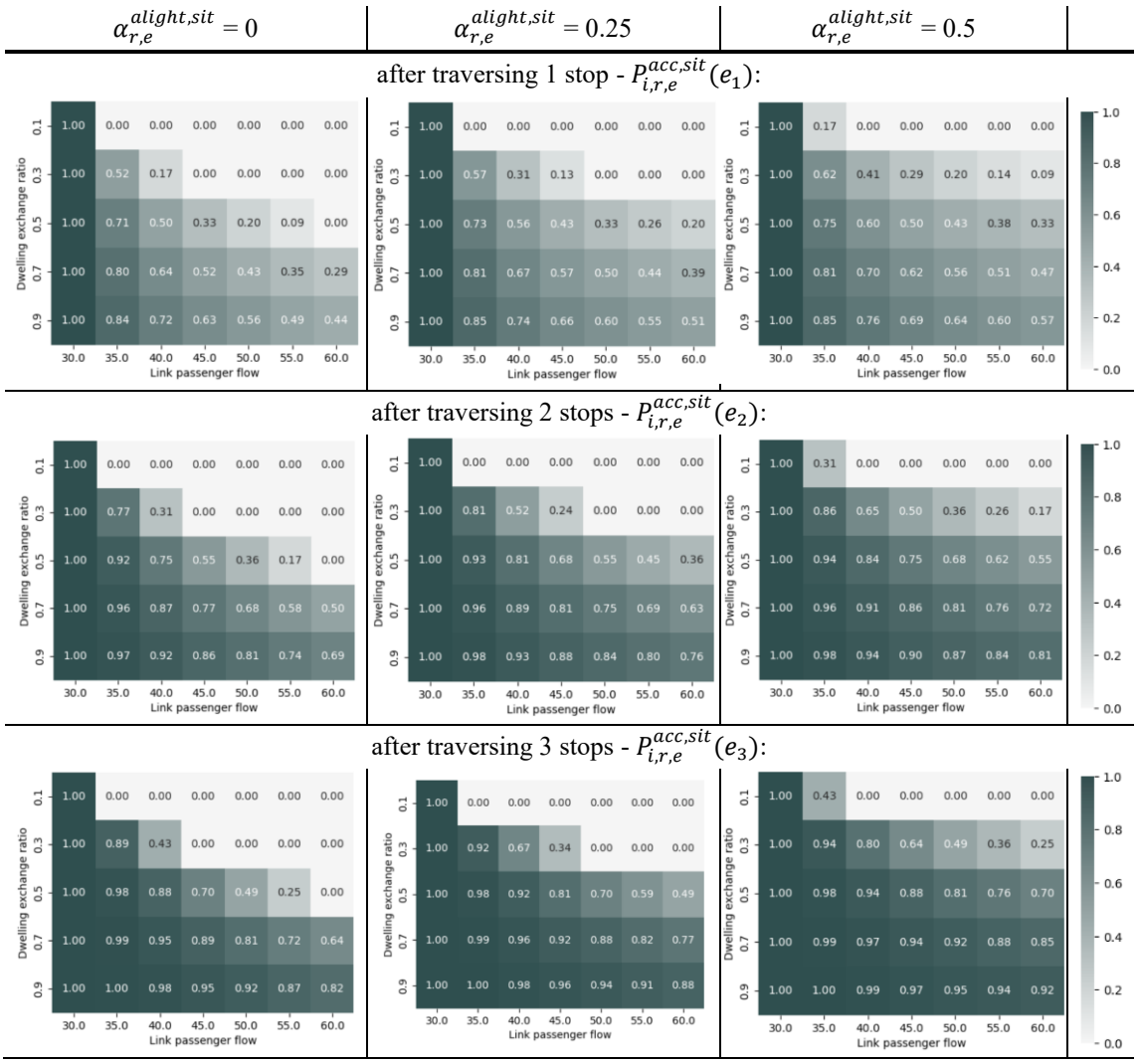
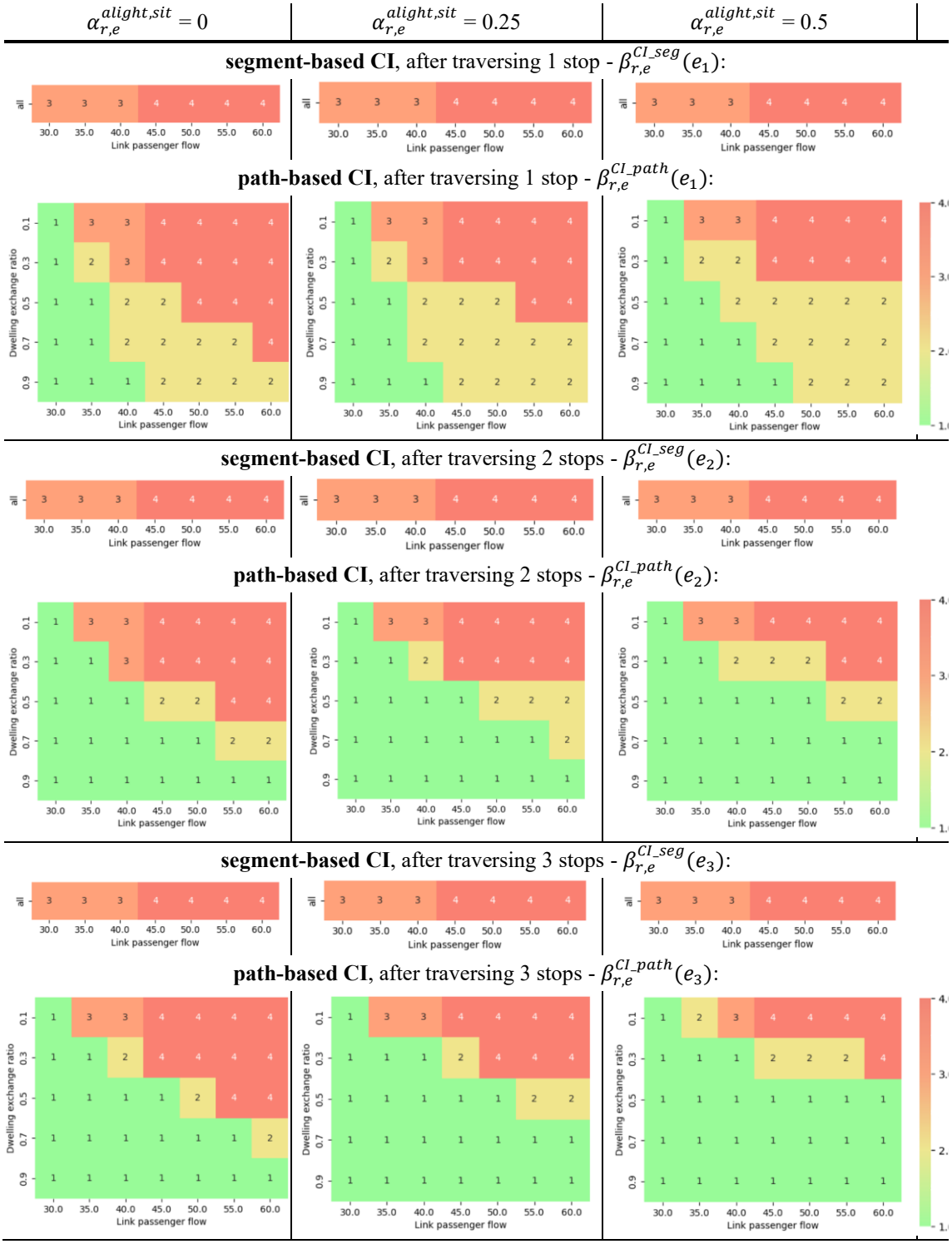


Figure 4: Toy-network results: segment- and path-based CI.



Real-world APC dataset demonstration

Next, to analyse the plausible network-wide effects of path-based CI, the framework was applied to an empirical APC dataset from the bus network of Rzeszów (Poland), a city of approx. 200k inhabitants (est. 400k metro area population). The city bus system consists of 54 bus lines operated by *MPK Rzeszów*, with over 80% of its bus fleet equipped with APC devices. We inferred the necessary input information from 2 principal data sources (1.) main data on passenger loads from the APC data registered in October 2023 (10 representative working days), and (2.) auxiliary data on trip distributions from the Rzeszów strategic transport model (calibrated to the 2022 – 2023 APC data) (City of Rzeszow, 2023). Data curation and matching of both datasets resulted in approx. 75% of O-D travel paths assigned with the corresponding APC records. We test our path-based CI framework for all the bus trips and peak-hour trips (i.e. filtered between 7:00 – 9:00 and 15:00 – 17:00), assuming the seated alighting share factor $\alpha_{r,e}^{alight,sit}$ of 0 or 0.5.

Comparison of segment- vs. path-based CI (tab. 3, tab. 4) shows that in approx. 17 – 23% of all the individual travel segments, the $\beta_{r,e}^{CI_seg}$ implies standing crowding (level 3 or 4) but $\beta_{r,e}^{CI_path}$ reveals now seating possibility (level 1 or 2). Benefits are particularly evident in case of higher (over)crowding conditions: path-based CI shows seating possibility for approx. 74% of instances with $\beta_{r,e}^{CI_seg}$ equal to level 3 and 37% of those equal to level 4 (assuming $\alpha_{r,e}^{alight,sit}$ = 0). For the factor of $\alpha_{r,e}^{alight,sit}$ = 0.5, the corresponding rates would rise to 85% and 43%. Moreover, path-based CI can also reassure travellers about higher seating probability ($\beta_{r,e}^{CI_path}$ of level 1) than segment-based CI ($\beta_{r,e}^{CI_seg}$ of level 2 or more) in ca. 43 - 53% of such instances.

Table 3: Case-study results: segment- vs. path-based CI (all bus trips).

Share of weighted OD path segments		**path-based CI $\beta_{r,e}^{CI_path}$**			
		{*}	{**}	{***}	{****}
segment-based CI $\beta_{r,e}^{CI_seg}$	{*}	45.6%	*n/a*	*n/a*	*n/a*
	{**}	27.6%	~0%		
	{***}	13.2%	1.8%	4.8%	
	{****}	2.1%	0.5%	~0%	4.5%

Table 4: Case-study results: segment- vs. path-based CI (peak-hour bus trips).

Share of weighted OD path segments		**path-based CI $\beta_{r,e}^{CI_path}$**			
		{*}	{**}	{***}	{****}
segment-based CI $\beta_{r,e}^{CI_seg}$	{*}	33.2%	*n/a*	*n/a*	*n/a*
	{**}	31.4%	~0%		
	{***}	16.8%	2.2%	6.6%	
	{****}	2.9%	0.7%	~0%	6.1%

4. CONCLUSIONS

This study introduces a novel concept of path-based crowding information (CI) in public transport (PT) networks. It extends the conventional segment-based CI on expected travel comfort conditions by accounting not only for in-vehicle passenger loads *between* the stops, but also for seat probability accumulated *at* the stops along a given O-D travel path. Principal contribution of this

study pertains to the mathematically grounded path-based CI generation algorithm based on the APC and O-D passenger flow data, including a custom seat competition model. Secondary contribution lies in the analytical insights from toy-network demonstrations (working mechanisms of path-based CI) and real-world dataset applications (prospective network effects of path-based CI).

Our results demonstrate that path-based CI can convey much more favourable information on travel comfort conditions than classical segment-based CI, even in overcrowded PT conditions. The toy-network illustrates that passengers can be reassured about new or higher seating possibility even with load volumes being 1.5 – 2.0 times higher than vehicle seat capacity, provided that seat capacity rotation at stops achieves around 30 – 50% at consecutive stops. The case-study application to APC dataset from Polish city of Rzeszów underscores that approx. 25% of all passenger trips and 40 – 80% of overcrowded travel instances could benefit from more favourable travel comfort information under the path-based CI. These benefits are largely sustained during peak hours, and already without assuming increased seat capacity rotation (i.e. $\alpha_{r,e}^{alight,sit} = 0$). Moreover, in about 40% of trip instances that benefit from path-based CI, passengers could be reassured about extra seating possibility for the majority of their O-D travel route.

Our research highlights the potential of path-based CI to serve as a helpful and reliable passenger information system, extending the capabilities of traditional, segment-based CI solutions. The framework is practically implementable with the modern-day passenger flow monitoring systems such as APC data (or analogous solutions). Customisation of CI for individual travel paths can contribute to improved perceived quality and satisfaction of PT journeys, especially in high-demand urban PT networks.

This study lays the groundwork for further research on personalized, path-based CI systems, and the contributions and insights of this study can be extended in a number of directions, including: performance evaluation measures for path-based CI systems; behavioural underpinning and empirical validation of the path-based CI framework; feasibility with other passenger flow data sources (e.g. AFC data); wider sensitivity analysis and application extended to the large-scale metropolitan PT networks.

ACKNOWLEDGEMENTS

Authors would like to express their gratitude to the Public Transport Authority of the City of Rzeszów (*ZTM Rzeszów*) and the Municipality of Rzeszów (*UM Rzeszów*) for sharing the input datasets for the purposes of this research, as well as to Mr. Dominik Zegar for his support in data preparation and processing. This research was partially funded by the National Science Centre, Poland (NCN), grant agreement no.: 2025/56/C/HS4/00296.

REFERENCES

City of Rzeszow [2023]. The 2023 Rzeszow Comprehensive Travel Surveys [Polish: *Rzeszowskie Badanie Ruchu 2023*.]

Darsena D, Gelli G, Iudice I, Verde F. [2023] Sensing technologies for crowd management, adaptation, and information dissemination in public transportation systems: A review. *IEEE Sensors Journal*. Nov 23;23(1):68-87.

Drabicki A, Kucharski R, Cats O, Szarata A. [2021] Modelling the effects of real-time crowding information in urban public transport systems. *Transportmetrica A: Transport Science*. Dec 10;17(4):675-713.

Drabicki A, Cats O, Kucharski R. [2025] Has the COVID-19 pandemic affected travellers' willingness to wait with real-time crowding information?. *Travel Behaviour and Society*. Jan 1;38:100895.
Flügel S, Hulleberg N. [2022] Aversion to In-vehicle Crowding before, during and after the COVID-19 Pandemic. *Findings*. Aug 15.
Kim KM, Hong SP, Ko SJ, Kim D. [2015] Does crowding affect the path choice of metro passengers?. *Transportation Research Part A: Policy and Practice*. Jul 1;77:292-304.
Jenelius E. [2018] Public transport experienced service reliability: Integrating travel time and travel conditions. *Transportation Research Part A: Policy and Practice*. Nov 1;117:275-91.
Jenelius E. [2020] Personalized predictive public transport crowding information with automated data sources. *Transportation Research Part C: Emerging Technologies*. Aug 1;117:102647.
Tirachini A, Hensher DA, Rose JM. [2013] Crowding in public transport systems: effects on users, operation and implications for the estimation of demand. *Transportation Research Part A: Policy and Practice*. Jul 1;53:36-52.
Yap M, Cats O, Van Arem B. [2020] Crowding valuation in urban tram and bus transportation based on smart card data. *Transportmetrica A: Transport Science*. Dec 20;16(1):23-42.
Yap M, Wong H, Cats O. [2025] Public transport crowding valuation in a post-pandemic era. *Transportation*. Feb;52(1):287-306.
Zhang Y, Jenelius E, Kottenhoff K. [2017] Impact of real-time crowding information: a Stockholm metro pilot study. *Public Transport*. Oct;9(3):483-499.