



Delft University of Technology

## Modelling the effects of real-time crowding information in urban public transport systems

Drabicki, Arkadiusz; Kucharski, Rafal; Cats, Oded; Szarata, Andrzej

### DOI

[10.1080/23249935.2020.1809547](https://doi.org/10.1080/23249935.2020.1809547)

### Publication date

2020

### Document Version

Final published version

### Published in

Transportmetrica A: Transport Science

### Citation (APA)

Drabicki, A., Kucharski, R., Cats, O., & Szarata, A. (2020). Modelling the effects of real-time crowding information in urban public transport systems. *Transportmetrica A: Transport Science*, 17 (2021)(4), 675-713. <https://doi.org/10.1080/23249935.2020.1809547>

### Important note

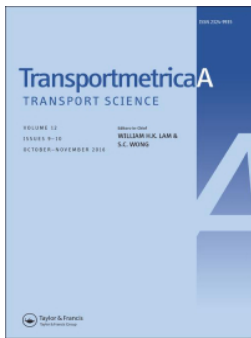
To cite this publication, please use the final published version (if applicable).  
Please check the document version above.

### Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

### Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.  
We will remove access to the work immediately and investigate your claim.



## Modelling the effects of real-time crowding information in urban public transport systems

Arkadiusz Drabicki, Rafał Kucharski, Oded Cats & Andrzej Szarata

To cite this article: Arkadiusz Drabicki, Rafał Kucharski, Oded Cats & Andrzej Szarata (2021) Modelling the effects of real-time crowding information in urban public transport systems, Transportmetrica A: Transport Science, 17:4, 675-713, DOI: [10.1080/23249935.2020.1809547](https://doi.org/10.1080/23249935.2020.1809547)

To link to this article: <https://doi.org/10.1080/23249935.2020.1809547>



© 2020 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 29 Aug 2020.



Submit your article to this journal [↗](#)



Article views: 1503



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 3 View citing articles [↗](#)

# Modelling the effects of real-time crowding information in urban public transport systems

Arkadiusz Drabicki <sup>a,b</sup>, Rafał Kucharski <sup>a</sup>, Oded Cats <sup>a</sup> and Andrzej Szarata <sup>b</sup>

<sup>a</sup>Department of Transport and Planning, Delft University of Technology, Delft, The Netherlands; <sup>b</sup>Department of Transportation Systems, Cracow University of Technology, Krakow, Poland

## ABSTRACT

Public transport (PT) overcrowding is a notorious problem in urban transport networks. Its negative effects upon travel experience can be potentially addressed by disseminating real-time crowding information (RTCI) to passengers. However, impacts of RTCI provision in urban PT networks remain largely unknown. This study aims to contribute by developing an extended dynamic PT simulation model that enables a thorough analysis of instantaneous RTCI consequences. In the model, RTCI is generated and disseminated across the network, and then utilised in passengers' sequential en-route choices. A case-study demonstration of the RTCI algorithm on urban PT network model of Kraków (Poland) shows that instantaneous RTCI has the potential to improve passengers' travel experience, although it is also susceptible to inaccuracy. RTCI provision can yield total travel utility improvements of 3% in typical PM peak-hour, with reduced impacts of the worst overcrowding effects (in terms of denied-boarding and in-vehicle travel disutility in overcrowded conditions) of 30%.

## Highlights:

- Real-time crowding information (RTCI) is an increasingly feasible solution in public transport.
- We introduce a novel framework for modelling the network effects of instantaneous RTCI.
- Instantaneous RTCI can result in improved travel experience but also substantial inaccuracy risk.
- Reduced impacts of the worst overcrowding experience amount to up to 30%.

## ARTICLE HISTORY

Received 7 January 2020  
Accepted 3 August 2020

## KEYWORDS

Real-time crowding information; RTCI; public transport; overcrowding; path choice

## 1. Introduction

Crowding in public transport (PT) is nowadays one of the main phenomena affecting passengers' travel experience and system performance. Especially in urban and metropolitan realm, negative impacts of passenger overcrowding are likely to determine the quality of

PT service in a substantial way. Despite multi-million ‘hard’ investment solutions, the ever-increasing demand for PT trips routinely outstrips the finite PT system supply, which is observable during peak-hour periods. For example, an ongoing investment programme in London, including a new Crossrail line (planned completion in 2021) and upgrades to the existing Tube network with a total cost of approx. 30–35 bn GBP (ca. 34–40 bn EUR), will increase rail capacity by ca. 30%. Yet, despite such massive-scale efforts, it is projected that system overcrowding will quickly bounce back to its pre-investment levels by the late 2020s, and will become even worse without further infrastructure upgrades (Mayor of London 2015).

Instead, there is a rising emphasis on ‘soft’ travel demand management solutions that potentially allow a more effective utilisation of available system capacity: by increasing both passengers’ and operators’ awareness of current travel conditions, they can help make more informed choices and alleviate the overcrowding experience. In turn, this calls for an in-depth understanding and quantification of crowding effects upon passengers’ travel behaviour, especially in light of innovative ICT solutions. Proper appraisal of overcrowding impacts is also vital for evaluating the PT schemes designed to alleviate them, as analytical models that neglect crowding may eventually underestimate potential benefits by even a margin of 30–60% (Leurent 2009; Tirachini, Hensher, and Rose 2013; van Oort et al. 2015; Cats, West, and Eliasson 2016). In this study, we present a modelling framework for describing the effects of an emerging solution, i.e. passenger real-time crowding information (RTCI) systems. RTCI is a travel demand management measure that can potentially help mitigate the negative experience of overcrowding in PT networks – albeit this claim is yet to be quantitatively and empirically underpinned.

### **1.1. Crowding in public transport systems: impacts and modelling approaches**

PT crowding has been shown to have a substantial impact upon passengers’ travel strategies. Crowding effects are all the more detrimental given that negative PT travel experiences tend to be particularly memorable (Abenzoza, Cats, and Susilo 2017), and in certain circumstances, overcrowding may actually become the main driver of PT travel dissatisfaction (Börjesson and Rubensson 2019). Essentially, crowding induces an externality cost which raises the marginal social cost of travelling (Hörcher, Graham, and Anderson 2017), and though it does not always affect the nominal journey times, it imposes substantial travel time disutility (Cats, West, and Eliasson 2016). This disutility is associated with multiple aspects such as reduced travel comfort (e.g. inability to travel seated), loss of travel time productivity, safety and security concerns, increased stress and anxiety (Kim et al. 2015), as well as raised perception of travel time unreliability and risk of arriving late at destination (Tirachini, Hensher, and Rose 2013). PT users may respond to crowding experience by adjusting their travel strategies (Tirachini, Hensher, and Rose 2013). These pertain to, among others, boarding a different train carriage (Pefitsi, Jenelius, and Cats 2020b), shifts in departure time choice, mode choice and route choice (Tirachini, Hensher, and Rose 2013), potentially leading to reduced trip frequency or even resignation from travelling altogether (Szarata 2014).

Passenger crowding effects are typically represented in PT assignment models using one of the following approaches. The first involves frequency-based PT assignment – mostly static and macroscopic-level models that represent crowding phenomena in a simplified,

implicit approach (Drabicki, Kucharski, and Szarata 2017b). Crowding impact is represented here as an additional path utility impedance, e.g. an in-vehicle travel time multiplier (Cats and Hartl 2016). As such, crowding penalty aims to reflect the disutility inflicted by on-board discomfort or denied boarding, but eventually, no strict capacity limits are enforced (Schmöcker et al. 2011). Uniform crowding penalty applied to total passenger flows implies that no differences in individual crowding experience can be considered. Crowding penalty is estimated based on average volume-to-capacity ratios without accounting for variability in individual vehicle loading levels (Cats and Hartl 2016). The second approach involves an explicit representation of overcrowding effects, utilised in scheduled-based PT assignment models (Drabicki, Kucharski, and Szarata 2017b). These are typically dynamic and microscopic-level models, able to capture the wide range of PT (over)crowding phenomena along with their inherent variability and stochasticity (Hamdouch et al. 2011; Cats, West, and Eliasson 2016), notably: strict capacity constraints of individual PT vehicles, denial-of-boarding and queuing phenomena, boarding and seating priority rules, demand-supply interactions (flow-dependent dwell times) and service reliability (Schmöcker et al. 2011; Cats, West, and Eliasson 2016; Cats and Jenelius 2018). Furthermore, dynamic simulation-based assignment models allow us to capture the effects of on-board discomfort upon travel utility at the individual passenger (agent) level, additionally distinguishing between standing vs. seating (dis)utility. Recent advancements in dynamic PT assignment (Pefitsi, Jenelius, and Cats 2020a) aim to reflect the passengers' considerations of (expected) load at the individual train carriage level in their travel choices. Consequently, dynamic scheduled-based PT assignment models are deemed better poised to evaluate the implications of overcrowding on passengers' en-route decisions, travel experience and overall system performance (Cats, West, and Eliasson 2016).

Impact of on-board crowding upon travel (dis)utility is typically represented by means of an additional in-vehicle travel time multiplier, the so-called **crowding penalty** (Tirachini, Hensher, and Rose 2013; Batarce et al. 2015). Crowding penalty is commonly deemed a non-decreasing function of volume-to-capacity ratio that becomes relevant for conditions corresponding to load factors of 50–90% (i.e. ratio between the number of passengers on-board and the respective vehicle seat capacity) (Wardman and Whelan 2011; Tirachini, Hensher, and Rose 2013) and rises non-linearly thereafter. As crowding disutility primarily relates to qualitative (descriptive) on-board conditions, there is a common agreement in the state-of-the-art that crowding penalty shall be expressed as a function of standees' density (i.e. expressed in terms of [number of passengers per square metre]) rather than nominal load factors (i.e. expressed as a [%] of seat capacity) (Tirachini, Hensher, and Rose 2013; Batarce, Muñoz, and de Dios Ortúzar 2016).

Crowding travel time valuations are commonly derived from stated-preference (SP) experiments, where respondents express their preferred trade-offs between crowding levels, journey times and other factors such as monetary fare, trip characteristics, personal traits, etc. Literature on SP crowding valuations is extensive and findings suggest that maximum mean crowding penalty values range from 1.4–1.5 to 2.1–2.5 (Rudnicki 1999; Whelan and Crockett 2009; Kroes et al. 2014; Haywood and Koning 2015; Batarce, Muñoz, and de Dios Ortúzar 2016; Li, Gao, and Tu 2017). This figure is interpreted as the ratio of disutility of travel time in overcrowded conditions, relative to disutility of travelling in normal conditions (i.e. without crowding discomfort). Individual sources report estimated values as high as 3.0–4.2, depending on model formulation (Tirachini, Hensher, and Rose 2013; Tirachini

et al. 2017). These are maximum values at the capacity limit, i.e. the so-called *crush capacity*, and describe the travel disutility of standing in overcrowded conditions. For seated passengers, max. crowding penalties are lower and, according to the cited SP estimates, range from 1.4 to 1.8. SP crowding valuations are however prone to overestimation bias and discrepancies associated with hypothetical choice situations (Yap, Cats, and van Arem 2020). Bansal et al. (2019) propose a different model formulation that accounts for flexible estimates of unobserved heterogeneity, and report crowding multipliers reaching up to 2.5–3.6 (seating penalty) and 3.2–5.8 (standing penalty).

Alternatively, crowding valuations can be estimated from revealed-preference (RP) data, where crowding penalties are derived from passengers' actual travel behaviour. Data sources for RP studies range from manual observations of passengers' travel choices (e.g. their movements at a station) to advanced evaluation techniques incorporating novel data collection sources, i.e. fusing automated fare collection (AFC) and automated vehicle location (AVL) systems. Crowding penalty values reported by RP studies are notably lower than those reported by SP experiments, with max. values of 1.55–1.95 at the crush capacity limit (Tirachini et al. 2016; Hörcher, Graham, and Anderson 2017; Yap, Cats, and van Arem 2020). Though the literature on RP crowding valuations is less extensive than SP experiments, RP valuations are seemingly more realistic, as they are inferred from actual (revealed) behavioural responses to PT overcrowding.

## 1.2. Real-time crowding information

Meanwhile, another emerging stream in PT assignment concerns modelling the effects of passenger real-time information (RTI) systems. Novel and disruptive ITS solutions offer a promising possibility of improving travel experience but also require adequate PT assignment tools, capable of representing the behavioural influence of RTI systems. As observed by Fonzone (2015) and Fonzone, Schmöcker, and Viti (2016), RTI provision increases the dynamics of passenger choice strategies, expands the set of travel alternatives, induces new decision-making objectives and increases the probability of en-route choice shifts. RTI solutions are widely expected to improve travel conditions as they can help shift the network from user equilibrium towards the optimum state (van Essen et al. 2016). The effects of RTI on travel times – a widespread solution in PT systems worldwide – were shown in agent-based PT assignment models (Cats et al. 2011; Fonzone and Schmöcker 2014) to have a potentially significant impact upon route choice and output travel utility in different scenarios. Perceptions of travel time value and its variability also play an important role in passengers' travel behaviour (Engelson and Fosgerau 2016).

A further possibility that seems increasingly feasible within ITS architecture (Fonzone, Schmöcker, and Viti 2016; Nuzzolo and Comi 2016) involves providing RTI on network passenger volumes, in form of **real-time crowding information (RTCI)**. Information on passenger flows and crowding levels can be nowadays collected from multiple sources, including APC systems (e.g. overhead passenger counters), AFC systems (e.g. tapping devices for smart cards/tickets), vehicle weight sensors, video recording data (e.g. CCTV cameras), mobile and wireless networks (e.g. Bluetooth and WiFi), crowd-sourcing data (e.g. user feedback in travel app) and other solutions (at proof-of-concept stage) that are planned for the near future. The RTCI solutions are reckoned to have potential to support more informed and effective travel choices among passengers (Gentile and Noekel

2016) and to develop demand management solutions that are more responsive to the actual use of PT system capacity (Fonzone, Schmöcker, and Viti 2016). RTCI is a novel solution in early research and implementation stages and little is known about its implications for network-wide passenger flow distribution and system performance. Practical implementation of RTCI systems has been hitherto confined to pilot projects and limited-size applications of providing RTI on train carriages' occupancy loads in Stockholm (Zhang, Jenelius, and Kottenhoff 2017), London (Schmitt 2017), Sydney (Susan 2018) and RTI on bus occupancy levels in Seoul (Seoul Metropolitan Government 2017). These implementations mostly involve communicating localised (station-level or train-level) crowding information, distributed to passengers via electronic station and/or on-board displays, with coverage limited to individual PT lines or line segments. Recent technological advancements already allow for automated processing of input data to generate crowding information. Consequently, a number of travel apps have recently started providing crowding information on PT departures, evaluated from historical user feedback on crowding experience – Google Maps Transit service (Google 2019), Moovit travel app (Moovit Inc 2019). Furthermore, individual apps – developed by the Dutch railways (Nederlandse Spoorwegen 2019), Tokyo railways (East Japan Railway Company 2019) and Singapore buses (Singapore LTA 2019) – now distribute crowding information based on real-time data on vehicle occupancy loads as well (e.g. from weight sensors).

From the research perspective, this is a relatively novel domain, and a limited number of studies have hitherto addressed specific aspects of RTCI systems in PT networks (a detailed summary of the state-of-the-art is presented in (Table 1)). One of challenges relates to generating (evaluating) information on crowding from existing datasets. State-of-the-practice RTCI solutions, mentioned earlier, provide crowding information either as an average historical (user-fed) data or as a 'raw', instantaneous crowding data. The notion of anticipatory (congested) travel time information has been already well-studied in the context of car traffic (e.g. Elhenawy, Chen, and Rakha 2014; Vlahogianni, Karlaftis, and Golias 2014; Woodard et al. 2017; Kucharski and Gentile 2019). Recently, research attention has turned also towards evaluating crowding information in PT networks in form of real-time prediction from multiple data sources, as examined by Jenelius (2018), Więcek et al. (2019), Jenelius (2019), Jenelius (2020) (Table 1). The goal of these analytical frameworks is to predict the crowding level of a PT vehicle with the highest achievable accuracy rate before its actual departure from the stop (station). In general, these studies demonstrate rising feasibility of predicting on-board crowding from the AVL and APC data sources that are increasingly available in PT networks. Crowding prediction can be projected from historical data with a moderate degree of accuracy. This can be improved further with real-time updates, enabling even up to 90% accurate crowding prediction shortly (i.e. a few minutes) before bus/train departure. Although these methods are shown to yield satisfactory results for line segments with regular passenger flows, an important challenge persists with regards to predicting crowding levels at busy, inner-city and transfer PT stops that are characterised by greater passenger flow variability. Overall, these studies provide a crucial methodological contribution towards estimating crowding prediction in practice. Yet, their findings are not fully transferable, as they are derived from case-study applications and not validated on a network scale.

Research on the behavioural impacts of RTCI systems is predominantly limited to stated choice surveys on passengers' willingness to wait in the event that information on crowding

**Table 1.** Summary of state-of-the-art research on the real-time crowding information (RTCI) systems in public transport (PT) networks.

| Study                                   | Focus and methodology   | Key findings   |
|---|---|--|
| <i>A. stated-preference surveys</i>     |   |  |
| Kim, Lee, and Oh (2009)                 | Effect of bus occupancy information on users' choice between boarding a first or second departure – Seoul (South Korea).<br>Descriptive information on on-board crowding (crowded, normal, seats available).  | Boarding probability influenced by trip purpose, travel time and sociodemographic variables. Increasing with information on seats available, while decreasing with higher in-vehicle travel time and bus crowdedness level. Commuting trips involving lower propensity to wait in general, while greater variability (choice heterogeneity) observable for non-commuting trips.  |
| Kroes et al. (2014)                     | Willingness to choose between taking a crowded bus/train service immediately, and waiting for the next, less-crowded service – Paris (France).<br>In-vehicle crowding represented on a 8-level scale.   | Waiting probability ranging from 13% – with the first train being hardly crowded, up to 75% – with the first train being fully crowded and seats available on-board the second departure.<br>Propensity to wait influenced mainly by crowding level in the first departure.  |
| Preston, Pritchard, and Waterson (2017) | Willingness to choose between the first or second regional rail departure with crowding information – South East England (UK).<br>Information on seats available on a 3-level scale (0%, 10% and 60%).  | Average acceptable waiting time ranging from 14 to 22 [mins], except for airport-bound travellers (max. 8 [mins]). Waiting probability also found to be higher for origin (termini) stations, (17–23 [mins]) than for intermediate stations (8–16 [mins]). Commuters somewhat less willing to wait than business or leisure travellers.<br>Significant propensity to wait if possible to get a seated place (for the regional rail trip). Influential factors: trip purpose, station amenities (facilities) and its position along the line. |
| Kattan and Bai (2018)                   | Users' choices between boarding vs. waiting at the platform with occupancy information on the nearest arrival of light rail transit (LRT) – Calgary (Canada).<br>Simplified information on crowding of the first departure only – (yes/no). (No crowding information available for the second departure.) | Propensity to skip the crowded first departure ranging from 45% (commuters) to 65% (non-commuters).<br>Factors influencing the willingness to wait: longer in-vehicle journey time, users above 25 y/old, perceived information unreliability, frequent LRT usage (familiarity with the PT network), origin (termini) stations, warm temperatures (15C and over).  |
| <i>B. real-world observations</i>       |   |  |
| Zhang, Jenelius, and Kottenhoff (2017)  | Pilot RTCI system implementation: observations of train carriage choice when boarding at a metro station – Stockholm (Sweden).<br>RTCI communicated instantaneously from an upstream station, represented on a 3-level scale for each train carriage (green, orange, red).                                | Observed limited RTCI impact on users' behaviour: with max. 4–8% shifts in boarding flows in crowded conditions. Main impact on passenger load of the most popular train carriage (reduced by 4% for crowded trains, increased by 4% for uncrowded trains).<br>Positive reception of RTCI system – 'traffic-light' style RTCI clearly understandable and positively received by most users. Visible preference towards a simplified, descriptive (3-level) RTCI scale.   |

(continued).



**Table 1.** Continued.

| Study                                     | Focus and methodology  | Key findings  |
|---|--|---|
| <i>C. empirical prediction algorithms</i> |  |   |
| Jenelius (2018); Jenelius (2020)          | <p>Framework for car-specific metro train crowding prediction, based on real-world load (APC) data – Stockholm (Sweden).</p> <p>Crowding prediction solved for a given train run, at a target station, at an individual train car level. Crowding predictors (input data): historical APC; real-time APC of previous trains; real-time APC at upstream stations.</p> <p>Methods considered: stepwise regression, lasso regression, boosted regression tree ensemble. Crowding evaluated either in 3 comfort levels, or in absolute number of passengers.</p>   | <p>Case study results: baseline accuracy up to ca. 60% – for prediction based on historical APC only, with twofold tendency to underestimate overcrowded runs, and to overestimate the least-crowded runs.</p> <p>Best crowding prediction for combination of all 3 (historical and current-day) data sources. Achievable baseline accuracy of up to 80%, increasing to over 90% shortly (2–5 min) before departure.</p> <p>Stepwise regression found as the most accurate method; similarly – lasso regression. In contrast, limited accuracy obtained with the boosted tree ensemble prediction method.</p>           |
| Jenelius (2019)                           | <p>Framework for bus crowding prediction from real-world location (AVL) and load (APC) data – Stockholm (Sweden).</p> <p>Crowding prediction solved for a given bus run, at a target stop. Crowding predictors (input data): historical APC; real-time AVL; real-time APC.</p> <p>Prediction method: lasso regression. Crowding evaluated either in 3 comfort levels, or in absolute number of passengers.</p>   | <p>Case study results: prediction based on historical APC data only – moderate accuracy (50–60%). with twofold tendency to underestimate overcrowded runs – and to overestimate the least-crowded runs.</p> <p>Best crowding prediction for combination of all 3 data sources. Achievable accuracy of up to 80% – 90% ca. 2–5 min before departure. Good level of accuracy also possible for crowding evaluated in absolute passenger volumes, with MAE error in single-digit numbers.</p> <p>Higher-order, non-linear regression models (e.g. quadratic terms) – no additional improvement in prediction accuracy.</p> |
| Więcek et al. (2019)                      | <p>Framework for on-board bus comfort level prediction from real-world passenger load (APC) data – Kraków (Poland). Comfort level prediction solved for a given bus run, at a target stop (line segment). Crowding predictors (input data): historical APC data.</p> <p>Bus crowding discretised into 6 comfort levels. Prediction method – homogenous Markov chain concept: bus comfort (crowding) level analysed as a sequence of discrete random variables. Probability of current state (i.e. bus comfort level) being solely a function of state attained in the previous time step (i.e. bus departures).</p> <p>Predicted comfort level determined primarily by the most recent observation(s), with diminishing impact of past states.</p> | <p>Case study findings: relatively high prediction accuracy. Comfort prediction error: a difference of max. 1 comfort level only in majority of cases.</p> <p>Markov chain concept – applicable for evaluating bus crowding on a categorised comfort scale, in case of bus line segments characterised by regular passenger flows.</p> <p>Discrepancy risk and RMSE increasing for the inner-city and transfer stops with higher variability of passenger arrivals.</p> <p>Recommendation for future research: application of heterogeneous Markov chain with transition matrix to improve prediction accuracy.</p>     |

(continued).

**Table 1.** Continued.

| Study  | Focus and methodology   | Key findings  |
|--|---|---|
| <i>D. simulation models</i><br>Nuzzolo et al. (2016) | Mesoscopic public transport assignment model, simulating the impact of predictive RTCI on long-term route (path) choices. Implemented within a day-to-day learning framework.<br>Predictive RTCI evaluated as a fixed-point problem solution – iterative outcome of in-vehicle loads (supply) merging towards choice probabilities (demand).<br>Final PT network state evaluated as a result of day-to-day learning process.  | Case-study simulation – final results after long-term learning process: significant departure time choice shifts (15–25%) and peak widening. Lower waiting time disutility (by 10%) and fail-to-board incidence (by 7%). Total travel utility improvements of 4% to 7%.<br>Very limited route choice impact (possibly due to network topology).   |
| Drabicki et al. (2017a)                              | Proof-of-concept algorithm for simulating the impact of RTCI on instantaneous route (path) choices. Incorporated within the mesoscopic public transport assignment model.<br>Instantaneous RTCI communicated on a 4-level descriptive scale.<br>Limited functionality, improved in this study.  | Preliminary toy-network demonstration: RTCI effects affected by: network saturation (demand) level, choice sensitivity, RTCI response rate.<br>Possible detrimental impact of instantaneous RTCI upon travel experience and information accuracy. Influenced by: limited network topology (e.g. choice limited to 2 bus lines); rising demand sensitivity and penetration rate (ubiquitous response to RTCI); higher network saturation level (moderate and high crowding).<br>Non-transferrable results – further validation needed. Insufficient replication and analysis of the wide-spectre RTCI effects, especially on a complex PT network. |
| Noursalehi, Koutsopoulos, and Zhao (2018, 2019)      | Predictive decision support model for generating RTCI on next train departures from a given station and modelling its impact on instantaneous departure choice.<br>Composed of: (1.) demand prediction module (passenger arrivals at stations) and (2.) on-line simulation model (passengers' instantaneous departure choices). Rolling horizon approach with network predictions updated every 15–30 min.<br>No route choice impact considered (nor day-to-day learning process).<br>Predictive RTCI communicated on a 3-level descriptive scale (boarding: guaranteed, likely, unlikely). | Case study results for London Underground: higher waiting acceptance (i.e. greater deference threshold) leading to lower denial-of-boarding, reduced experience of train overcrowding – and consequently, improved travel comfort.<br>Good prediction accuracy, with minor underestimation risk mainly for low deference threshold. Considered only localised predictive RTCI impact (i.e. crowding information available at a specific station only).  |

levels of the next PT departures is provisioned (Table 1). The stated willingness to wait with RTCI is influenced by trip purpose, necessity to arrive on-time at destination, travel time and sociodemographic variables (Kim, Lee, and Oh 2009), station amenities (facilities) and its position along the PT line (i.e. termini stations) (Preston, Pritchard, and Waterson 2017), weather conditions and familiarity with PT service (Kattan and Bai 2018). It is also primarily driven by the possibility to avoid overcrowding in the next train departure (Kroes et al. 2014) or get a seat in exchange for longer waiting time (Preston, Pritchard, and Waterson 2017). Otherwise, an empirical study of Zhang, Jenelius, and Kottenhoff (2017) concerning the pilot implementation of RTCI on individual train carriage loads at a Stockholm metro station is probably the only RP investigation in this research field so far. Although crowding information has a minor impact on boarding volumes of specific train carriages (maximum changes of 4–8%), the study also observes an overall positive reception of a simplified, descriptive (3-level) RTCI system.

Finally, only a few recent studies have proposed frameworks for simulating the effects of crowding information in PT assignment (Table 1). Nuzzolo et al. (2016) develop a mesoscopic PT simulation approach whose objective is to model the long-term effects of predictive RTCI in day-to-day assignment process, which is evaluated as a fixed-point problem solution. However, the model does not reveal the impact of predictive RTCI on within-day path choice shifts, nor the potential consequences of providing ‘raw’ RTCI, i.e. instantaneous crowding information without a prediction scheme. In another series of works, Noursalehi, Koutsopoulos, and Zhao (2018, 2019) propose a predictive decision support platform that simulates boarding probability with real-time crowding prediction on the next train departures from the station. However, the effects of RTCI were assumed to be limited to departure choice, and hence no route choice shifts nor day-to-day impacts are considered.

Moreover, in an earlier work, we proposed a proof-of-concept framework for modelling the instantaneous effects of RTCI on route choices (Drabicki et al. 2017a). Based on a preliminary experimental setting, it was concluded that RTCI effects can be affected by network saturation level, demand sensitivity, RTCI response rate and RTCI evaluation algorithm. These limited findings are extended in the current study with an improved and generically applicable methodology, a comprehensive set of experiments and an in-depth discussion.

### **1.3. Research gap and contribution**

In this study, we address the gap related to state-of-the-art methods for describing the effects of real-time crowding information (RTCI) in public transport (PT) networks. While a number of recent studies has dealt with certain aspects of modelling the impacts of real-time passenger information in PT networks, a proper analytical framework of instantaneous RTCI and its within-day effects is still missing. We argue that RTCI consequences for PT systems require further research attention. Understanding how RTCI influences passengers’ travel strategies in real-time, how this translates into journey experience, and what is the magnitude of changes in current system performance, are crucial for adequately devising and assessing the RTCI. Also, little remains known whether access to RTCI can be beneficial on a system-wide scale without considering cooperative or anticipatory capabilities.

Yet, to the best of our knowledge, these research questions have not been satisfactorily answered. Specifically, existing PT assignment models do not allow to represent the instantaneous effects of RTCI generation, dissemination and utilisation. Lack of such knowledge

and adequate analytical tools forms a major research gap that can also hamper practical RTCI implementation. Thus, an extended travel behaviour model, capable of representing dynamic changes in passengers' decisions in response to instantaneous RTCI of PT services is needed. Such a model shall be implemented within a PT assignment framework that is able to reproduce the ensuing phenomena in PT system performance.

In this work, we introduce a complete framework for simulating the phenomena arising in the PT system once passengers have access to instantaneous RTCI of PT services. The key methodological contribution lies in extending the dynamic passenger path choice model to describe the within-day influence of RTCI upon passengers' travel decisions, and further ramifications upon service performance and passenger flow distribution. Application results, firstly on toy-networks and followed then by simulations on a real-world model of a city PT network (Kraków, Poland), demonstrate possible changes in passengers' choice patterns and consequences for the real-time performance of PT service once passengers respond to currently available RTCI. These experimental schemes provide an insight into the potential efficacy of instantaneous RTCI in congested urban networks, indicating some of its advantages and shortcomings.

Our study focuses on representing the impacts of RTCI systems which have not been yet captured in state-of-the-art assignment models. Within our framework, crowding information is directly collected, disseminated and then utilised by PT passengers. This also allows to analyse the potential consequences of providing RTCI to passengers without considering prior crowding experience. Understanding this can be relevant especially in the context of untypical network conditions such as service disruptions or unfamiliarity with the PT system.

The remainder of this paper is organised as follows. Section 3 describes the methodology of this study, introducing the general modelling framework and requirements, followed thereafter by a detailed mathematical formulation of the proposed RTCI algorithm. Section 4 presents results – firstly on simple toy networks which demonstrate the overall capabilities of the RTCI algorithm, and secondly from its application to a city-scale public transport model (Kraków, Poland). Section 4 draws conclusions, in terms of methodological aspects and implementation findings. We conclude by highlighting the practical implications of our works and indicating directions for further research works.

## 2. Method

### 2.1. Modelling requirements

The representation of crowding information effects requires an assignment framework that models the PT system performance in an explicit, disaggregate way. Such a model needs to cover the wide spectrum of dynamic interactions between individual PT system components which are associated with the (over)crowding phenomena. Firstly, on the supply side, how vehicle dwell times at stops increase with passenger flows (boarding and alighting volumes) and with in-vehicle crowding levels (which are bounded by explicit capacity limits). Secondly, on the demand side, how crowding influences the travel utility perceived by passengers (i.e. decrease in perceived travel comfort with rising volume-to-capacity ratio), and in severe cases leads to denied boarding (once capacity constraints become binding). Finally, the model should capture the mutual demand-supply interdependencies in the PT

**Table 2.** Notation.

| Symbol                          | Definition   |
|---------------------------------|--|
| <i>Sets</i>                     |  |
| $G(S, E)$                       | PT network directed connected graph  |
| $S$                             | Set of nodes (stops) $s$   |
| $E$                             | Set of links (trip segments) $e$   |
| $A$                             | Set of actions $a$   |
| $I_a$                           | Set of paths $i$ , associated with undertaking action $a$  |
| $L$                             | Line, i.e. a set of ordered sequence of stops $L = \{s_1, s_2, \dots, s_n\}$ , served by a set of runs (trips, departures) $r$ |
| $K_{o,d}$                       | OD flow, i.e. set of passengers $k_{o,d}$ travelling from origin $o$ to the destination $d$                                    |
| $K_{r,e}$                       | Passenger load on-board the run (trip) $r$ along the segment $e$   |
| <i>Variables and constants:</i> |  |
| $s$                             | Node, i.e. stop between consecutive links $e^-$ and $e^+$  |
| $e$                             | Link, i.e. trip segment between consecutive tail stop $s^-$ and head stop $s^+$  |
| $r$                             | Run (trip, departure) belonging to line $r \in L$  |
| $a$                             | Alternative action at network node $s$   |
| $a_{k,s}$                       | Action chosen by passenger $k$ at network node $s$   |
| $i$                             | (Downstream) path from network node $s$ to destination $d$ ; $i = \{s, s_1, s_2, \dots, s_n, d\}$                              |
| $k_{o,d}$                       | Passenger (agent) $k$ belonging to OD flow $K_{o,d}$   |
| $o$                             | Origin node  |
| $d$                             | Destination node   |
| $t_{a,k}$                       | Decision time instance of action $a$ choice by passenger $k$   |
| $t_{r,e}$                       | Entering time instance of run $r$ at segment $e$   |
| $p_{a,k}$                       | Choice probability of action $a$ for passenger $k$   |
| $u_{i,k}$                       | (Expected) utility of path $i$ for passenger $k$   |
| $t_e^{ivt}$                     | (Expected) in-vehicle travel time of trip segment $e$  |
| $t_e^{wt}$                      | (Expected) wait time at the tail stop $s^-$ of trip segment $e$  |
| $t_e^{wkt}$                     | (Expected) walk time to the tail stop $s^-$ of trip segment $e$  |
| $n_i^{tr}$                      | (Expected) number of transfers along path $i$  |
| $\beta_e^x$                     | (Expected) coefficient of utility component $x$ at trip segment $e$  |
| $z$                             | Monetary valuation factor of travel time $t$   |
| $\beta_{r,e}$                   | Recorded RTCI value of run $r$ along trip segment $e$  |
| $\beta_{l,e}(t)$                | Generated RTCI value of line $L$ along trip segment $e$ , valid at time $t$  |
| <i>Input parameters:</i>        |  |
| $\eta_{r,e}$                    | Seat capacity of vehicle run $r$ (along trip segment $e$ )   |
| $\kappa_{r,e}$                  | Crush capacity of vehicle run $r$ (along trip segment $e$ )  |
| <i>Outputs (KPIs):</i>          |  |
| $w$                             | Passenger welfare, i.e. generalised passenger travel cost in monetary terms  |
| $c^0$                           | Share of pass. decisions with accurate RTCI  |
| $c^+$                           | Share of pass. decisions with inaccurate RTCI (crowding overestimation)  |
| $c^-$                           | Share of pass. decisions with inaccurate RTCI (crowding underestimation)   |

network. This pertains to e.g. how fluctuations in passenger flows and deteriorating service regularity reinforce each other in overcrowded PT networks (i.e. the well-known bunching phenomenon).

Another modelling requirement pertains to reproducing the working principles of an RTCI system in a PT network. These comprise the processes (mechanisms) of generating, disseminating and utilising the crowding information. The main challenges here relate to modelling how crowding levels of PT vehicles are recorded in real-time; how crowding information is generated and updated instantaneously from the information sources available, i.e. real-time and/or historical crowding data (and possibly – simulated future prediction); and how crowding information is mapped to (user-tailored) Advanced Travel Information System framework and then disseminated in real-time across the whole PT network.

Finally, the PT simulation framework should reliably reproduce the acquisition and utilisation of RTCI on the demand side. This relates to modelling the ubiquitous (system-wide) reaction of passengers currently travelling in the PT network and their responsiveness to the RTCI (i.e. penetration rate, choice sensitivity). This task is behaviourally challenging and requires a detailed consideration of passengers' decision-making process, where RTCI is interpreted and traded off against their own expectations of other travel attributes. Importantly, dynamic and sequential properties of path choice process should be satisfied, as RTCI can be received both pre-trip or en-route and potentially influence passengers' choices at any stage of their journey. Since RTCI becomes strictly valid for a specific spatiotemporal trip instance, passengers' decisions are determined by the currently available crowding advice and travel paths towards their destination. Each passenger should be represented in a way that at any journey stage he/she may either follow the current path, or reconsider it and take a re-routing decision instead. Such shifts in travel patterns can occur instantaneously, e.g. in response to the latest update on PT network (over)crowding provided by the RTCI.

Consequently, representing these phenomena requires a combination of local formulas to calculate choice probabilities in the presence of RTCI, and embedding them within a network simulation model to evaluate their system-wide impact. At the passenger level, the aim is to simulate how the newly obtained crowding information – an additional factor in the decision process – influences the expected utility of travel choices and the resultant travel decisions. Meanwhile, at the network level, the focus is on estimating passenger flows and travel costs that are an aggregate outcome of all these individual choices. Representing the complexity of RTCI effects will be only feasible within a PT assignment model that can properly analyse all these properties of PT operations and passengers' travel choices.

## 2.2. PT agent-based simulation model platform

The mesoscopic agent-based *BusMezzo* public transport (PT) simulation model (Toledo et al. 2010; Cats 2011; Cats et al. 2011) is utilised in this study to model the instantaneous effects of passengers' responses to RTCI in PT networks. The *BusMezzo* model has been already applied to examine the impacts of RTI on travel times (Cats et al. 2011; Cats and Jenelius 2014), PT capacity reductions (Cats and Jenelius 2018) and (over)crowding in PT system (Cats, West, and Eliasson 2016; Drabicki et al. 2017a) on route choices and is capable of replicating the three principal categories of crowding effects: denied boarding; on-board discomfort; and irregular vehicle arrivals. Capacity limits are explicitly observed (Cats, West, and Eliasson 2016; Gavrilidou and Cats 2019), which implies that excessive passenger flows are strictly denied the boarding beyond the assumed crush capacity limit  $\kappa_{r,e}$ . In-vehicle travel discomfort is influenced by rising volume-to-capacity ratio, distinguishing travel disutility experienced by standing and seated passengers. (Further elaboration is given in the Subsection 3.4.) The model assumes that passengers prefer to sit, utilising the available seat capacity  $\eta_{r,e}$  before standing on-board. Seating priority rules are observed, meaning that on-board passengers are able to take a seat before boarding passengers, and those travelling further downstream have a priority to sit over passengers alighting earlier. Finally, the progression of PT vehicles is determined by flow-dependent dwell times at stops. This implies that fluctuating (and rising) passenger flows may impede service regularity and

induce a negative feedback phenomena such as e.g. bus bunching (Moreira-Matias et al. 2016).

The model essentially consists of PT supply (transport network and PT lines operating on this network) and PT demand (passengers travelling between their origins and destinations), along with mutual dynamic interactions between them (Cats, West, and Eliasson 2016; Laskaris et al. 2019) (Table 2).

The PT network is represented by a directed connected graph  $G(S, E)$ , where  $S$  is the set of nodes – corresponding to stops and  $E$  is the set of links representing line (trip) segments and walking links (access, egress and transfer links). Line segment  $e$  of a PT line  $L$  connects its tail stop  $e^- \in S$  with its head stop  $e^+ \in S$ . A PT line  $L$ , in turn, is defined as an ordered sequence of stops  $L = (s, s_1, s_2, \dots, s_n)$  or, equivalently, as an ordered sequence of line segments  $L = (e, e_1, e_2, \dots, e_{n-1})$ . While a stop can be served by several lines, a segment is uniquely assigned to the specific line. Optionally, walking link  $e$  stretches between tail stop  $e^- \in S$  and head stop  $e^+ \in S$  of different PT lines to allow for transfers between stops.

The time-dependent PT service supply is modelled through line runs (departures)  $r$  dispatched from the first stop and serving consecutive stops of line  $L$ . Each run is operated by a vehicle characterised by limit values of seat capacity  $\eta_{r,e}$ , and the so-called crush capacity  $\kappa_{r,e}$ , i.e. maximum number of passengers allowed on-board. Travel time of run  $r$  consists of two components: riding times  $t_{r,e}$  along line service segments and dwell times  $t_{r,s}$  at stops. Segment ride times in this study are assumed fixed, while stop dwell times are flow-dependent, namely passenger boarding and alighting flows and in-vehicle crowding levels. Consequently, while vehicles are dispatched from the first stop according to the fixed timetable, departure times from consecutive run segments  $t_{r,e}$  are influenced by the evolving network performance and may differ from the nominal schedule.

The PT passenger demand is represented by an OD matrix at the zone-level through  $K_{o,d}$  – the number of passengers travelling from the origin stop  $o$  to the destination stop  $d$  during a given time period. Passengers  $k_{o,d}$  are initiated at an origin stop  $o$  according to a given passenger inflow distribution (we assume a constant passenger inflow rate during the simulation) and assigned to a destination stop  $d$ . Each passenger  $k$ , while making decisions, considers taking an action  $a$ , i.e. travel choice that involves travelling along the path set  $I_a$ . This path set consists of multiple individual paths, where path  $i$  is essentially a sequence of stops (or equivalently, line segments) connecting his/her current decision-making point  $s$  with his/her destination stop  $s_d$ , that is  $i = (s, s_1, s_2, \dots, s_n, s_d)$ . The path may consist of multiple line segments, transfer connections between lines within stops and/or walking links between stops. The approach in the *BusMezzo* model is to define path  $i$  as a set of all the alternatives that imply the same chain of stops with equivalent link attributes (Cats, West, and Eliasson 2016; Cats and West 2020). Additionally, we relax the default dominance and filtering rules, in order to increase the choice set size and include additional paths that might only become attractive with access to RTCL.

The *BusMezzo* is an event-based choice model, with passengers' choices determined sequentially as they progress through the network. At each decision node  $s$ , passenger  $k$  makes a decision involving successive action  $a_{k,s}$  out of possible action set  $A$ . These decisions are grouped into connection, boarding and alighting decisions. At the origin, each passenger chooses which of the available stops to walk to (connection decision). Each time a vehicle arrives at the stop, each waiting passenger decides whether to board it or stay and wait for another vehicle (boarding decision). Each time before the vehicle arrives at next

downstream stops, passengers on-board decide whether to alight from it or stay on-board and ride further (alighting decision). Each passenger who chooses to alight may decide to walk to another stop or stay and wait at the same stop – (another connection decision instance). While evaluating the boarding decision, passenger considers all the possible downstream paths associated with boarding a given PT run and compares them with paths available if remaining at this stop. Similarly, while evaluating the alighting decision, passenger considers all the possible downstream paths available at alighting stop and compares them with paths available if staying in the vehicle.

The resulting origin-destination path is an outcome of the sequence of actions  $a_{k,s}$  undertaken at consecutive decision points  $s$  at decision time instances  $t_{a,k}$ . Consequently, paths are not predetermined pre-trip – but rather adaptive, i.e. may evolve as passengers respond to travel conditions changing en-route. Importantly, the dynamic and sequential decision-making pattern of the path choice model implies that all of these decisions might be reconsidered en-route, as traveller may obtain up-to-date travel information on current or anticipated PT service conditions.

The choice model is based on the utility maximisation principle, i.e. passengers aim to minimise the perceived travel cost of travel action  $a$  (which is equivalent with the aim to maximise the perceived utility of travel action  $a$ ), and thus to minimise the travel cost of remaining path  $i$  of their journey. The perceived travel cost includes time components (in-vehicle time, wait time, walk time) and other aspects (e.g. number of transfers, in-vehicle crowding discomfort). In this paper, we represent this decision process with a discrete choice model, namely a multinomial logit model (MNL), but alternative models are also applicable.

The utility  $u_{a,k}$  of action  $a$  for passenger  $k$  is obtained as the utility of path set  $I_a$  associated with that particular action, in case of MNL expressed as follows:

$$u_{a,k} = \ln \sum_{i \in I_a} \exp(u_{i,k}). \quad (1)$$

The path utility  $u_{i,k}$  is, in turn, sum of systematic part of path utility  $v_{i,k}$  plus a random error term component  $\varepsilon_k$  perceived by passenger  $k$ :

$$u_{i,k} = v_{i,k} + \varepsilon_k = \sum_{e \in i} \beta_{e,k}^x \cdot t_e^x + \varepsilon_k. \quad (2)$$

Systematic path utility component is the sum of expected travel time attributes  $t_e^x$  of trip components  $x$ , multiplied by their relative perceived (dis)utility coefficients  $\beta_e^x$  (for the sake of simplicity, we assume uniform disutility coefficient values across the whole demand population, i.e.  $\beta_{e,k}^x = \beta_e^x$ ). Essentially, this comprises total expected travel utility related to in-vehicle travel time  $t_e^{ivt}$ , wait time  $t_e^{wt}$ , walk time  $t_e^{wkt}$  of all trip segments  $e$  belonging to path  $i$ , plus the number of transfers  $n_i^{tr}$  along path  $i$ :

$$v_{i,k} = \sum_{e \in i} \beta_e^{ivt} \cdot t_e^{ivt} + \sum_{e \in i} \beta_e^{wt} \cdot t_e^{wt} + \sum_{e \in i} \beta_e^{wkt} \cdot t_e^{wkt} + \beta^{tr} \cdot n_i^{tr}. \quad (3)$$

From Equation (3) one can distinguish perceived journey time (PJT) disutility components associated with in-vehicle time (IVT) disutility  $\sum_{e \in i} \beta_e^{ivt} \cdot t_e^{ivt}$ , wait time (WT) disutility  $\sum_{e \in i} \beta_e^{wt} \cdot t_e^{wt}$ , walk time (WKT) disutility  $\sum_{e \in i} \beta_e^{wkt} \cdot t_e^{wkt}$  and transfer penalty (TRP) disutility



$\beta^{tr} \cdot n_i^{tr}$  of path  $i$ . We assume here perceived disutility coefficients of wait time  $\beta_e^{wt}$  and walk time  $\beta_e^{wkt}$  equal to double the perceived disutility coefficient rate of uncrowded in-vehicle travel time, i.e.  $\beta_e^{ivt} = -1.0$  and  $\beta_e^{wt} = \beta_e^{wkt} = -2.0$ . Perceived transfer disutility coefficient  $\beta^{tr}$  is assumed to be equivalent to extra five minutes of the uncrowded in-vehicle travel time, i.e. every transfer imposes additional disutility equal to the  $\beta_e^{ivt} = -1.0$  multiplied by 5 min. These weights are assumed according to state-of-the-art findings (de Dios Ortuzar and Willumsen 2011; Gentile and Noekel 2016; Cats, West, and Eliasson 2016; Yap, Cats, and van Arem 2020). Since all perceived utility coefficients (Equation (3)) are negative, the resultant path utility is a negative value. Crucially, the  $\beta_e^{ivt}$  rate is itself a dependent variable of real-time crowding information (RTCI), as explained in the Subsection 2.4. below.

Finally, passenger at a given decision point  $s$  makes one of possible decisions  $a_{k,s} \in A$  according to the probabilistic MNL formula:

$$p_{a_{k,s}} = \frac{\exp(u_{a_{k,s}})}{\sum_{A' \in \Delta} \exp(u_{A'})} \quad (4)$$

Microscopic decisions of single agents can be aggregated to obtain passenger flows  $K_{o,d}$  in the network, crucial for estimating on-board crowding conditions. At each decision point, the number of passengers deciding to board  $K_{r,s}^b$  and alight  $K_{r,s}^a$  at consecutive stops  $s$  of vehicle run  $r$  can be determined as an aggregate outcome of multiple agents' decisions.

Network-wide total travel costs are measurable in terms of passenger welfare  $w$ , expressing total travel utility of all passengers  $K_{o,d}$  weighted by monetary travel time valuation factor  $z$ :

$$w = -z \cdot \sum_{k \in K_{o,d}} u_{i,k} \quad (5)$$

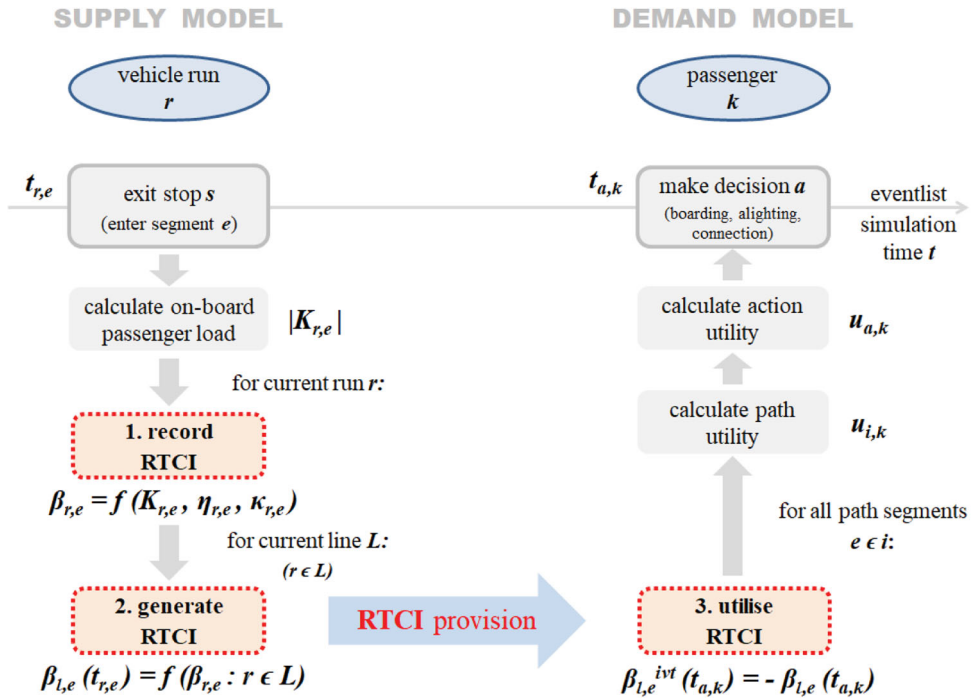
Passenger welfare  $w$  is understood here as the total generalised travel cost under the fixed-demand assumption (Cats, West, and Eliasson 2016) and, in such form, does not account for induced and/or suppressed passenger demand.

In the remainder of this section, we describe the new modelling functionalities, introduced in the *BusMezzo* model to enable the representation of instantaneous RTCI phenomena and their consequences for PT system performance.

### 2.3. Simulating real-time crowding information (RTCI)

Representation of the RTCI impact on the supply and demand sides that we incorporate in our model comprises of three main aspects (Figure 1):

- initially, how the crowding information of PT vehicles is recorded (observed) in real-time as they propagate through the network;
- subsequently, how the recorded in-vehicle crowding data is used to generate crowding information disseminated in real-time to passengers;
- finally, how passengers acquire and utilise the real-time crowding information, in order to update the expected utility of considered path options, which may impact their travel decisions.



**Figure 1.** Flowchart of the instantaneous RTCI algorithm, as implemented in the BusMezzo model.

A central variable of the introduced RTCI model is  $\beta_{l,e}^{RTCI}$  – **the in-vehicle passenger crowding rate**. This variable can be understood as the value-of-time multiplier due to increased (over)crowding conditions (Table 3), which is equivalent to negative value of in-vehicle travel time (dis)utility co-efficient  $-\beta_{l,e}^{ivt}$ . This is a common crowding variable across the consecutive stages of generating, disseminating and utilising the crowding information:

















1. **Recording the RTCI:** Firstly, the in-vehicle crowding level  $\beta_{r,e}$  is recorded at each stop-exit time instance  $t_{r,e}$ , i.e. whenever run  $r$  enters a line segment  $e$  (or, equivalently, whenever it departs from the tail stop  $s = e^-$ ):

$$\beta_{r,e} = f(|K_{r,e}|, \eta_{r,e}, \kappa_{r,e}) \quad (6)$$

The  $\beta_{r,e}$  is a function of the vehicle seat capacity  $\eta_{r,e}$ , total capacity (so-called *crush capacity*)  $\kappa_{r,e}$  and passenger load  $|K_{r,e}|$  on-board a given run (trip) segment. In this paper, we use a step-wise RTCI mapping scale, defined in detail further below (Table 3).

2. **Generating the RTCI:** Subsequently, the recorded crowding levels  $\beta_{r,e}$  are used to generate crowding information that is to be disseminated to passengers. As stated earlier, information is generated and available for passengers at the line-segment level  $\beta_{l,e}$ . In principle, the generated crowding information of line segment  $\beta_{l,e}$  at time  $t$  may be a function of crowding levels for all runs of line  $r \in l$  that have already traversed the line

**Table 3.** The assumed RTCI mapping function – 4-level crowding scale parameterisation. Right column contains crowding factors that are used to evaluate anticipated and experienced IVT disutility.

| Descriptive interpretation of on-board crowding conditions:  | Volume-capacity ratio with respect to: |                                       | RTCI factor<br>$\beta_{r,e}, (\beta_{l,e})$ |
|--|--|---------------------------------------|---|
|  | Seat cap. $ K_{r,e}  / \eta_{r,e}$     | Crush cap. $ K_{r,e}  / \kappa_{r,e}$ |   |
|     <b>Uncrowded conditions</b> (plenty of seats available), $\leq \sim 80\%$ seat capacity                  | $\leq 0.8$                             | $\leq 0.8$                            | 1.0   |
|     Individual seats available, $\leq 100\%$ seat capacity   | $\leq 1.0$                             |                                       | 1.2   |
|     Need to stand, but in comfortable conditions, $\leq 80\%$ crush capacity                                 | $> 1.0$                                |                                       | 1.5*  |
|     <b>Overcrowded conditions</b> (excessive crowding, denial-of-boarding risk), $\leq 100\%$ crush capacity |  | $\leq 1.0$                            | 1.8*  |

Note: \* 1.2 for experienced IVT disutility if passenger is seated

segment  $e$ , and thus for which the crowding on segment  $e$  was already recorded, that is:

$$\beta_{l,e}(t) = f(\beta_{r,e} : r \in l, t_{r,e} \leq t) \quad (7)$$

The specific function to apply may either differ from simple instantaneous updates, through smoothing the multiple recent runs (Drabicki et al. 2017a), up to fusing with historical data or applying predictive models. In this paper, we assume that generated information of line  $\beta_{l,e}$  at time  $t$  is equal to the recorded information of the latest (most recent) vehicle run  $r$  that has traversed line segment  $e$  – i.e. **instantaneously** updated RTCI:

$$\beta_{l,e}(t) = \beta_{r,e} ; r \in l, t_{r,e} \leq t \wedge t_{r+1,e} > t \quad (8)$$

The instantaneously generated RTCI  $\beta_{l,e}$  valid at time  $t$  is then disseminated across the whole PT network as a uniform value for all passengers.

3. **Utilising the RTCI:** Once the generated RTCI of a given network segment  $e$  is being disseminated, it is utilised by passenger  $k$  who considers travelling along that segment at decision time instance  $t_{a,k}$ . Hereby, we assume that 100% of passengers incorporate the RTCI provisioned in their decision making process. RTCI information becomes utilised whenever the passenger takes an action  $a$  associated with boarding, alighting or making a connection decision. Once path utility calculation process is triggered, passengers may acquire the up-to-date crowding information  $\beta_{l,e}$  valid at time  $t$  for the currently considered segment  $e$  of line  $L$ . Passengers utilise crowding information by updating the expected utility of considered path  $u_{i,k}$ . Technically, this is achieved through applying  $\beta_{l,e}^{ivt}(t) = -\beta_{l,e}(t)$  as a perceived in-vehicle travel time (IVT) disutility multiplier in the utility formula (Equation (3)):

$$v_{i,k}(t_{a,k}) = \sum_{e \in i} \beta_{l,e}^{ivt}(t_{a,k}) \cdot t_e^{ivt} + \sum_{e \in i} \beta_e^{wt} \cdot t_e^{wt} + \sum_{e \in i} \beta_e^{wkt} \cdot t_e^{wkt} + \beta^{tr} \cdot n_i^{tr} \quad (9)$$

These three stages constitute the proposed modelling framework of RTCI systems in PT network, whereby crowding is recorded in vehicles  $\beta_{r,e}$ , disseminated to passengers at the line-segment level  $\beta_{l,e}$ , and used to update the action (path) utility with a crowding discomfort rate  $\beta_{l,e}^{ivt}(t)$  (Figure 1). This is done in accordance with state-of-the-practice of modelling the PT crowding effects, whereby crowding penalty is included as an additional in-vehicle travel time (IVT) multiplier in passenger path choice model (e.g. Tirachini, Hensher, and Rose 2013; Gentile and Noekel 2016; Yap, Cats, and van Arem 2020). Additionally, we assume ubiquitous coverage of RTCI (i.e. 100% penetration rate among PT passengers), but our modelling framework allows to specify a variable RTCI penetration rate. In practice, the share of passengers who utilise RTCI in their path choices may depend on a variety of factors, such as the availability of broadcasting (dissemination) means, the technology of RTCI system and perceived information reliability.

The form and parameterisation of the RTCI-related functions as defined in Equations (6) – (9) will determine the ultimate impact of the RTCI. For instance, Equation (6) represents how the actual in-vehicle crowding is translated into crowding information disseminated to users. Here, we propose a discrete, 4-level crowding information scale (Table 3), with the baseline value  $\beta_{l,e} = 1.0$  denoting uncrowded conditions, and step-wise increasing  $\beta_{l,e}$  value as a function of rising on-board crowding conditions, reaching a maximum value of  $\beta_{l,e} = 1.8$  for overcrowded conditions when passenger load is approaching the vehicle capacity limit. Assumptions regarding the 4-level crowding information scale and corresponding  $\beta_{l,e}$  values are inferred based on multiple literature sources on crowding valuations in PT journeys (as summarised previously in the Subsection 1.1.), especially the works of Rudnicki (1999), Whelan and Crockett (2009), Tirachini, Hensher, and Rose (2013), Kroes et al. (2014), Hörcher, Graham, and Anderson (2017), Yap, Cats, and van Arem (2020). Our proposed  $\beta_{l,e}$  values (Table 3) are slightly lower than average, accounting for the fact that majority of studies contain valuations mostly from SP studies (prone to certain exaggeration bias, especially at higher overcrowding levels) and rail systems (where crowding impact is presumably more significant due to longer trip distance/duration) – whereas our focus is predominantly on urban PT networks.

The  $\beta_{r,e}$  values presented in Table 3 are also used for evaluating the users' experience of on-board journey conditions, i.e. IVT disutility (Equation (3)). Importantly, we distinguish the travel disutility experienced differently by seated vs. standing traveller inside the same vehicle. Hence, the maximum *experienced* IVT disutility multiplier is  $\beta_{r,e} = 1.2$  if passenger is seated, while standing overcrowding conditions may imply a max.  $\beta_{r,e} = 1.8$  for the traveller that is standing on-board the same run  $r$  along segment  $e$ .

The modelling framework formulated above enables us to simulate RTCI and its impact on passenger path decisions. Integrating RTCI into passenger behaviour model alters their behaviour which, in turn, yields shifts in passenger flows and, consequently, in the overall network performance. To assess how RTCI impacts PT network performance, two alternative measures are hereby proposed.

The first measurement of network-wide system performance relates to **passengers' travel experience**, measureable with path utility (Equations (1) – (3)) or travel welfare (Equation (5)). These can be measured either as a total (global) value (Equation (2)), and/or through their decomposition into specific trip stages as shown in (Equation (3)).

The second aspect measuring the RTCI system performance is the **crowding information accuracy**. Crowding information provided to passengers and supporting their decisions

should be credible and consistent with the actual travel conditions they eventually experience. We measure the consistency between information on projected crowding conditions  $\beta_{l,e}(t_{a,k})$  along segment  $e$  of line  $L$  – disseminated to the passenger  $k$  and utilised by him/her while making a decision at action time  $t_{a,k}$  – against the actual crowding conditions  $\beta_{r,e}(t_{r,e})$  – observed by passenger once on-board the run  $r$  traversing that particular segment  $e$ .

Accuracy is evaluated for each single line segment  $e$  traversed by a passenger during his/her trip, then aggregated for the whole path  $l$ , and finally computed as a network-wide RTCI accuracy rate  $c^0$  for all the passengers  $k \in K$ . Due to the sequential decision-making process and the recurrently updated path utility at each decision point, measuring accuracy of all decisions is complex. Here, we apply certain heuristics and propose to consider RTCI accuracy only for decision instances directly preceding passengers' boarding (or staying on-board) actions. More intuitively, information accuracy experienced along the line segment  $e$  is measured only for such decision node  $s = e^-$ , which is the last possible instance at which the passenger is able to consider the RTCI of this line segment  $e$  (i.e. boarding/alighting decision involving entering a given line segment  $e$  on-board a vehicle run  $r$ ).

A simple yet sufficient RTCI accuracy metric is the share of *accurate* passenger decision instances  $c^0$  for which projected crowding information  $\beta_{l,e}(t_{a,k})$  was in agreement with observed crowding conditions  $\beta_{r,e}(t_{r,e})$  (Equation (10)). Complementary values  $c^-$  and  $c^+$  denote then share of *inaccurate* passenger decision instances where RTCI underestimated or overestimated the observed crowding conditions, respectively:

$$\begin{aligned} c^+ &= \frac{\sum_{k \in K} \sum_{e \in k} \delta_e^+}{\sum_{k \in K} \sum_{e \in k} (\delta_e^+ + \delta_e^0 + \delta_e^-)}, \delta_e^+ = \begin{cases} 1 & \text{if } \beta_{l,e}(t_{a,k}) > \beta_{r,e}(t_{r,e}) \\ 0 & \text{otherwise} \end{cases} \\ c^0 &= \frac{\sum_{k \in K} \sum_{e \in k} \delta_e^0}{\sum_{k \in K} \sum_{e \in k} (\delta_e^+ + \delta_e^0 + \delta_e^-)}, \delta_e^0 = \begin{cases} 1 & \text{if } \beta_{l,e}(t_{a,k}) = \beta_{r,e}(t_{r,e}) \\ 0 & \text{otherwise} \end{cases} \\ c^- &= \frac{\sum_{k \in K} \sum_{e \in k} \delta_e^-}{\sum_{k \in K} \sum_{e \in k} (\delta_e^+ + \delta_e^0 + \delta_e^-)}, \delta_e^- = \begin{cases} 1 & \text{if } \beta_{l,e}(t_{a,k}) < \beta_{r,e}(t_{r,e}) \\ 0 & \text{otherwise} \end{cases} \end{aligned} \quad (10)$$

The distinction made in Equation (10) allows evaluating the relative impact of an under- or over-estimation of the actual on-board crowding conditions on the perceived RTCI accuracy. Arguably, the former (i.e. underestimation risk, denoted as  $c^-$ ) has particularly negative implications for passengers' travel experience. In such instances, travel conditions turn out to be more adverse than anticipated. This might especially undermine the credibility of RTCI accuracy, and in a broader perspective, call into question the passengers' eventual trust and therefore reliance on crowding information.

Ultimately, the effectiveness of RTCI system in PT simulation can be measured as follows:

1. The first criterion is to maximise passenger welfare as defined in Equation (5).
2. The second criterion is to maximise the share of accurate passenger decisions as given in Equation (10), and especially, to minimise the share of inaccurate decisions due to crowding underestimation.

### 3. Application

#### 3.1. Simulations' outline

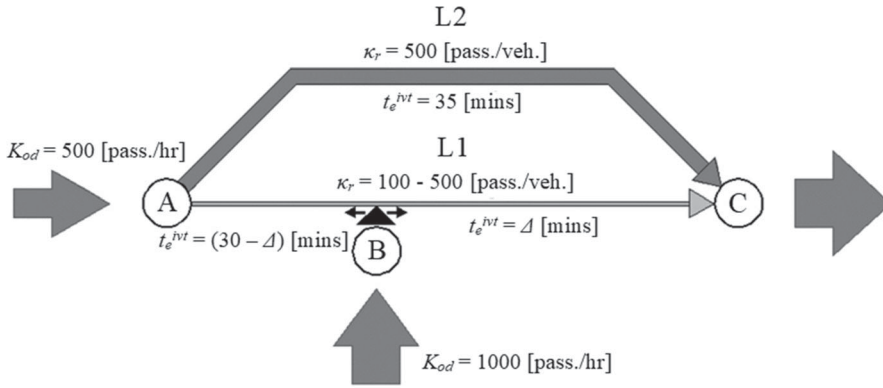
To illustrate how the proposed algorithm models the impact of instantaneous real-time crowding information (RTCI) we introduce the following experimental simulations. The experiments are designed to test the validity of the proposed RTCI algorithm and its capability of representing a chain of reactions likely to emerge with the RTCI system active in PT networks – specifically:

- (1) how crowding information (RTCI) is generated from PT supply data in real-time – and then instantaneously disseminated to passengers,
- (2) whether (and how) passengers' choice patterns change (either pre-trip and/or en-route) and whether they shift to alternative, less-crowded paths once they acquire information about network crowding along their paths,
- (3) how passengers' reactions interplay with variable rates of crowding conditions and travel times on alternative lines,
- (4) whether the RTCI allows to improve passengers' utility and thus overall system performance and travel welfare (i.e. a combined outcome of both travel times and travel comfort),
- (5) what is the resultant accuracy of RTCI, and whether the dissemination time lag of instantaneous RTCI leads to negative consequences for passenger utility and system welfare.

The experimental demonstration involves a set of 2 toy PT networks with analogous network layout. For each toy-network scenario, we analyse how RTCI influences the results with respect to the reference ('no RTCI') case with identical demand and supply assumptions – except for the RTCI availability: in the 'no RTCI' case, passengers evaluate choice utilities without crowding information; whereas in the 'RTCI' case, we assume ubiquitous response to crowding information, i.e. 100% of passengers have access to and utilise the RTCI, and choice utilities are evaluated with the extended path choice model. Since passengers do not have any prior expectations of (over)crowding emerging downstream, toy-network experiments illustrate how they respond to non-anticipatory RTCI information, representing e.g. sudden network events or travellers unfamiliar with the PT system.

Since the simulations are stochastic, outputs are reported as an average of 10 replications with randomised seed (a sufficiently representative sample with acceptably low variability in results). We simulate a period of two hours, with supply model (PT lines) operating for the whole period, and demand being generated from the origin(s) during an intermediate 60-minute period – i.e. starting at the 30th minute and lasting until the 90<sup>th</sup> minute. Seat capacity  $\eta_{r,e}$  is assumed uniformly for all PT vehicles, equal to 60% of the crush capacity  $\kappa_{r,e}$ .

In Subsection 3.3., we run experiments on a real-size PT network model of the urban PT system in Kraków, Poland to simulate, quantify, measure and visualise RTCI impact. Simulation is performed for a typical weekday PM peak hour and depicts impact of RTCI in practical setting indicating both its benefits and shortcomings.



**Figure 2.** Topology of the first toy network. Service headways: L1 and L2 – every 5 min.

### 3.2. Toy network experiments

#### 3.2.1. Toy network no. 1

Figure 2 presents a simple PT network, comprised of two O-D pairs (A-C and B-C) and two PT lines (L1 and L2). L1 line is a direct service connecting A and B origin stops with the same destination stop C, with a total travel time of 30 min. L2 line serves the A-C pair only and has a longer travel time (35 min) but much higher capacity (well above the demand). Both lines operate at a service headway of 5 min. L1 line is served by PT vehicles of low capacity of 100 passengers per vehicle, except for an intermediate peak period (ca. 20 min long) when vehicles of higher capacity are used (500 passengers per vehicle). The origin stop A generates demand of 500 passengers per hour, while the busier origin stop B generates a higher demand volume of 1,000 passengers per hour.

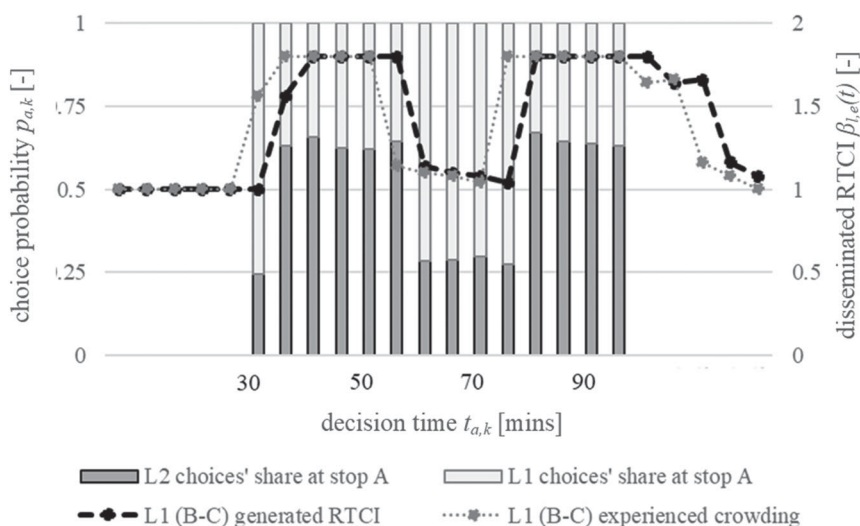
In the following experiments, we investigate how RTCI provision affects simulation output – in accordance with the methodology introduced above. Firstly, the number of passengers  $|K_{r,e}|$  is observed for consecutive runs  $r$  of lines L1 and L2 along their line segments  $e$ . Secondly, by comparing on-board passenger volumes against vehicle capacity and applying the mapping procedure (Table 3), the recorded in-vehicle crowding levels  $\beta_{r,e}$  are obtained (Equation (6)). Thirdly, crowding data from consecutive runs is being instantaneously translated into RTCI further disseminated to passengers  $\beta_{l,e}$  as stated in Equation (8). Finally, passengers become aware of the current RTCI  $\beta_{l,e}$  available at time instance  $t$  and utilise this information at every decision stage (Equation (9)).

The following points summarise the main observations from our experiments on the first toy-network.

#### (a) From real-time crowding information (RTCI) to passengers' choices:

Figure 3 depicts the utilisation phase of RTCI – passenger choices at the origin stop A that result from RTCI generated in a given time instance. Without availability of RTCI, passenger choices remain constant throughout simulation time, with less than 25% of origin A demand taking the L2 route. Their decision is solely based on the expected shorter travel time of the L1 line that becomes their preferred choice, as crowding conditions along





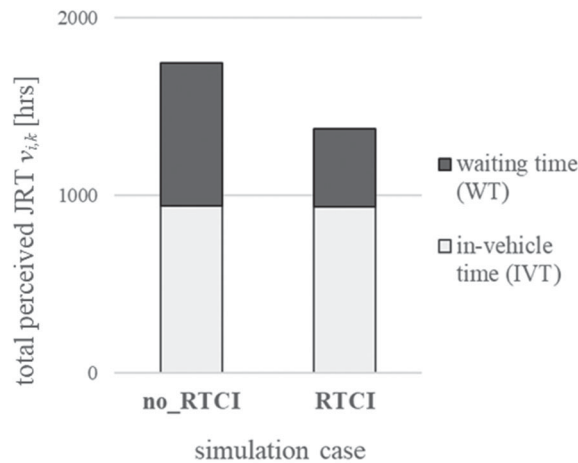
**Figure 3.** Generated RTCI and passenger choices at origin stop A. Dark bars denote the patronage rate of a longer, but uncrowded line L2. Initially low, it rises substantially once passengers receive RTCI on overcrowding along L1 line, then falls back again when overcrowding subsides along L1 (due to increased capacity) and rises back again thereafter. Thick black line denotes RTCI available at stop A during passengers' decisions, while dotted grey line denotes actual crowding experienced later.

that route are not considered by them. Figure 3 demonstrates how their decisions are influenced by availability of RTCI. Passengers are informed about severe in-vehicle crowding ahead, i.e. along the (B-C) segment, and incorporate this knowledge in the path utility formula, where the relative disutility of L1 in-vehicle travel time climbs up by a  $\beta_{l,e}$  multiplication factor of 1.5 - 1.8. Consequently, this leads to a significant rise in the L2 line patronage from 25% up to even 60% – implying that the longer but uncrowded L2 route is now the preferred choice for origin A demand. In the short intermediate period when larger PT vehicles are in operation, overcrowding on the L1 line subsides and its  $\beta_{l,e}$  factor decreases to around 1.0, meaning that L1 line becomes the dominant choice again. Afterwards, once smaller vehicles are back in operation, crowding emerges anew on the L1 (B-C) segment and more passengers at stop A shift towards L2 line again.

(b) *RTCI impact on travel utility:*

Investigation of resultant passenger travel disutility (Equation (4)) shows a ca. 20% reduction in perceived journey time with the introduction of RTCI in the first toy-network (Figure 4). A vast amount of these are attributable to savings in waiting time disutility (WT) which result from a significantly lower occurrence of denial-of-boarding at stop B: as the RTCI on downstream L1 (B-C) crowding is available, passengers at origin A are encouraged to shift towards the L2 route, and thus L1 runs arriving at stop B have now higher residual capacity. Additionally, the longer travel time along the L2 route is compensated by the lower perceived in-vehicle travel time disutility (IVT) experienced by passengers who take this route from stop A; hence, total IVT disutility also decreases, though by a marginal rate. Additionally, RTCI contributes to higher probability of travelling in more comfortable (less





**Figure 4.** Total travel disutility in the first toy-network (perceived journey time). Introducing RTCI reduces both travel costs due to waiting time (45%) and in vehicle time (1%).

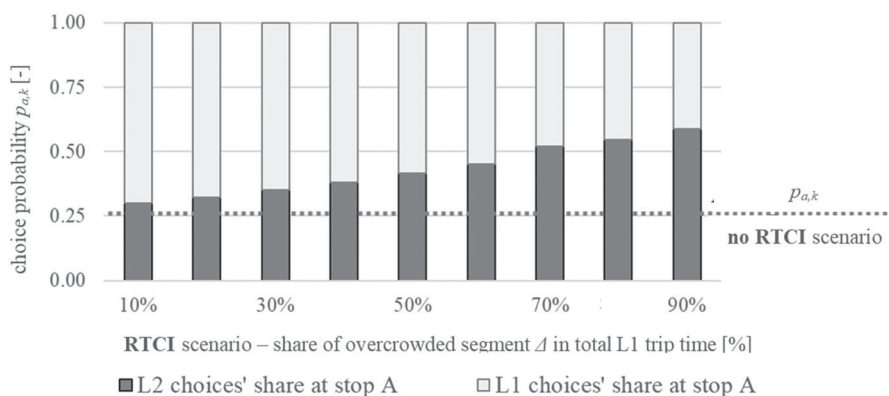
crowded) conditions: for example, the share of total IVT spent in uncrowded conditions, equivalent to  $\beta_{I,e} = 1.0$ , rises from 21% to 32% with the RTCI.

(c) *Passengers' reactions to longer travel time in overcrowding conditions:*

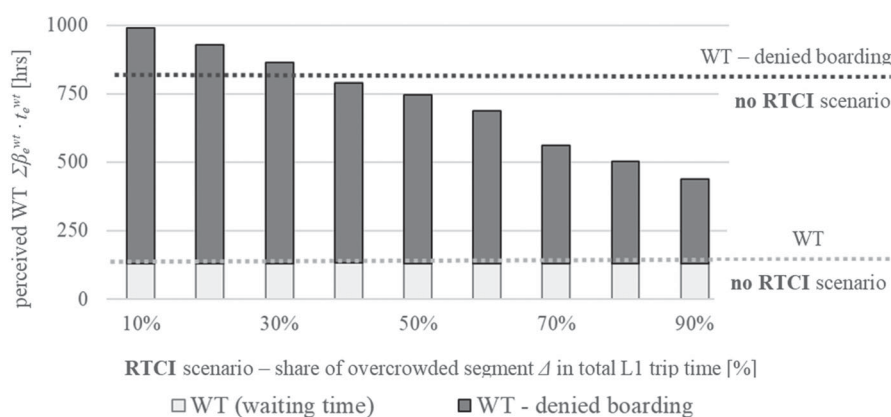
To enrich the above picture, let us illustrate how passenger choices change with expectations of rising travel time in overcrowded conditions. We now assume various lengths  $\Delta$  of the (B-C) segment of line L1, i.e. by shifting gradually the relative position of stop B, it ranges from 10% to 90% of the total L1 trip time. Since the capacity of L1 line along the (B-C) segment is insufficient, especially due to high boarding volumes at stop B, this segment is especially prone to overcrowding. Hence, the L1 travel utility depends on shifting length  $\Delta$  of the (B-C) segment trip time. It can be seen in Figure 5 that as the position of stop B is moved upstream towards stop A and travel time  $\Delta$  in overcrowded conditions along the (B-C) increases, the perceived L1 travel disutility becomes higher, leading to a significant drop in the probability of choosing line L1 at stop A, in comparison to scenario without RTCI. For  $\Delta$  equal to 3 min (i.e. 10% of L1 travel time), RTCI provision already induces an extra 5% shift towards line L2 at stop A, and for the most extreme case of the  $\Delta$  equal to 27 min (i.e. 90% of L1 travel time) – about additional 35% of passengers at stop A opt for L2 line (Figure 5).

(d) *Positive PT network-wide effects:*

While Figure 5 depicts localised impact of RTCI upon choice shifts at origin stop A, we now take a closer look how these can contribute towards the maximisation of total, network-wide travel utility. Figure 6 shows that decisions of passengers at stop A – who experience improved IVT utility themselves – have relevant and positive impact for travel utility of passengers at stop B as well: as fewer passengers select the overcrowded L1 line at stop A, the denial-of-boarding risk at stop B decreases substantially. In contrast, without access to RTCI, boarding denial at stop B has up to 60% greater negative impact for travel



**Figure 5.** Passenger choices at stop A as a function of rising travel time  $\Delta$  in overcrowded conditions along the L1(B-C) segment. Dark bars denote patronage rate of the longer, yet uncrowded L2 line, which is positively correlated with  $\Delta$ : while chosen by ca. 30% of passengers at stop A if only 10% of L1 trip time is overcrowded, this figure rises over 60% when 90% of L1 trip time is overcrowded – compared to a constant L2 probability rate of 25% for the ‘no RTCI’ scenario.



**Figure 6.** Perceived waiting time (WT) disutility at stop B, distinguishing the WT delay due to denial-of-boarding. Dark bars show a substantial decrease in denial-of-boarding delay, as more residual capacity becomes available on the L1 runs - which, in turn, is caused by a lower number of passengers travelling from stop A with the L1 line due to the rising (anticipated) travel time  $\Delta$  in overcrowded conditions.

utility, resulting in an average additional waiting time of max. 6 min per each passenger at stop B (Figure 6).

This simple experiment demonstrates how RTCI information may help alleviate overcrowding in critical parts of the network and improve network performance in wider spectrum – an issue especially important in urban PT networks, where network capacity is often limited due to high passenger volumes already carried from upstream line segments. Such phenomenon is an example of positive ‘ripple-effects’, i.e. wider network benefits induced by localised response to the RTCI. Consequently, downstream passengers willing to board at the latter stops are more likely to experience higher travel discomfort, both

in overcrowded conditions (greater denial-of-boarding risk) as well as in normal network crowding (lower chance of getting a seat).

(e) *Improved vehicle capacity utilisation vs. positive travel experience:*

One of the purported objectives of RTCI system should be to produce a more even utilisation of PT capacity, as information on crowding should encourage passenger flow shifts towards less-crowded paths and reduce loads on the most overcrowded segments. Table 4 presents how resultant changes in travel disutility, measured by perceived journey time (PJT), correlate with the shifting length of L1 (B-C) segment. Importantly, RTCI provision at first translates into worse average journey times. This stems from the fact that passengers at stop A indeed reduce their perceived IVT disutility (i.e. no crowding along L2 line), yet this implies longer journey times (i.e. due to choosing L2 line) and is not compensated enough by initial gains in denial-of-boarding disutility at stop B. However, as L1 (B-C) segment length increases, RTCI becomes more advantageous in terms of both absolute as well as perceived journey times (max. decreases of ca. 10% and 20% respectively). On-board journey conditions improve for all passengers (Table 4) with RTCI: notably, for the final case where (B-C) segment comprises 90% of L1 trip time, share of total perceived IVT in the worst conditions ( $\beta_{I,e} = 1.8$ ) drops from 44% to 37%.

### 3.2.2. Toy network no. 2

To reveal additional phenomena associated with instantaneous RTCI, we introduce a second toy-network with a different topology (Figure 7). It consists now of a single O-D pair only, served by two parallel lines, L1 and L2 – with equal running times (20 min per line segment, 40 min per line in total) but different service frequencies and vehicle capacity. The L1 line is directly connected to the origin (via stop A1) and to the destination (via stop C1). It operates every 10 min with low-capacity vehicles (100 pass./veh.). The parallel L2 line involves an extra 10-minute walk to access its origin stop A2, has a lower headway of 30 min, yet operates with high-capacity vehicles (500 pass./veh.). As in the previous network, seat capacity is assumed equal to 60% of vehicle crush capacity. Additionally, an intermediate transfer is available, connecting both lines at stops B1 and B2, that involves an extra 5-minute walk. Passenger demand is generated at 1500 pass./hour. When RTCI is not available, line L1 would be naturally the preferred travel choice, but passengers becoming aware of downstream overcrowding may migrate towards line L2 – that is, either re-route at the origin (at stop A1) or transfer en-route (at stop B1).

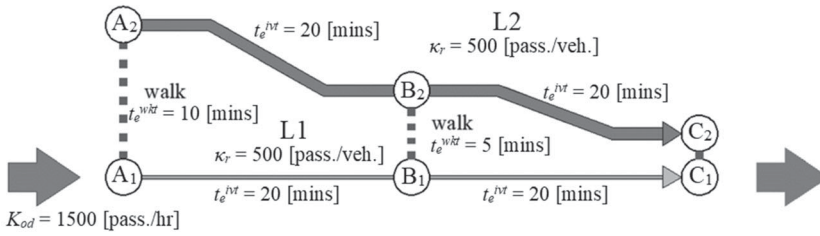
The following points summarise main observations from experiments on the second toy-network:

(f) *Origin passenger shifts evolving with RTCI:*

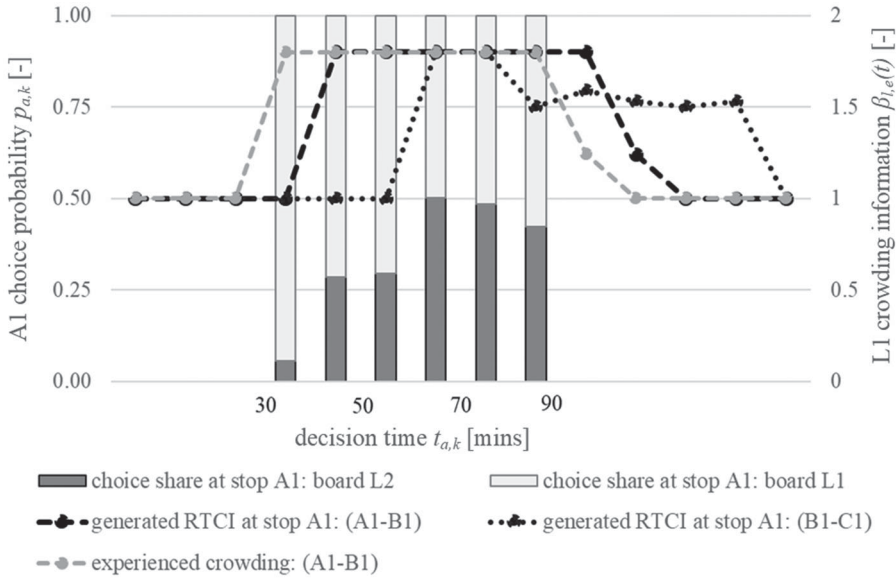
Figure 8 shows origin choice patterns as passengers receive variable RTCI from the downstream network. Three characteristic time intervals can be distinguished here. In the first interval (before the 40<sup>th</sup> minute), the system has not recorded any passenger loads yet, and thus passengers do not anticipate any crowding – over 90% of them choose the L1 line, which leads to overcrowding. Then, in the second interval, overcrowding along the first (A1-A2) segment becomes visible to passengers entering stop A from the 40<sup>th</sup> minute onwards and clearly affects their choices: L2 patronage rate at the origin goes up from 10% to ca.

**Table 4.** Mean journey time and on-board comfort experience, i.e. perceived IVT measured for specific crowding levels. While average perceived journey time (PJT) grows from 47 up to 72 min with increasing travel time  $\Delta$  in overcrowded conditions, it oscillates between 54 and 56 min with RTCI. Though initially RTCI leads to ca. 15% higher mean PJT (i.e. for  $\Delta$  equal to 10–30% of total L1 trip time), it results in up to 22% lower PJT when  $\Delta$  reaches 70–90% of the L1 trip time.

| Scenario                          |                            | travel time of the overcrowded B-C segment $\Delta$ [mins]<br>- share of total L1 trip time [%] |            |               |            |                |            |                |            |                |            |
|-----------------------------------|----------------------------|---|------------|---------------|------------|----------------|------------|----------------|------------|----------------|------------|
|                                   |                            | 3 [mins]: 10%   |            | 9 [mins]: 30% |            | 15 [mins]: 50% |            | 21 [mins]: 70% |            | 27 [mins]: 90% |            |
|                                   |                            | no  | yes        | no            | yes        | no             | Yes        | no             | yes        | no             | yes        |
| RTCI                              |                            |   |            |               |            |                |            |                |            |                |            |
| Absolute journey time             | mean [mins]:               | 27.1  | 29.1       | 33.4          | 34.0       | 39.6           | 38.8       | 45.9           | 42.8       | 52.2           | 47.3       |
|                                   | rel. change vs. 'no RTCI': | 107.4%  |            | 101.8%        |            | 98.0%          |            | 93.2%          |            | 90.6%          |            |
| Perceived total disutility        | mean [mins]:               | 47.1  | 54.3       | 53.4          | 55.5       | 59.7           | 56.8       | 66.0           | 55.4       | 72.2           | 56.3       |
|                                   | rel. change vs. 'no RTCI': | 115.3%  |            | 104.0%        |            | 95.1%          |            | 83.9%          |            | 77.9%          |            |
| Share of perceived IVT disutility | $\beta_{I,e} = 1.8$        | <b>14%</b>  | <b>14%</b> | <b>28%</b>    | <b>27%</b> | <b>36%</b>     | <b>33%</b> | <b>41%</b>     | <b>37%</b> | <b>44%</b>     | <b>37%</b> |
|                                   | $\beta_{I,e} = 1.5$        | 0%  | 0%         | 0%            | 0%         | 0%             | 0%         | 0%             | 0%         | 0%             | 1%         |
| In crowding conditions:           | $\beta_{I,e} = 1.2$        | 14%   | 14%        | 30%           | 29%        | 37%            | 37%        | 42%            | 39%        | 46%            | 42%        |
|                                   | $\beta_{I,e} = 1.0$        | 72%   | 72%        | 42%           | 43%        | 27%            | 30%        | 17%            | 24%        | 11%            | 19%        |



**Figure 7.** Topology of the second toy network. Service headways: L1 –10 min, L2 –30 min.

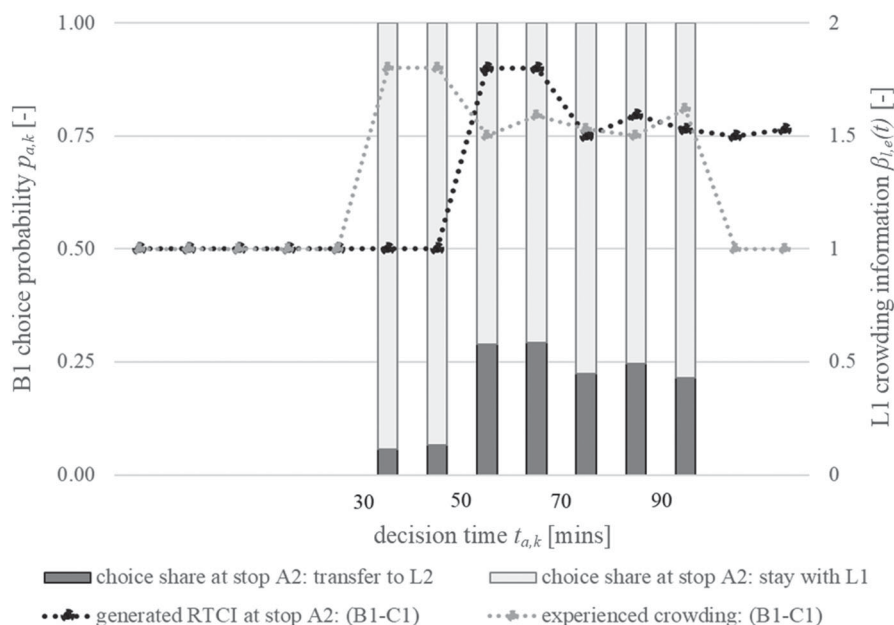


**Figure 8.** Origin choices evolving with RTCI. Dark bars denote the patronage rate of the longer but uncrowded L2 line. Initially low at barely 5% of O-D flows, it increases to ca. 25% as passengers receive information on overcrowding along the first L1 segment – and reaches a maximum of 60% once RTCI shows overcrowding on both L1 segments.

27%. In the third interval, once overcrowded L1 runs enter the second segment (A2-A3) of L1 line, i.e. from the 60<sup>th</sup> minute onwards, RTCI at the origin indicates now overcrowding on both L1 line segments, and the share of passengers choosing the L2 rises further up to almost 50%.

#### (g) En-route passenger shifts due to RTCI:

Importantly, RTCI affects not only origin choices but also induces shifts en-route, affecting alighting and connecting decisions as hereby illustrated. Bars in Figure 9 represent choices among passengers on-board the L1 runs as they approach the intermediate stop B1. This is the instance when alighting decision is being triggered: the expected utility of staying on-board L1 line is re-evaluated against alighting at stop B1 and transferring towards L2 line. As transfers require additional walking and waiting times, without RTCI the transfer probability is under 10%. However, once RTCI available to passengers reports overcrowding



**Figure 9.** En-route choices among passengers on-board the L1 line along the first segment (A1-A2). Dark bars denote the probability of transferring from L1 to L2 line, which involves a transfer between B1 and B2 stops and longer overall travel time but allows to avoid L1 overcrowding. Initially low at barely 5% of L1 passengers, it rises up to 20–25% as passengers receive information on overcrowding along second (A2–A3) segment of L1.

on L1 line along the downstream (A2-A3) segment, the disutility of staying on-board the L1 service becomes higher. Consequently, now ca. 25% of passengers decide to get off at stop A2 and take the L2 line to reach their destination.

(h) *Dissemination time lag of instantaneous RTCI:*

Importantly, simulation outputs also underscore a crucial feature associated with instantaneous RTCI system, i.e. dissemination time lag of crowding information. In Figures 8 and 9 we plot the anticipated crowding conditions  $\beta_{l,e}$  of line L1 segments (based on RTCI-provided information), against the crowding conditions  $\beta_{r,e}$  actually experienced by passengers. In certain cases, this leads to discrepancies as visible e.g. for origin passengers boarding the first bus trip at the 30th minute: since no passenger flows were reported yet in the PT network, they expect an uncrowded trip ( $\beta_{l,e} = 1.0$ ), but as an aggregate result of multiple such decisions, each of them eventually ends up on-board a fully overcrowded bus ( $\beta_{r,e} = 1.8$ ) departing from stop A1. Similarly, passengers approaching the stop A2 do not expect downstream crowding according to current RTCI, but since they fully comply with this information, staying on-board results in overcrowded travel conditions along the next (A2-A3) segment. Only after the 60th minute – i.e. once overcrowding has been reported along both L1 line segments – will the next approaching passengers receive the updated RTCI information and thus have the incentive to transfer towards line L2. This inevitable information lag is an important limitation of instantaneous RTCI system.

In specific settings, decisions made based on anticipated crowding will not be consistent with actually experienced crowding, thereby exposing a major deficiency in the RTCI credibility. This is especially relevant in contexts where demand levels are characterised by abrupt changes.

(i) *Information accuracy:*

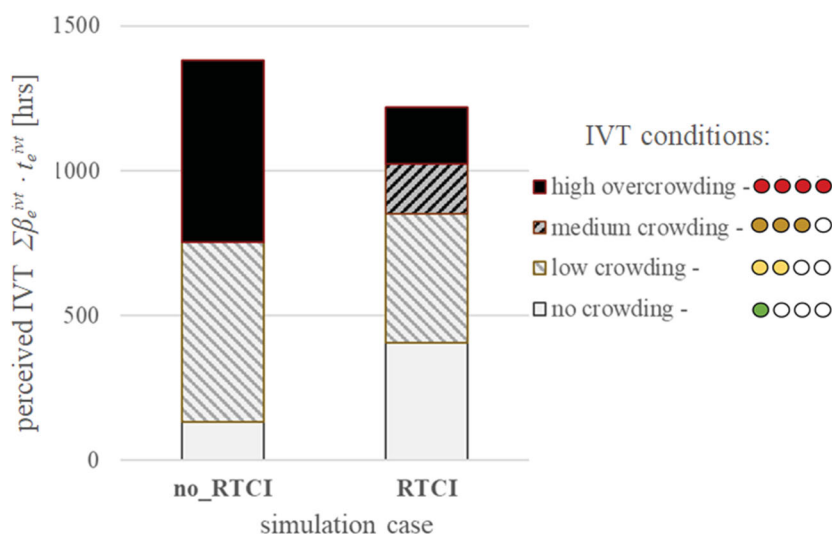
Finally, investigation of RTCI accuracy as defined in Equation (10) reveals that majority of travellers in the second toy-network receive crowding information consistent with their subsequent on-board experience. Globally, crowding information is found to be accurate ( $c^0$ ) for 80% of cases. Around 12% of passengers' decisions are based on RTCI which underestimates the actual crowding ( $c^-$ ) – which can be deemed especially negative for both the PT passengers and operators. Majority of these underestimation cases are observed for passengers that board the L1 services along the (A1–B1) and (B1–C1) segments early-on in the simulation. No crowding was reported yet by the RTCI system, but because of universal obedience of the instantaneously generated RTCI – the first L1 bus trips along these segments are experienced as unexpectedly overcrowded. Conversely, in 8% of the cases, passengers expected higher crowding conditions than actually experienced ( $c^+$ ) – i.e. overestimation which has non-adverse consequences for travel utility, but nevertheless negative for RTCI system credibility. Conversely, crowding overestimation is typically observable in the latter simulation period, once demand flows along the L1 begin to decrease but upstream passengers are still being notified of L1 overcrowding. Both overestimation and underestimation instances are therefore a direct consequence of the above-mentioned dissemination time lag of instantaneous RTCI. Clearly, this also underscores the shortcomings of non-anticipatory RTCI that is utilised by travellers in a non-cooperative manner. While in majority of instances the access to RTCI seems to be beneficial, the noticeable inaccuracy risk implies that in certain cases it can become actually counterproductive.

(j) *Output travel experience benefits with RTCI:*

Similarly to results reported for the first toy-network, introducing RTCI improves average travel utility by ca. 8%. A major share of these benefits are evidently reflected in the perceived IVT (Figure 10) – especially, a substantially lower share of passengers travelling in overcrowded conditions: total perceived IVT equivalent to  $\beta_{I,e} = 1.8$  reduces by 60%. Meanwhile, the relative share of two most 'comfortable' perceived IVT categories ( $\beta_{I,e} \leq 1.2$ ) – reflecting conditions when every passenger can get a seated place – goes up from 55% to 63% with the RTCI access, and ca. 15% of perceived IVT is now associated with medium crowding conditions ( $\beta_{I,e} = 1.5$ ), which are non-observable at all (0% share) in the 'no RTCI' case. These findings show that crowding information access, though it does not eliminate overcrowding in the PT network, allows to smoothen the on-board comfort travel experience. The risk of travelling in extreme crowding conditions is significantly mitigated and passenger loads are distributed more evenly across the L1 and L2 departures.

### **3.3. Real-size PT network simulation of instantaneous RTCI – case study of Kraków**

After analysing the main capabilities of our proposed model on toy-networks, we demonstrate the effects of instantaneous RTCI algorithm for a real-world PT network. We simulate

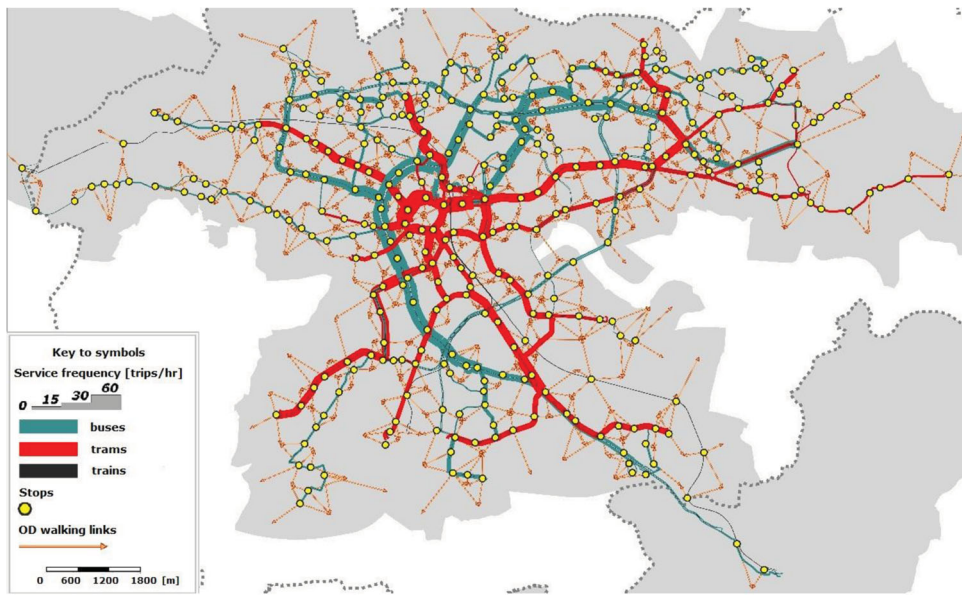


**Figure 10.** Perceived IVT in the second toy-network, decomposed between four crowding levels. RTIC not only decreases the overall IVT disutility, but also allows to utilise vehicle capacity more efficiently – note the substantial reduction in overcrowding experience (equivalent to  $\beta_{I,e} = 1.8$ ), coupled by rising share of travel time spent in more comfortable conditions ( $\beta_{I,e} < 1.8$ ).

a simplified model of the core urban PT system of Kraków city, Poland (i.e. the second-largest city in Poland with ca. 750k inhabitants): the supply model consisting of over 800 stops served by 136 bus, tram and urban rail (SKA) lines (Figure 11), and the demand model representing the typical PM peak-hour pattern (between 3 and 4 pm), with about 65k passenger trips distributed between ca. 10k OD pairs. We apply the same methodology as in toy-network studies and simulation output is again compared between ‘RTIC’ scenario (assuming 100% compliance with crowding information) vs. ‘no RTIC’ scenario. Output welfare estimations assume monetary time valuations as specified by JASPERS (2015) for the 2019 data, which weighted across three essential trip categories (business, commuting and leisure trips), give an average rate of ca. 33.1 [PLN/hr], or 7.5 [EUR/hr] (assuming a conversion rate of 4.4 [PLN/EUR]). PT demand and supply data is assumed based on a recent comprehensive travel survey and the Kraków strategic transport model (Szarata 2015), which served as a reference point for calibrating the Kraków model applied for our simulation works. The model was validated with respect to main characteristics on the demand (OD passenger volumes, travel times) and supply sides (PT network, service provision, segment run times) to ensure its credible representation of the city PT system.

Our simulations on a real-world PT network model show that passengers’ obedience of instantaneous RTIC can contribute to a more efficient distribution of passenger loads in the urban PT network. In the RTIC scenario, a visible increase in number of passengers able to get a seated place – coupled with substantial reduction in passenger volumes exposed to the worst overcrowding conditions – is traceable along multiple PT network segments in the PM peak hour (Figure 12). Interestingly, this becomes evident along inner-city, busy PT routes, which tend to be highly crowded during the PM peak. Such changes are an outcome of a number of phenomena already revealed in toy network simulations, emerging



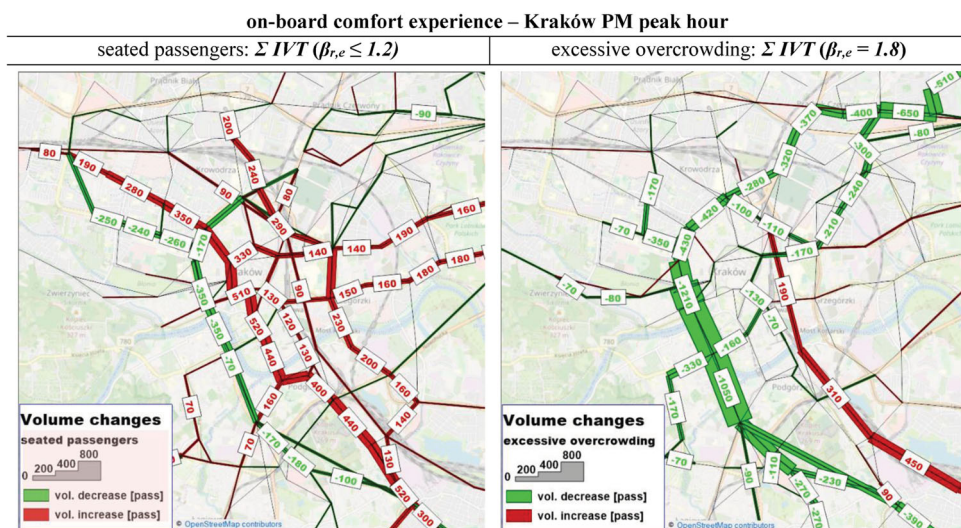


**Figure 11.** Case-study network – topology of the core urban PT system in Kraków (Poland).

in passengers' path choices due to acquisition and utilisation of RTCI. Among others, RTCI encourages passengers to use less-crowded services, operating in close proximity or as adequate alternatives to the most popular PT routes. It also increases the visibility of parallel service corridors, which are of low popularity due to their lower service frequency (requiring extra waiting time) or extra detour from the O-D travel route (requiring extra walking distance and/or transfers) – hence e.g. the rising patronage rate of the suburban (SKA) train line. Consequently, provision of RTCI in the PM peak hour allows passengers to utilise the available system capacity which is under-utilised otherwise if they are unable to foresee downstream crowding conditions.

The resultant passenger flow shifts translate into better journey experience, reducing the overall travel disutility (PJT) by ca. 2.5% with RTCI availability (Table 5). This is principally attributable to lower waiting time (WT) and in-vehicle time (IVT) disutility, while changes in walking time (WKT) and transfer (TRP) disutility are relatively smaller (yet still favourable). An important share of PJT savings stems from significantly lower incidence of waiting time due to denied boarding – a major source of PT overcrowding disutility – whose decline thanks to RTCI provision reaches about 30% (Table 5).

Moreover, a detailed inspection of perceived IVT (Figure 13) further confirms that access to RTCI brings network-wide improvement of on-board travel comfort experienced by passengers. The share of weighted travel time spent in excessive overcrowding conditions ( $\beta_{I,e} = 1.8$ ) decreases by the highest margin among all the 4 distinguished crowding conditions (on our RTCI scale defined in Table 3) – in aggregate terms, ca. 30% less than without access to RTCI; share of travel time spent in moderate standing conditions ( $\beta_{I,e} = 1.5$ ) is down by 9%; and ultimately, ca. 4% more weighted travel time is associated with travelling in seated conditions ( $\beta_{I,e} = 1.0$  or  $\beta_{I,e} = 1.2$ ). Once PJT changes are translated into projected passenger welfare benefits, total travel disutility savings, enabled by RTCI availability



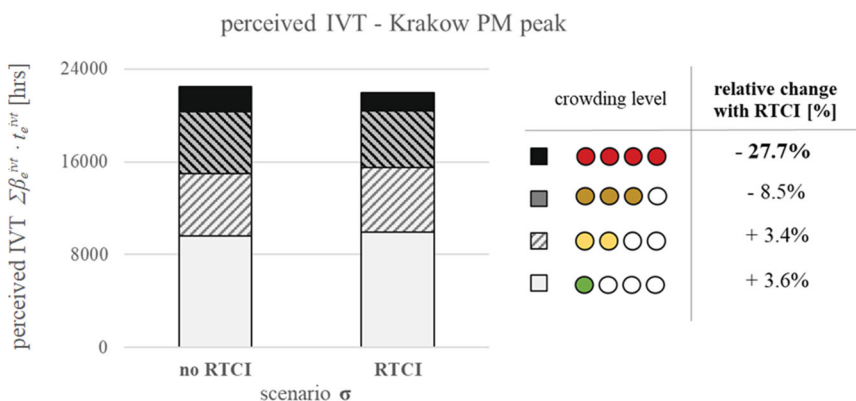
**Figure 12.** Kraków PM peak results – impact of RTCI on recorded passenger flows in urban PT network. Shifts in network flows indicate that access to RTCI allows more passengers to get a seated place (*left*) and simultaneously avoid excessive overcrowding (*right*) during their journey.

**Table 5.** Kraków PM peak results – output travel experience, measured by perceived journey time (PJT) changes. Access to RTCI improves all the PJT components, though by a limited amount, and total PJT decreases by ca. 2.3%. Notably, substantial improvements are visible in terms of denial-of-boarding disutility, which is reduced by 30%.

| Passenger travel disutility – PM peak |  | Scenario      |               | Relative change $\Delta$ [%] |
|---------------------------------------|--|---------------|---------------|------------------------------|
|                                       |  | no RTCI       | RTCI          |                              |
| total weighted [hrs]                  | in-vehicle time ( <i>IVT</i> )                                     | 22 485        | 21 966        | –2.3%                        |
|                                       | waiting time ( <i>WT</i> )   | 19 199        | 18 238        | –5.0%                        |
|                                       | walk time ( <i>WKT</i> )   | 18 138        | 18 025        | –0.1%                        |
|                                       | transfer penalties ( <i>TRP</i> )                                  | 8 409         | 8 311         | –1.2%                        |
|                                       | <b>total (<i>PJT</i>)</b>  | <b>68 132</b> | <b>66 540</b> | <b>–2.3%</b>                 |
|                                       | <b>waiting time – denied boarding (<i>WT<sub>denied</sub></i>)</b> | <b>2 532</b>  | <b>1 762</b>  | <b>–30.4%</b>                |
| average [mins]                        | absolute journey time ( <i>JRT</i> )                               | 35.3          | 35.1          | –0.8%                        |
|                                       | perceived journey time ( <i>PJT</i> )                              | 63.5          | 62.0          | –2.3%                        |

in the Kraków urban PT network, are equivalent to approx. 12 000 EUR (52 800 PLN) per PM peak hour on a typical weekday. This implies potential savings of approx. 0.2 EUR (0.8 PLN) per passenger trip, or an equivalent of almost 25% of a single-ticket fare. In annual terms, such RTCI-induced welfare benefits – during a single PM peak hour only – would amount to over 3.6 m EUR (15.8 m PLN).

Finally, accuracy reported in our Kraków case-study indicates that RTCI is fully consistent with passengers' subsequent observations of on-board crowding only in just about  $c^0 = 56\%$  of decision instances (Table 6). Importantly, there was a  $c^- = 31\%$  probability of observing worse on-board conditions than expected (i.e. RTCI underestimation risk), whereas the opposite case of passengers experiencing a more comfortable trip than



**Figure 13.** Kraków PM peak results – output changes in on-board comfort experience, measured by perceived in-vehicle travel time (IVT). Though RTCI reduces overall IVT disutility decreases by just ca. 2.3%, it contributes to substantially lower experience of the worst overcrowding experience ( $\beta_{r,e} = 1.8$ ), which is down by almost 30%. Meanwhile, probability of comfortable (seated) travel conditions ( $\beta_{r,e} \leq 1.2$ ) is also higher in the RTCI scenario.

**Table 6.** Kraków PM peak results – RTCI accuracy is satisfied in only ca. 56% of passengers’ decision instances. Importantly, in almost 31% of all cases, passengers’ output decisions were made based on RTCI which underestimated the actually observed crowding conditions. In about 20% of decision instances, some seats were still supposed to be available inside the PT vehicle, but ultimately all seats became occupied once passengers got on-board.

| Share of pass. decision instances $c(a_{k,s})$ |            | Predicted RTCI $\beta_{r,e}(t_{a,k})$ |             |              |             |
|--|------------|---------------------------------------|-------------|--------------|-------------|
|  |            | 1.0                                   | 1.2         | 1.5          | 1.8         |
| Observed RTCI $\beta_{r,e}(t_{r,e})$           | 1.0        | <b>36.4%</b>                          | 2.8%        | 4.1%         | 0.3%        |
|  | <b>1.2</b> | 6.1%                                  | <b>1.3%</b> | 2.4%         | 0.2%        |
|  | <b>1.5</b> | 13.5%                                 | 3.6%        | <b>14.9%</b> | 3.2%        |
|  | <b>1.8</b> | 2.6%                                  | 0.7%        | 4.4%         | <b>3.5%</b> |

expected was about three times less likely ( $c^+ = 13\%$ ). This reflects a significant risk for instantaneous RTCI reliability is related to its inability to anticipate fluctuating demand conditions. Overall, our findings illustrate that despite shortcoming in information accuracy, RTCI can reduce passengers’ travel times. Notwithstanding, there is a risk that users will not perceive this information trustworthy in the long run, jeopardising its potential benefits.

#### 4. Discussion

This paper introduces a novel modelling approach for describing the effects of instantaneous real-time crowding information (RTCI) systems in public transport (PT) networks. The principal contribution of this work lies within the proposed methodology – firstly, an extended passenger path choice model accounting for influence of RTCI upon travel decisions, and secondly – assignment framework simulating processes both in PT supply and PT demand system, as a consequence of instantaneous generation, dissemination and utilisation of RTCI. The RTCI algorithm is implemented within a dynamic, agent-based

PT assignment model, and can be therefore used to investigate the wider spectrum of consequences for PT passengers and operators.

The secondary contribution of this study lies in the application of the proposed RTCI algorithm, which reveals implications of passenger flow shifts potentially emerging once passengers follow the information on current crowding conditions of their PT services. Decision shifts are likely to occur at any journey stage (either pre-trip or en-route), producing route choice patterns that are more dynamic and evolve instantaneously. Eventually, the magnitude of these changes alters the performance of PT systems, and thus a feedback loop arises between RTCI-based passengers' travel choices, output passenger flows, PT network conditions, and subsequently generated crowding information. Our results, firstly on toy-network schemes and then for the PT network of Kraków, demonstrate ramifications of such phenomena – focusing on the spatiotemporal evolution of network crowding and passengers' decisions, network capacity utilisation, accuracy of provided information and passengers' resultant travel utility (welfare).

In particular, case-study simulations on the Kraków PT model shed more light onto potential consequences in urban PT networks. Apparently, ubiquitous (100%) response to instantaneous RTCI leads to improvements in passengers' current journey experience – our results obtained for a typical PM peak hour in Kraków indicate visible yet limited benefits in perceived travel disutility of 2 - 3% in global terms. Perhaps more importantly, RTCI mitigates the worst overcrowding experience: share of the worst on-board overcrowding experience (i.e. total passenger-hours corresponding to the highest crowding level of  $\beta_{r,e} = 1.8$ ) decreases by 27%, share of moderate crowding experience (i.e.  $\beta_{r,e} = 1.5$ ) decreases by 9% and waiting time due to denied boarding reduces by 30%. When translated into projected welfare gains, these figures imply monetary savings equivalent to approx. 3.6 million EUR when calculated annually for the PM peak-hour only.

However, obtained results also underscore risks related to RTCI accuracy, stemming from the inherent property of instantaneous RTCI: namely, the dissemination time lag of crowding information, and the impact of information provision and resultant passengers' responses on the validity of provisioned RTCI, which leads to discrepancy between anticipated vs. experienced on-board crowding. This major shortcoming of instantaneous RTCI system reliability concerns its inability to forecast sufficiently the abrupt fluctuations in network saturation conditions. Consequently, crowding information turned out to be accurate only for ca. 56% of passenger decision instances during the Kraków PM peak hour, and for as many as 31% of travel decisions the RTCI system underestimated the actually observed crowding conditions. In particular, 20% of all travel decision instances involved the expectation of seats available that turned out to be incorrect once passengers got on-board the PT vehicle. More insight can help understand the extent to which this dissemination time lag can impact the overall PT system effectiveness, and when it might cause the RTCI system to become actually counterproductive. This also calls for the development of anticipatory RTCI techniques that can take inspiration from studies on travel information provision e.g. in the context of car traffic and help overcome the inaccuracy issues.

The proposed model allows to predict the insofar unreported phenomena of novel RTCI solutions and illustrate their future performance in PT networks. As such, it can serve as a useful decision support system tool for researchers, transport engineers and authorities, being applicable for both strategic planning, tactical and real-time operations. Apart

from simulating the RTCI-induced phenomena and reproducing shifts in PT system dynamics, it can be directly utilised to measure the effectiveness of instantaneous RTCI under various conditions, as well as its implications for PT system performance and passengers' travel experience. Our case-study findings already indicate its certain advantages and shortcomings: positive changes in journey experience that might be however coupled with substantial inaccuracy risks. This might raise concerns as to whether passengers would ultimately trust and follow such instantaneously generated crowding information. Importantly, in simulations we assume a universal (100%) response rate to the RTCI, since our objective is to illustrate the generic effects of RTCI provision and contrast them with the 'no RTCI' network. In reality, not all the passengers will ultimately obey the RTCI, and so the influence of variable RTCI penetration rates needs to be further explored.

While we hope that our study provides an enriching contribution towards the novel RTCI solutions, it can be also seen as a groundwork paving the way for follow-up works necessary in this interesting and extensive research field. The methodology presented in this study focuses specifically on instantaneous RTCI scheme that is generated from a single (most recent) PT departure but it has been formulated in a universal way and can be directly extended to implement various feasible RTCI generation strategies, e.g. weighted crowding information of real-time services and/or historical data, as well as inclusion of crowding prediction algorithms. While demonstrating the general capabilities of our RTCI algorithm, application findings presented in this paper are formulated based on specific case-study conditions. More research is needed to understand the effects of RTCI systems in relation to different scenario assumptions and derive universal (transferable) conclusions, considering e.g. network coverage, and variable (over)crowding conditions. Results are specific to our RTCI penalty function (Table 3), which should be validated in empirical investigation. For example, perceptions of crowding penalty need not be only a function of absolute volume-capacity ratios, but can also include probability of getting a seat at a certain journey stage (e.g. Sumalee, Tan, and Lam 2009). This would then imply a relative decrease in crowding penalty for longer trips.

Crucially, our framework assumes that passengers do not have any prior expectations of crowding information and can only follow the currently available RTCI. This can be deemed representative for RTCI impacts in case of travellers unfamiliar with the PT system or for unexpected network events, but may also overestimate the potential RTCI benefits. An extended RTCI algorithm should reflect the passengers' long-term accumulation of travel experience, acquisition of crowding experience vs. crowding information and day-to-day adaptation of travel strategies. This can be further reformulated as an equilibrium seeking process between disseminated crowding information vs. resultant travel choices, where system operator may learn its dynamics and seek to provide predictive RTCI, whereas PT users may use their experience to anticipate actual realisation (i.e. travel conditions) from the RTCI. Solution of this iterative problem will help understand how RTCI provision can contribute to shifting the PT network from non-cooperative user equilibrium towards double-anticipatory user equilibrium. The latter is expected to result with conditions which are more similar to those obtained under system optimum conditions. Thus, another interesting and worthwhile research direction concerns the application of RTCI simulation models to plan such (pro)active control and demand management strategies, which might improve network performance (and simultaneously e.g. reduce the risk of information inaccuracy or service unreliability).



Moreover, other demand-side determinants can also play a relevant role in overall assessment of the RTCI effects, including the heterogeneity in travel behaviour (with regards to both reading of RTCI and choice response) or demand elasticity impacts of RTCI. Additionally, in this study we focus on the spatial dimension of RTCI impact upon travel decisions, i.e. how it affects the choice probability of a specific PT route. These considerations can be extended to temporal effects, i.e. how RTCI may also influence the choice probability of individual PT departures, either instantaneously and/or in the long-term. Finally, validation of actual (revealed) passenger behaviour in the future will be an essential step towards developing proposed model into a reliable, evidence-based analytical tool for the RTCI systems in PT networks.

## Acknowledgement

This study has been supported by research grant received within the Iwanowska Programme, organised by the Polish National Agency for Academic Exchange (NAWA) (agreement no. PPN/IWA/2018/1/00084). The authors would also like to express their gratitude to anonymous reviewers, whose insightful comments and remarks contributed to the final version of this paper.

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## Funding

This study has been supported by research grant received within the Iwanowska Programme, organised by the Polish National Agency for Academic Exchange (NAWA) (agreement no. PPN/IWA/2018/1/00084).

## ORCID

Arkadiusz Drabicki  <http://orcid.org/0000-0001-8191-6886>

Rafał Kucharski  <http://orcid.org/0000-0002-9767-8883>

Oded Cats  <http://orcid.org/0000-0002-4506-0459>

Andrzej Szarata  <http://orcid.org/0000-0002-0057-3023>

## References

- Abenzo, R. F., O. Cats, and Y. O. Susilo. 2017. "Travel Satisfaction with Public Transport: Determinants, User Classes, Regional Disparities and Their Evolution." *Transportation Research Part A: Policy and Practice* 95: 64–84.
- Bansal, P., R. Hurtubia, A. Tirachini, and R. A. Daziano. 2019. "Flexible Estimates of Heterogeneity in Crowding Valuation in the New York City Subway." *Journal of Choice Modelling* 31: 124–140.
- Batarce, M., J. C. Muñoz, and J. de Dios Ortúzar. 2016. "Valuing Crowding in Public Transport: Implications for Cost-Benefit Analysis." *Transportation Research Part A: Policy and Practice* 91: 358–378.
- Batarce, M., J. C. Muñoz, J. de Dios Ortúzar, S. Raveau, C. Mojica, and R. A. Ríos. 2015. "Use of Mixed Stated and Revealed Preference Data for Crowding Valuation on Public Transport in Santiago, Chile." *Transportation Research Record* 2535 (1): 73–78.
- Börjesson, M., and I. Rubensson. 2019. "Satisfaction with Crowding and Other Attributes in Public Transport." *Transport Policy* 79: 213–222.
- Cats, O. 2011. *Dynamic Modelling of Transit Operations and Passenger Decisions*. Doctoral dissertation. Stockholm: KTH Royal Institute of Technology.

- Cats, O., and M. Hartl. 2016. "Modelling Public Transport on-Board Congestion: Comparing Schedule-based and Agent-Based Assignment Approaches and Their Implications." *Journal of Advanced Transportation* 50 (6): 1209–1224.
- Cats, O., and E. Jenelius. 2014. "Dynamic Vulnerability Analysis of Public Transport Networks: Mitigation Effects of Real-Time Information." *Networks and Spatial Economics* 14 (3–4): 435–463.
- Cats, O., and E. Jenelius. 2018. "Beyond a Complete Failure: the Impact of Partial Capacity Degradation on Public Transport Network Vulnerability." *Transportmetrica B: Transport Dynamics* 6 (2): 77–96.
- Cats, O., H. N. Koutsopoulos, W. Burghout, and T. Toledo. 2011. "Effect of Real-Time Transit Information on Dynamic Path Choice of Passengers." *Transportation Research Record* 2217 (1): 46–54.
- Cats, O., and J. West. 2020. "Learning and Adaptation in Dynamic Transit Assignment Models for Congested Networks." *Transportation Research Record* 2674: 1.
- Cats, O., J. West, and J. Eliasson. 2016. "A Dynamic Stochastic Model for Evaluating Congestion and Crowding Effects in Transit Systems." *Transportation Research Part B: Methodological* 89: 43–57.
- de Dios Ortuzar, J., and L. G. Willumsen. 2011. *Modelling Transport*. Chichester, UK: John Wiley & Sons.
- Drabicki, A., R. Kucharski, O. Cats, and A. Fonzone. 2017a. "Simulating the Effects of Real-Time Crowding Information in Public Transport Networks." In 2017 5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS) (pp. 675–680). IEEE.
- Drabicki, A., R. Kucharski, and A. Szarata. 2017b. "Modelling the Public Transport Capacity Constraints' Impact on Passenger Path Choices in Transit Assignment Models." *Archives of Transport* 43: 7–28.
- East Japan Railway Company. 2019. "JR-EAST Train Info. [Mobile Application Software]." Accessed December 18, 2019. <http://www.jreast-app.jp/en/>.
- Elhenawy, M., H. Chen, and H. A. Rakha. 2014. "Dynamic Travel Time Prediction Using Data Clustering and Genetic Programming." *Transportation Research Part C: Emerging Technologies* 42: 82–98.
- Engelson, L., and M. Fosgerau. 2016. "The Cost of Travel Time Variability: Three Measures with Properties." *Transportation Research Part B: Methodological* 91: 555–564.
- Fonzone, A. 2015. "What do you do with Your App? Study of Bus Rider Decision Making with Real-time Passenger Information." *Transportation Research Record* 2535 (1): 15–24.
- Fonzone, A., and J. D. Schmöcker. 2014. "Effects of Transit Real-time Information Usage Strategies." *Transportation Research Record* 2417 (1): 121–129.
- Fonzone, A., J. D. Schmöcker, and F. Viti. 2016. "New Services, New Travelers, Old Models? Directions to Pioneer Public Transport Models in the Era of Big Data." *Journal of Intelligent Transportation Systems* 20 (4): 311–315.
- Gavrilidou, A., and O. Cats. 2019. "Reconciling Transfer Synchronization and Service Regularity: Real-time Control Strategies Using Passenger Data." *Transportmetrica A: Transport Science* 15 (2): 215–243.
- Gentile, G., and K. Noekel. 2016. *Modelling Public Transport Passenger Flows in the Era of Intelligent Transport Systems*. New York: Springer International.
- Google, L. L. C. 2019. "Google Maps – Transit & Food [Mobile Application Software]." Accessed December 18, 2019. <https://apps.apple.com/us/app/google-maps-transit-food/id585027354>.
- Hamdouch, Y., H. W. Ho, A. Sumalee, and G. Wang. 2011. "Schedule-based Transit Assignment Model with Vehicle Capacity and Seat Availability." *Transportation Research Part B: Methodological* 45 (10): 1805–1830.
- Haywood, L., and M. Koning. 2015. "The Distribution of Crowding Costs in Public Transport: New Evidence from Paris." *Transportation Research Part A: Policy and Practice* 77: 182–201.
- Hörcher, D., D. J. Graham, and R. J. Anderson. 2017. "Crowding Cost Estimation with Large Scale Smart Card and Vehicle Location Data." *Transportation Research Part B: Methodological* 95: 105–125.
- JASPERS. 2015. *Blue Book – Road Infrastructure*. Luxembourg: Joint Assistance to Support Projects in European Regions (JASPERS).
- Jenelius, E. 2018. "Car-Specific Metro Train Crowding Prediction Based on Real-Time Load Data." In 2018 21st IEEE International Conference on Intelligent Transportation Systems (ITSC) (pp. 78–83).
- Jenelius, E. 2019. "Data-Driven Bus Crowding Prediction Based on Real-Time Passenger Counts and Vehicle Locations." In 2019 6th IEEE International Conference on Models and Technologies for Intelligent Transport Systems (MT-ITS 2019).

- Jenelius, E. 2020. "Data-Driven Metro Train Crowding Prediction Based on Real-Time Load Data." *IEEE Transactions on Intelligent Transportation Systems* 21 (6): 2254–2265.
- Kattan, L., and Y. Bai. 2018. "LRT Passengers' Responses to Advanced Passenger Information System (APIS) in Case of Information Inconsistency and Train Crowding." *Canadian Journal of Civil Engineering* 45 (7): 583–593.
- Kim, K. M., S. P. Hong, S. J. Ko, and D. Kim. 2015. "Does Crowding Affect the Path Choice of Metro Passengers?" *Transportation Research Part A: Policy and Practice* 77: 292–304.
- Kim, J. K., B. Lee, and S. Oh. 2009. "Passenger Choice Models for Analysis of Impacts of Real-Time bus Information on Crowdedness." *Transportation Research Record* 2112 (1): 119–126.
- Kroes, E., M. Kouwenhoven, L. Debrincat, and N. Pauget. 2014. "Value of Crowding on Public Transport in île-de-France, France." *Transportation Research Record* 2417 (1): 37–45.
- Kucharski, R., and G. Gentile. 2019. "Simulation of Rerouting Phenomena in Dynamic Traffic Assignment with the Information Comply Model." *Transportation Research Part B: Methodological* 126: 414–441.
- Laskaris, G., O. Cats, E. Jenelius, M. Rinaldi, and F. Viti. 2019. "Multiline Holding Based Control for Lines Merging to a Shared Transit Corridor." *Transportmetrica B: Transport Dynamics* 7 (1): 1062–1095.
- Leurent, F. 2009. "On seat Congestion, Passenger Comfort and Route Choice in Urban Transit: A Network Equilibrium Assignment Model with Application to Paris." In 88th Annual Meeting of the Transportation Research Board, Washington D.C., USA.
- Li, H., K. Gao, and H. Tu. 2017. "Variations in Mode-Specific Valuations of Travel Time Reliability and in-Vehicle Crowding: Implications for Demand Estimation." *Transportation Research Part A: Policy and Practice* 103: 250–263.
- Mayor of London. 2015. "London Infrastructure Plan 2050: Transport Supporting Paper. (Updated in 2015)." Accessed April 30, 2020. [https://www.london.gov.uk/sites/default/files/Transport%20Supporting%20Paper\\_3.pdf](https://www.london.gov.uk/sites/default/files/Transport%20Supporting%20Paper_3.pdf).
- Moovit Inc. 2019. "Moovit App [Mobile Application Software]." Accessed December 18, 2019. <https://moovitapp.com/>.
- Moreira-Matias, L., O. Cats, J. Gama, J. Mendes-Moreira, and J. F. de Sousa. 2016. "An Online Learning Approach to Eliminate Bus Bunching in Real-Time." *Applied Soft Computing* 47: 460–482.
- Nederlandse Spoorwegen. 2019. "NS Reisplanner Xtra. [Mobile Application Software]." Accessed December 18, 2019. <https://www.ns.nl/en/travel-information/ns-on-your-mobile/ns-app.html>.
- Noursalehi, P., H. N. Koutsopoulos, and J. Zhao. 2018. "Real Time Transit Demand Prediction Capturing Station Interactions and Impact of Special Events." *Transportation Research Part C: Emerging Technologies* 97: 277–300.
- Noursalehi, P., H. N. Koutsopoulos, and J. Zhao. 2019. "Predictive Decision Support for Real-Time Crowding Prediction and Information Generation." In 98th Annual Meeting of Transportation Research Board, Washington D.C., USA.
- Nuzzolo, A., and A. Comi. 2016. "Advanced Public Transport and Intelligent Transport Systems: New Modelling Challenges." *Transportmetrica A: Transport Science* 12 (8): 674–699.
- Nuzzolo, A., U. Crisalli, A. Comi, and L. Rosati. 2016. "A Mesoscopic Transit Assignment Model Including Real-Time Predictive Information on Crowding." *Journal of Intelligent Transportation Systems* 20 (4): 316–333.
- Peftitsi, S., E. Jenelius, and O. Cats. 2020a. "Evaluating Crowding in Individual Train Cars Using a Dynamic Transit Assignment Model." In 99th Annual Meeting of the Transportation Research Board, Washington D.C., USA.
- Peftitsi, S., E. Jenelius, and O. Cats. 2020b. "Determinants of Passengers' Metro Car Choice Revealed Through Automated Data Sources: A Stockholm Case Study." *Transportmetrica A: Transport Science* 16 (3): 529–549.
- Preston, J., J. Pritchard, and B. Waterson. 2017. "Train Overcrowding: Investigation of the Provision of Better Information to Mitigate the Issues." *Transportation Research Record* 2649 (1): 1–8.
- Rudnicki, A. 1999. [in Polish] *Jakość Komunikacji Miejskiej* [Quality of Urban Public Transport]. Krakow: SITK RP.



- Schmitt, A. 2017. "These London Trains Have Real-Time Displays to Reduce Crowding." Streets-Blog USA, August 3. Accessed December 18, 2019. <https://usa.streetsblog.org/2017/08/03/these-london-trains-have-real-time-displays-to-reduce-crowding/>.
- Schmöcker, J. D., A. Fonzone, H. Shimamoto, F. Kurauchi, and M. G. Bell. 2011. "Frequency-based Transit Assignment Considering Seat Capacities." *Transportation Research Part B: Methodological* 45 (2): 392–408.
- Seoul Metropolitan Government. 2017. "New Service Allows Passengers to Check Congestion Level of Buses in Seoul, May 24." Accessed December 18, 2019. <http://english.seoul.go.kr/avoid-congested-buses-take-vacant-buses/?cat=29>.
- Singapore LTA (Land Transport Authority). 2019. "MyTransportSG [Mobile Application Software]." Accessed December 18, 2019. <https://www.mytransport.sg/content/mytransport/home.html>.
- Sumalee, A., Z. Tan, and W. H. Lam. 2009. "Dynamic Stochastic Transit Assignment with Explicit Seat Allocation Model." *Transportation Research Part B: Methodological* 43 (8–9): 895–912.
- Susan. 2018. Sydney Provides Real-Time Info on Train Crowding. *The Transit Wire*, May 17. Accessed December 18, 2019. <http://www.thetransitwire.com/2018/05/17/sydney-provides-real-time-info-on-train-crowding/>.
- Szarata, A. 2014. [in Polish] *Modelowanie podróży wzbudzonych oraz tłumionych zmianą stanu infrastruktury transportowej* [Modelling Induced and Suppressed Trips Resultant from Changes in Transport Infrastructure Conditions]. Krakow: Cracow University of Technology.
- Szarata, A. 2015. "[in Polish] Wyniki badań podróży w Krakowie – KBR 2013 [2013 Comprehensive Travel Survey Results in Krakow]." *Transport Miejski i Regionalny* 2015 (5): 4–8.
- Tirachini, A., D. A. Hensher, and J. M. Rose. 2013. "Crowding in Public Transport Systems: Effects on Users, Operation and Implications for the Estimation of Demand." *Transportation Research Part A: Policy and Practice* 53: 36–52.
- Tirachini, A., R. Hurtubia, T. Dekker, and R. A. Daziano. 2017. "Estimation of Crowding Discomfort in Public Transport: Results from Santiago de Chile." *Transportation Research Part A: Policy and Practice* 103: 311–326.
- Tirachini, A., L. Sun, A. Erath, and A. Chakirov. 2016. "Valuation of Sitting and Standing in Metro Trains Using Revealed Preferences." *Transport Policy* 47: 94–104.
- Toledo, T., O. Cats, W. Burghout, and H. N. Koutsopoulos. 2010. "Mesoscopic Simulation for Transit Operations." *Transportation Research Part C: Emerging Technologies* 18 (6): 896–908.
- van Essen, M., T. Thomas, E. van Berkum, and C. Chorus. 2016. "From User Equilibrium to System Optimum: A Literature Review on the Role of Travel Information, Bounded Rationality and Non-Selfish Behaviour at the Network and Individual Levels." *Transport Reviews* 36 (4): 527–548.
- van Oort, N., M. Drost, T. Brands, and M. Yap. 2015. "Data-Driven Public Transport Ridership Prediction Approach Including Comfort Aspects." In 13th CASPT Conference, Rotterdam, The Netherlands.
- Vlahogianni, E. I., M. G. Karlaftis, and J. C. Golias. 2014. "Short-term Traffic Forecasting: Where We Are and Where We're Going." *Transportation Research Part C: Emerging Technologies* 43: 3–19.
- Wardman, M., and G. Whelan. 2011. "Twenty Years of Rail Crowding Valuation Studies: Evidence and Lessons from British Experience." *Transport Reviews* 31 (3): 379–398.
- Whelan, G., and J. Crockett. 2009, April. "An Investigation of the Willingness to Pay to Reduce Rail Overcrowding." In Proceedings of the First International Conference on Choice Modelling, Harrogate, England (Vol. 30).
- Więcek, P., D. Kubek, J. H. Aleksandrowicz, and A. Strózek. 2019. "Framework for Onboard Bus Comfort Level Predictions Using the Markov Chain Concept." *Symmetry* 11 (6): 755.
- Woodard, D., G. Nogin, P. Koch, D. Racz, M. Goldszmidt, and E. Horvitz. 2017. "Predicting Travel Time Reliability Using Mobile Phone GPS Data." *Transportation Research Part C: Emerging Technologies* 75: 30–44.
- Yap, M., O. Cats, and B. van Arem. 2020. "Crowding Valuation in Urban Tram and Bus Transportation Based on Smart Card Data." *Transportmetrica A: Transport Science* 16 (1): 23–42.
- Zhang, Y., E. Jenelius, and K. Kottenhoff. 2017. "Impact of Real-time Crowding Information: A Stockholm Metro Pilot Study." *Public Transport* 9 (3): 483–499.