Assessing contributions of passenger groups to public transportation crowding

Skoufas, Anastasios; Cebecauer, Matej; Burghout, Wilco; Jenelius, Erik; Cats, Oded

DOI
10.1016/j.jpubtr.2024.100110

Publication date
2024

Document Version
Final published version

Published in
Journal of Public Transportation

Journal of Public Transportation 26 (2024) 100110

Contents lists available at ScienceDirect

Journal of Public Transportation

journal homepage: www.journals.elsevier.com/journal-of-public-transportation

Assessing contributions of passenger groups to public transportation crowding

Anastasios Skoufas [a,*], Matej Cebecauer [a], Wilco Burghout [a], Erik Jenelius [a], Oded Cats [a,b]

[a] Department of Civil and Architectural Engineering, KTH Royal Institute of Technology, Stockholm SE-10044, Sweden
[b] Department of Transport and Planning, Delft University of Technology, Delft 2600 AA, the Netherlands

ARTICLE INFO

Keywords:
Public transportation
Crowding
Passenger group
Smart card data

ABSTRACT

On-board crowding in public transportation has a significant impact on passengers' travel experience. However, there is little knowledge of how different passenger groups contribute to on-board crowding. Empirical knowledge of specific passenger groups' impact on the system facilitates more effective tuning of policy instruments such as new fare structures, dedicated public transportation services, infrastructure investments, and capacity provision. We propose a method to capture the crowding contributions from selected passenger groups by means of smart card data analytics. Two crowding contribution metrics at the passenger journey level are proposed: (1) *time-weighted contribution to load factor* and (2) *maximum contribution to load factor*. We apply the proposed method to the multimodal public transportation system of Region Stockholm, Sweden. We demonstrate the method for two groups: school students, and passengers traversing Stockholm's inner city. Our findings indicate that school students and passengers traversing the inner city have similar crowding contributions, utilizing 15 % and 11 % of the seating capacity across all modes during the AM and the PM peak, respectively. The commuter rail network, as well as some of the areas neighboring it, experience on average more than 70 % and 90 % utilization of their seating capacity during the AM peak, by school students and passengers traversing the inner city, respectively.

1. Introduction

On-board crowding is a key problem in large cities since it can negatively affect passengers' travel experience (Kim et al., 2015) as well as increase travel time variability and waiting times (Tirachini et al., 2013). This can impose additional pressure on the public transportation system, resulting in higher operational costs, especially during peak hours. In common practice, the demand component of crowding is investigated as a single quantity, thereby missing the perspective on how different groups contribute to crowding. Such information can facilitate the design of targeted policy measures towards crowding reduction and thereby reduce the pressure on the public transportation system.

The impacts of crowding on the demand side of the public transportation system have been subject to extensive research. Crowding is associated with passenger anxiety (Y.-H. Cheng, 2010), stress (Kim et al., 2015), and exhaustion (Mohd Mahudin et al., 2012), negative safety and security perceptions (Katz and Rahman, 2010), as well as feelings related to privacy invasion (Wardman and Whelan, 2011). For passengers working while commuting, crowding is related to productivity loss (Gripsrud and Hjorthol, 2012). The well-being effect of crowding can lead passengers to adjust their behavior in terms of mode, route choice, and departure time (Cheng et al., 2020). Several studies highlight the importance of crowding valuation since it can reveal how crowding affects passengers' choices (e.g. Hörcher et al., 2017; Yap et al., 2020; Yap et al., 2023).

From the standpoint of public transportation demand-supply interactions, crowding can affect dwell times and the total in-vehicle time due to difficult in-vehicle passenger movement (Tirachini, 2011). Specifically, for buses, increased dwell times can trigger bus bunching if no control measures (e.g., bus holding) are applied (Sáez et al., 2012). Passengers' waiting times are also related to crowding, given that passengers can be left behind due to capacity limitations (Sipetas et al., 2020). Furthermore, crowding externalities can be translated as an increase in the in-vehicle passenger cost, frequency, and fare optimization problems (Tirachini et al., 2010). Last, crowding can also be considered in public transportation line design, as it can further facilitate better

* Corresponding author.
E-mail addresses: skoufas@kth.se (A. Skoufas), matejc@kth.se (M. Cebecauer), wilco@kth.se (W. Burghout), jenelius@kth.se (E. Jenelius), oded.cats@kth.se, O.Cats@tudelft.nl (O. Cats).

https://doi.org/10.1016/j.jpubtr.2024.100110
Received 15 July 2024; Received in revised form 18 October 2024; Accepted 4 November 2024
Available online 15 November 2024

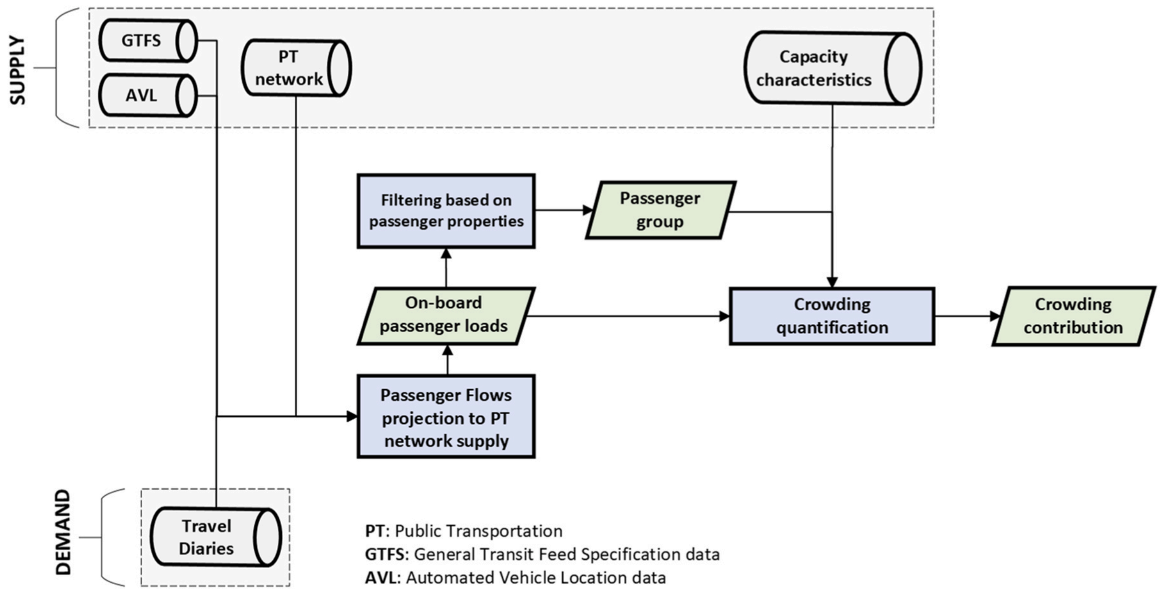

Fig. 1. Methodological framework (elliptic tubes: datasets, rectangles: modules, parallelogram: intermediate databases).

service frequency and capacity provision (Jara-Diaz and Gschwender, 2003).

Given the importance of crowding for public transportation planning and management, an array of crowding measures has been proposed to quantify its extent. On-board crowding in public transportation is defined as "having a significant number of people sharing a limited space while using a public transportation service" (Tirachini et al., 2013). It can be evaluated from two different perspectives; objective measures such as passenger density or capacity utilization, and subjective measures of passengers' experienced or perceived crowding (Evans and Wener, 2007).

An objective measure of on-board crowding is the *number of standing passengers per square meter,* for which the benchmarks defining unacceptable crowding levels vary across the world. For example, four standing passengers per square meter is the benchmark in Europe (The International Association of Public Transport (UITP), 2009) and in Australia (Diec et al., 2010). For the United States of America (USA), this number increases to five standees per square meter (National Academies of Sciences Engineering and Medicine, 2006) whereas in China, this threshold increases up to eight standees per square meter in buses (General Administration of Quality Supervision Inspection and Quarantine (AQSIQ), 2004) and up to six passengers per square meter in urban rail transit modes (Ministry of Housing and Urban-Rural Development of the People's Republic of China, 2013). Another objective measure is the load factor which is computed by dividing the passenger load by the vehicle's seating capacity. Based on this metric, Level of Service (LOS) standards can be defined. LOS between A and C refers to load factors of less than one, while LOS F indicates a '*crush load*' with load factors greater than 1.5 (Transit Cooperative Research Program, 2003). Different crowding levels (1–7) are defined based on the load factor percentage. Crowding levels 1, and 2 correspond to load factors less than 1, while crowding level 7 corresponds to load factor values greater than 2 (Wardman and Whelan, 2011).

The increasing availability of automated and passively collected data sources in the public transportation industry, such as smart card data, enables the understanding of the daily face-to-face encounters (Sun et al., 2013) and the network-wide estimation of on-board crowding. Automated data sources can facilitate route choice generation (e.g., in Skoufas et al., 2024), enabling the valuation of on-board crowding. Hörcher et al. (2017) propose the metrics of density of crowding (standing passengers/m^2) and standing probability for calculating crowding costs. In a similar way, Yap et al. (2020) introduce the metric of the time-weighted load factor and the standing density in order to quantify the valuation of public transportation crowding. Jenelius (2020) highlights the need to introduce predictive, personalized crowding measures incorporating the changing seat availability along a passenger journey. He proposes using the metrics of the probability of getting a seat upon boarding, emphasizing the necessity of having a seat for some passengers, and the expected travel time standing, incorporating the cumulative probability that a passenger will not secure a seat at each segment of the journey.

The recent COVID-19 pandemic brought the on-board crowding in public transportation into the spotlight, due to the social distancing recommendations and obligations being imposed worldwide. In this context, Basso et al. (2023) propose three novel objective crowding metrics to capture virus exposure in public transportation. Specifically, the average time that each bus service line has more than one passenger per square meter, the average number of passengers per square meter for all buses passing at each bus stop, and the total number of other people that each passenger meets during his/her trip are proposed. Lin et al. (2023) investigate the distribution and crowding exposure across different socioeconomic groups, introducing an equity perspective in the field of public transportation crowding.

In the literature, existing crowding metrics mainly refer to the total passenger load, and do not consider the heterogeneous mix of travelers that contribute to the on-board crowding. Different traveler groups are sensitive to different service attributes and contextual factors (e.g., fares, private car accessibility, work hours). In order to design effective policy measures, incentives, or supply adjustments to reduce crowding, it is therefore important to understand the composition of travelers on each line segment across the public transportation system and target the relevant groups.

In this study, we propose a methodology and a set of novel metrics to assess and understand the on-board crowding contributions that selected passenger groups inflict on the rest of the passenger journeys. We select the specific passenger groups to demonstrate the proposed method which is also applicable for other groups. For the selection, we considered groups for which tailored policy implications, such as new fare structures or dedicated public transportation services, could be potentially relevant. The method facilitates the interpretation of the results in space, further utilizing the capabilities of smart card data to provide a better understanding of on-board crowding. Importantly, the method is tailored for mobility traces such as smart card data, and is reproducible subject to data availability, thereby enhancing its added

Table 1
Notation.

Notation	Definition
g	Set of journeys made by passenger group
i	Passenger journey
d	Day of the week
t	Time of the day
z	Spatial unit (zone)
m	Mode of transport
a	Network segment (stop-to-stop link)
A_i	Set of network segments traversed by passenger i
κ_a	Vehicle seating capacity on segment a
l_{ag}	Passenger group load on segment a
t_a^{ivt}	Travel time on network segment a
qt_{ig}	Time-weighted contribution to load factor of group g to passenger journey i
$\overline{qt}_{g,d,t}$	Mean value of the time-weighted crowding contribution qt_{ig} on day d during time t
$\overline{qt}_{g,z,t}$	Mean value of the time-weighted crowding contribution qt_{ig} during time t for zone z
f_{ig}^{max}	Maximum contribution to load factor of group g to passenger journey i
$\overline{f^{max}}_{g,d,t,m}$	Mean value of the maximum contribution of group g to load factor on day d during time t in transport mode m

value. This study contributes to identifying and quantifying how, when, and where specific passenger groups contribute to on-board crowding. The results obtained can shed light on the demand component of crowding (e.g., who is contributing, and when?), the supply component (e.g., is the capacity provided adequate?) as well as the social dimension thereof (e.g., equity aspects related to crowding contributions).

We demonstrate our method by investigating the crowding contributions of two distinctive passenger groups, namely school students, and passengers traversing the inner city, given their distinct spatio-temporal travel patterns. Our application to the Region Stockholm, Sweden utilizes large-scale panel data in the form of smart card transactions. Potential applications of our approach include assessing the impact of new development areas on network congestion, analyzing secondary impacts of service changes due to the re-distribution of passenger flows, and the design of demand management and pricing instruments aimed at targeting specific user groups.

The remainder of this paper is organized as follows. Sections 2 and 3 present the proposed method and the case study area, respectively. In Section 4, the key findings of the study and their implications are presented. Section 5 discusses the implications of the results, reflects on future research directions and relevant research limitations, and presents the conclusions of the study.

2. Methodology

In this section, we present the methodology for assessing the crowding contributions of passenger groups. The most crucial part of the developed method is the proposed novel metrics capturing the crowding contributions induced by selected passenger groups (subsection 2.1). Existing crowding measures capture the total crowding in the system, while our proposed ones capture the crowding induced by a selected passenger group, therefore revealing a different perspective of crowding both in space and in time. The methodology is framed with the data description subsection, describing how the necessary attributes are obtained (subsection 2.2).

Fig. 1 presents the methodological framework for estimating the contribution to the on-board crowding from a selected passenger group. The notation used in the paper is summarized in Table 1.

2.1. On-board crowding contribution

The proposed metrics quantify the crowding contribution of a selected traveler group based on the load factor of the traveler group at the line segment (stop-to-stop link) level. The load factor of traveler group g on segment a is the ratio between the load of group g, l_{ag}, and the seating capacity of the vehicle, κ_a. The metrics are computed at the passenger journey level, and they capture i) the time-weighted average crowding contribution from the selected passenger group throughout the rest of the passenger journeys, and ii) the maximum segment crowding contribution in the rest of the passenger journeys. In calculating the proposed metrics, it is assumed that passengers prefer sitting over standing.

2.1.1. Time-weighted contribution to load factor

The time-weighted contribution to load factor (qt_{ig}) of passenger group g to passenger journey i not belonging to group g uses the travel time t_a^{ivt} for each segment traversed by journey i as a weight, thus highlighting the effect of travel time on the experienced on-board crowding. The components (l_{ag}, κ_a, t_a^{ivt}) of the time-weighted contribution to load factor (qt_{ig}) are known for all network segments traversed by passenger journey i (A_i). We calculate the metric qt_{ig} for each passenger journey i not belonging to the group g using formula (1):

$$qt_{ig} = \frac{\sum_{a \in A_i} \frac{l_{ag}}{\kappa_a} * t_a^{ivt}}{\sum_{a \in A_i} t_a^{ivt}}, \quad \forall i \notin g \tag{1}$$

Given that qt_{ig} is calculated for all passenger journeys i ($i \notin g$) the mean value for any given time of day t ($\overline{qt}_{g,d,t}$) is calculated using formula (2):

$$\overline{qt}_{g,d,t} = \frac{1}{|n_{g,d,t}|} \sum_{i \in n_{g,d,t}} qt_{ig} \tag{2}$$

where $n_{g,d,t}$ is the set of passenger journeys not in g on day d during time period t.

The proposed method also facilitates spatially aggregating the results (e.g., to statistical census zones). For every passenger journey i ($i \notin g$), the origin and destination at the stop level are known. We can aggregate the crowding contributions across all days, into geographical units according to formula (3):

$$\overline{qt}_{g,z,t} = \frac{1}{\sum_d |n_{g,d,t,z}|} \sum_d \sum_{i \in n_{g,d,t,z}} qt_{ig} \tag{3}$$

where $n_{g,d,z,t}$ is the set of passenger journeys i ($i \notin g$) on day d during time period t starting (or ending) in zone z.

2.1.2. Maximum contribution to load factor

The second proposed measure is the maximum contribution to the load factor across traversed network segments f_{ig}^{max}. The motivation for this metric lies in the fact that passengers tend to disproportionally recall negative travel experiences (Abenoza et al., 2017). Special emphasis is given when the travel conditions exceed a certain discomfort threshold (Börjesson and Rubensson, 2019). We are therefore interested in capturing: 1) the highest level of discomfort in a passenger journey i ($i \notin g$) due to the presence of the passenger group g and 2) network segment a (and transport mode m) where it was experienced throughout the journey. We calculate the f_{ig}^{max} for each passenger journey i using formula (4).

$$f_{ig}^{max} = \max_{A_i}\left(\frac{l_{ag}}{\kappa_a}\right), \quad \mathrm{a} \in \mathrm{Ai}, \quad \forall i \notin g \tag{4}$$

We can compute the mean value of f_{ig}^{max} across time of day and different transport modes m according to formula (5):

$$\overline{f^{max}}_{g,d,t,m} = \frac{1}{|n_{g,d,t,m}|} \sum_{i \in n_{g,d,t,m}} f_{ig}^{max} \tag{5}$$

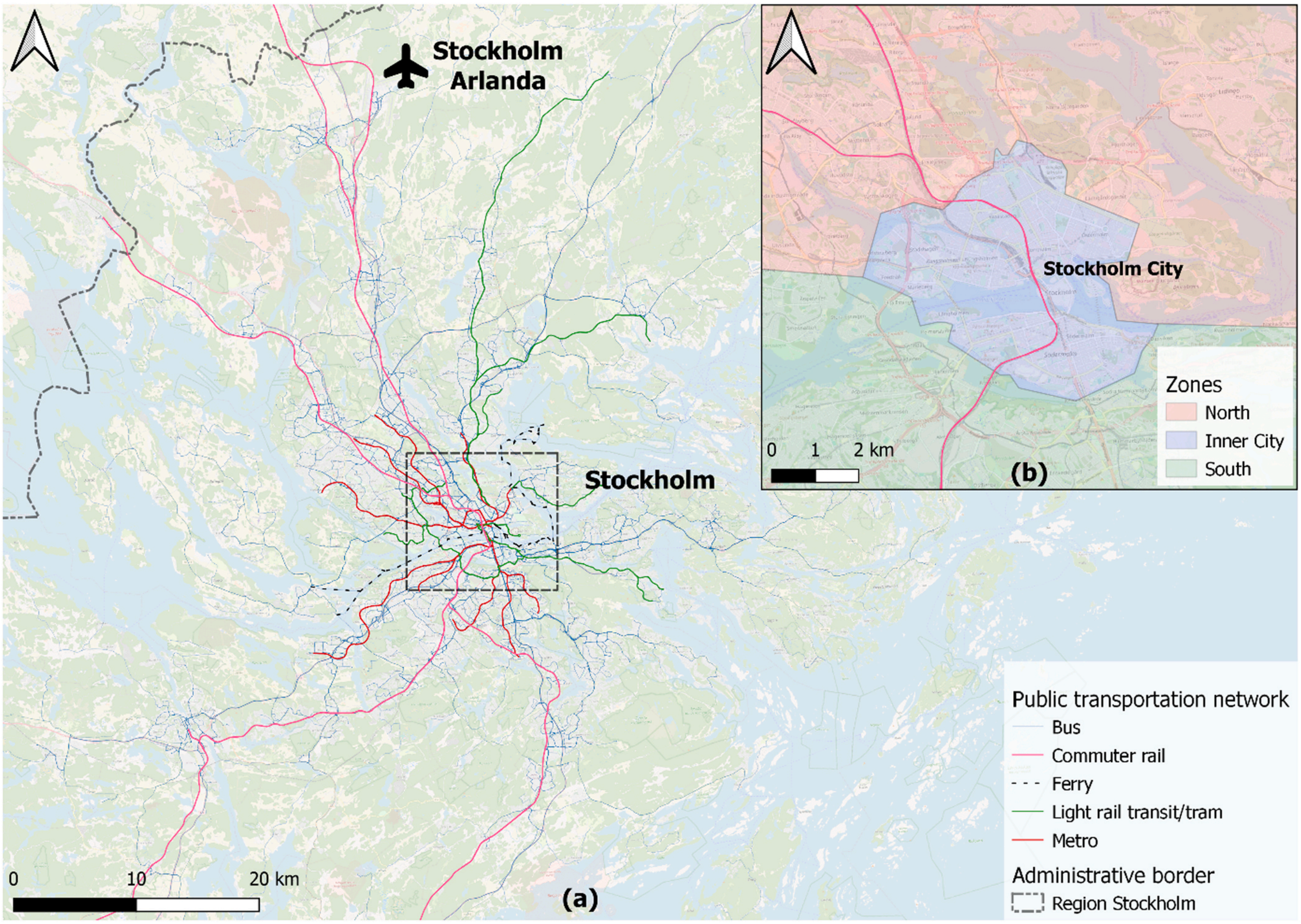

Fig. 2. The public transportation network of Region Stockholm (a) and the defined North, Inner city and South zones (b).

where $n_{g,d,t,m}$ is the set of passenger journeys i ($i \notin g$) on day d during time period t experiencing their maximum contribution to load factor from passenger group g on transport mode m.

This metric may not reflect the overall crowding condition in the mode. However, it can guide targeted policy-making (e.g., demand management strategies) to alleviate high-crowding contributions from a selected passenger group. Last, the metric's high level of detail, including the network segment a (and transport mode m) where the maximum crowding contribution was experienced, can guide decision-making at a macroscopic level (e.g., (re-)design of a new line). The abovementioned enhances the added value of the f_{ig}^{max} metric, especially since cost-efficient public transportation supply provision is a challenge in many cities worldwide.

2.2. Data description

In the proposed method, we utilize the full travel diaries of passengers. In the case of tap-in only public transportation systems, this implies that the tap-out locations (transfer stations and final destinations) need to be inferred. There are well-established techniques in the literature for inferring the tap-out locations of passenger trips (Kholodov et al., 2021; Munizaga and Palma, 2012; Trépanier et al., 2007).

2.2.1. Passenger flows projection to public transportation network supply

When tap-in and tap-out locations are known, the mode, line number, vehicle, departure and arrival times, and travel time for each passenger journey i can be inferred by fusing Automated Vehicle Location (AVL) and Automated Face Collection (AFC) data. When tap-in and tap-out on certain vehicle trip/departure are known, the travel time can be inferred by considering observed AVL departure and arrival times. For modes where passengers tap in at the gate or the platform, each trip is assigned to the first departure found in AVL data after the tap-in. Even if walking times from the gate to the platform may be neglected, this assumption is in line with the principle that passengers aim at minimizing their waiting times and, therefore, their total travel time. In the case of an urban rail network with a limited route choice set, this assumption does not critically affect passenger flow projection to the public transportation supply. GTFS data can replace missing or incomplete vehicle departures since they contain complete sets of scheduled departures for all lines across the network. In addition, each passenger journey i is connected to a unique smart card identifier, which enables the identification of the journeys belonging to the passenger group g, and the remaining ones, for which we estimate the proposed metrics.

2.2.2. Estimation of passenger loads

By projecting all passenger trips inferred in the public transportation system, we are able to estimate passenger loads for each network segment a traversed by each vehicle departure (run) in the system, including the passenger loads from the selected passenger group g (l_{ag}). In addition, travel times for each segment a (t_a^{ivt}) are known. Regarding the estimation of the seating capacity κ_a, each vehicle departure is connected with the vehicle's characteristics. Therefore, the seating capacity κ_a is known across the network segments traversed by each vehicle departure, and it can be used for estimating the proposed metrics for each passenger journey i which is associated with riding this particular vehicle ($i \notin g$).

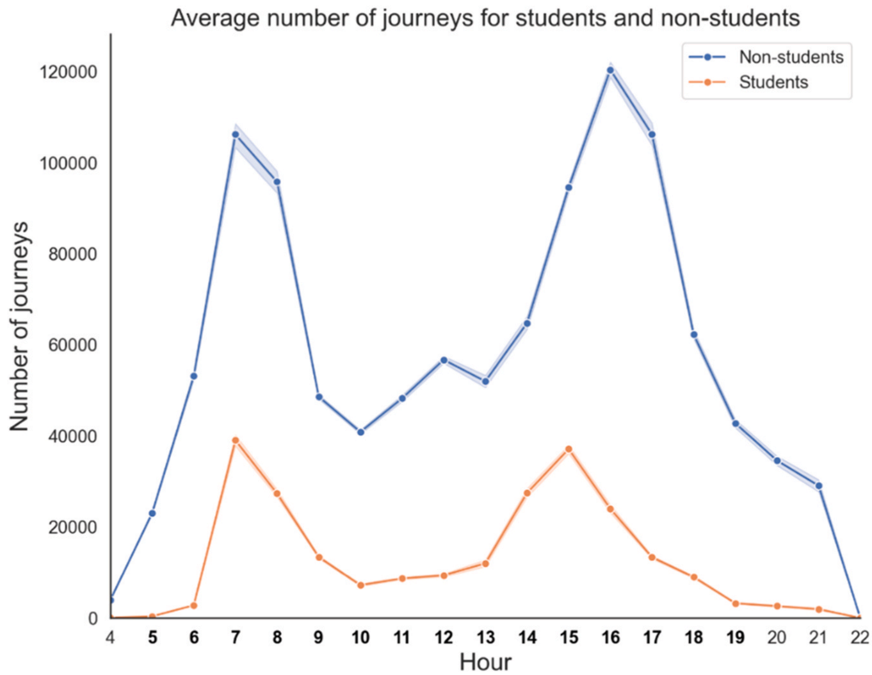

Fig. 3. Average hourly demand of students and non-students during workdays (highlighted hours on x-axis highlight the validity of the student ticket).

2.2.3. *Definition of passenger group g*

After assigning each passenger to a specific vehicle departure, the passenger loads across all the public transportation network can be computed. Smart card data contains information on all the transactions made with their respective time and location stamps as well as the associated subscription type. This enables the segmentation of smart card records based on subscription type, spatial or temporal criteria, or a combination thereof (Cats, 2023). For example, one can define the group of passengers of interest g as all those who depart within a certain time window, all those with a monthly pass, or all those who have been identified as members of a certain cluster based on long-term travel patterns such as ‘early birds’ (Cats and Ferranti, 2022a), or ‘local’ travelers (Cats and Ferranti, 2022b), or travelers associated with a certain home zone and certain associated socio-economic characteristics (Kolkowski et al., 2023).

3. Case study application – Region Stockholm

Region Stockholm has the most extensive multimodal public transportation network in Sweden, covering ca. 2.3 million residents. Two million public transportation journeys occur daily in the Region. The network consists of 5549 stations and 700 lines of metro, bus, commuter rail, light rail transit/tram, and ferries. Fig. 2a presents the public transportation network of the Region and Fig. 2b shows the North, Inner city, and South zones as defined in this study.

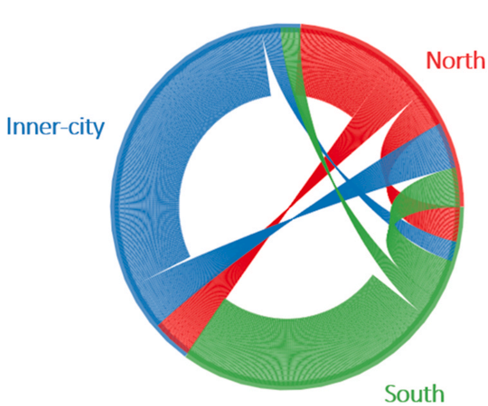

Fig. 4. Student flows among North, Inner city, and South of Stockholm.

Stockholm is located on an archipelago (lakes and waterways make up for 30 % of the total area) between Lake Mälaren and the Baltic Sea. Stockholm is well known for its monocentric planning, and is a great example of a radial public transportation system, given the geographical water barriers that split the built-up area (Cats et al., 2015). Consequently, there are few north-south connections due to the limitations of the local topology. The existing connections create a few well-defined bottlenecks.

Region Stockholm has an open tap-in-only public transportation ticket system, meaning that passengers’ tap-out locations need to be inferred. The tap-in location of a trip j is recorded in the AFC data, and the inference is implemented by searching for a stop within a search radius of the tap-in location of the next trip $j+1$, given that this is made within a specific time window (excluding intermediate activities between trips) and it can be facilitated by matching public transportation lines (Cats et al., 2019). Furthermore, demographic data is available by coupling census data from the Swedish counterparts of census tracts in the United States (*Basområden* or *DeSo* zones).

We analyze travel demand data from passenger journeys made in autumn (September-November) 2022. In total, 60 working days are selected. The time period of several weeks covered by smart card data is deemed sufficient to capture the behavioral patterns of the passenger groups (Goulet-Langlois et al., 2016).

Regarding the selection of the passenger groups, we define two different passenger groups based on (i) the fare product type, and (ii) travel spatial patterns. We choose the following passenger groups based on the aforementioned criteria:

3.1. *School students’ journeys*

There is free school choice in Stockholm, meaning that students can be enrolled in a school outside of their district. School-related journeys represent ca. 25 % of the total demand during the morning and afternoon peak periods, making school students a non-negligible passenger group. We identify school students in the smart card data using the fare product type (school youth ticket). Eligible users are students under

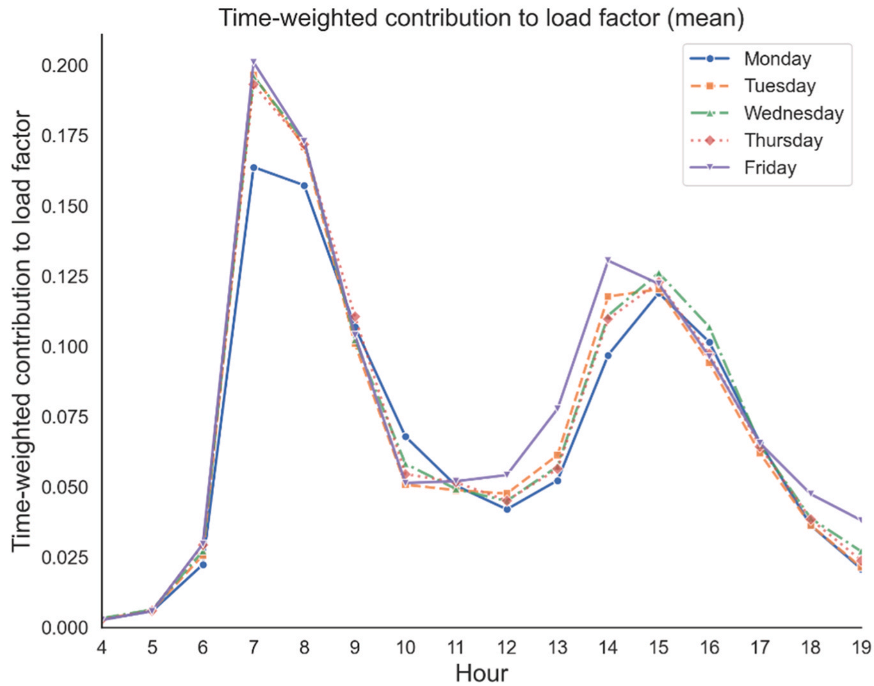

Fig. 5. Time-weighted contribution to load factor metric across time of day for all non-student journeys.

twenty (20) years old. School tickets are valid for journeys conducted during workdays (Monday to Friday) between 04:30–19:00, facilitating school-related trips in Region Stockholm (Stockholms Lokaltrafik (SL), 2022). Our analysis period (60 working days) contains 11,262,038 student journeys and 44,524,054 non-student journeys.

Fig. 3 presents the temporal variation of student and non-student demand across the day.

Students have slightly different AM and PM peaks compared to the rest of the passengers, given the difference between school hours and business hours. During both the students' AM peak (07:00–09:00) and PM peak (14:00–16:00) periods, they make up on average 24 % of the total demand in the network.

Given the free school choice in Region Stockholm, it is relevant to investigate where students travel for school purposes. Fig. 4 presents the student passenger flows within and between the north, inner city, and south zones. The results reveal that the majority of the flows are intrazonal (64 %), and 36 % are interzonal flows, reflecting relatively long travel distances for educational (school) purposes. Last, a significant proportion of these students commute to the city center (Statistiska Centralbyrån (SCB), 2017), making this group particularly interesting as a subject for analysis of crowding contributions.

3.2. Passenger journeys traversing the inner city

The local topology (see Section 3, paragraph 2) and the planning of public transportation services in Stockholm imply that most passengers traveling from the south to the north (and vice versa) must travel through the inner city, using either the metro or the commuter rail (see Fig. 2b), and thereby add to the passenger volumes in this most congested part of the network. In recent years, the establishment of many companies in the central and northern parts of Stockholm (e.g., Solna, Kista, Sundbyberg) and many new residencies built in the southern part have changed the travel patterns between the south and the north (Trafikanalys, 2011). Commuting times using public transportation in Stockholm have increased by 5.4 % (44 minutes per commuting journey in 2015) (Bastian and Börjesson, 2017). In addition, even though passenger journeys traversing the inner city constitute about 9 % of all passenger journeys in our dataset, they affect the vast majority of the rest of the passenger journeys (85 %). A passenger journey is concerned as affected only if the qt_{ig} has a non-zero value. Many journeys are affected since the traversing journeys are relatively long (19.5 kilometers on average, compared to 8.5 kilometers for all journeys), overlapping in time and space with many journeys. For these reasons, this passenger group is of special interest with regard to their potential contribution to crowding.

Journeys traversing the inner city can be identified based on their tap-in and their inferred tap-out location. We identify 5,574,332 such journeys and 55,161,008 other journeys, with roughly equal shares of journeys from the south to the north and those from the north to the south. The share of passenger journeys traversing the inner city is stable at about 11 % on average throughout the day (05:00–22:00).

4. Results

In this section we present the crowding contribution results for the two selected passenger groups. We calculate the crowding contribution metrics for each passenger i not belonging to the selected passenger group g. We use the mean across all such passengers to visualize the results in time and space. Emphasis is put on the AM peak period since it is the time period during which crowding is most prevalent and thus most relevant for policy making.

4.1. School students

The student group contributes to the crowding, in the sense that the time-weighted contribution to load factor is higher than zero, for the overwhelming majority of non-student journeys (94 %). Fig. 5 shows the mean time-weighted contribution to the load factor across time of day $\overline{qt}_{g,d,t}$ (see Eq. 2). Across all non-student journeys, school students utilize, on average, 15 % and 11 % of the seating capacity during the AM peak (07:00–09:00) and PM peak (14:00–16:00) periods, respectively.

To understand how crowding contributions are distributed in space, we aggregate the results in the DeSo zoning system. Specifically, the first aggregation concerns the zones as origins, and the second concerns zones as destinations, both for non-student journeys i. Figs. 6 and 7 present the mean of the time-weighted crowding contribution metric for

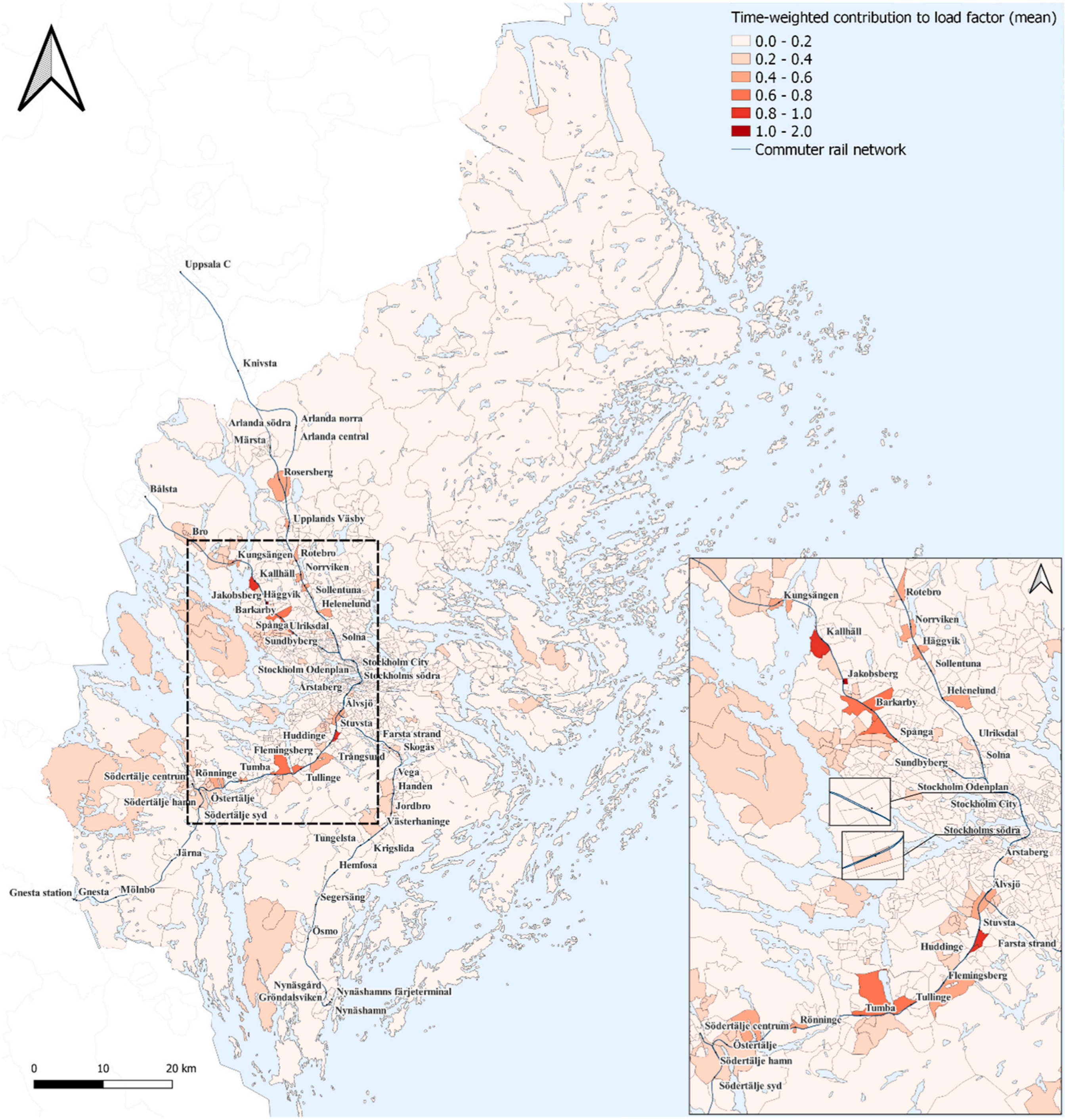

Fig. 6. Spatiotemporal aggregation of the time-weighted contribution to load factor (mean) during AM peak (07:00–09:00) – zones as origins.

the AM peak $\overline{qt}_{g,z,t}$ (see Eq. 3, zones as origins and destinations, respectively).

It can be observed that when results are aggregated based on the origin of the journeys, crowding contributions are, as expected, higher in Stockholm's suburbs than in the city center. Specifically, passengers originating from zones such as Barkarby in the north and Huddinge in the south experience more than 40 % seating capacity utilization by school students throughout their journey. It is also worth mentioning that the origin zones experiencing significant contributions are located in proximity to the commuter rail network. The opposite pattern can be observed when results are aggregated based on the destination of the journeys, i.e. crowding contributions are higher in Stockholm inner city zones (e.g., Stockholm Odenplan, Stockholm City, Stockholms södra) compared to the suburban zones.

It is also relevant to capture the highest contribution to the load factor (see Eq. 4), since passengers can disproportionally recall negative experiences in terms of discomfort. Fig. 8 presents the mean across time of day per transport mode $\overline{f^{\max}}_{g,d,t,m}$ (see Eq. 5). Results indicate that there are no significant differences across the days. Specifically, non-students experience the highest contribution to the load factor when traveling with the commuter rail during both AM (0.72) and PM (0.39) peaks, i.e. indicating that students occupy 72 % and 39 % of the seating capacity in the network segments with the highest student load. Last, the $\overline{f^{\max}}_{g,d,t,m}$ is significantly lower (<20 %) for all other modes across all

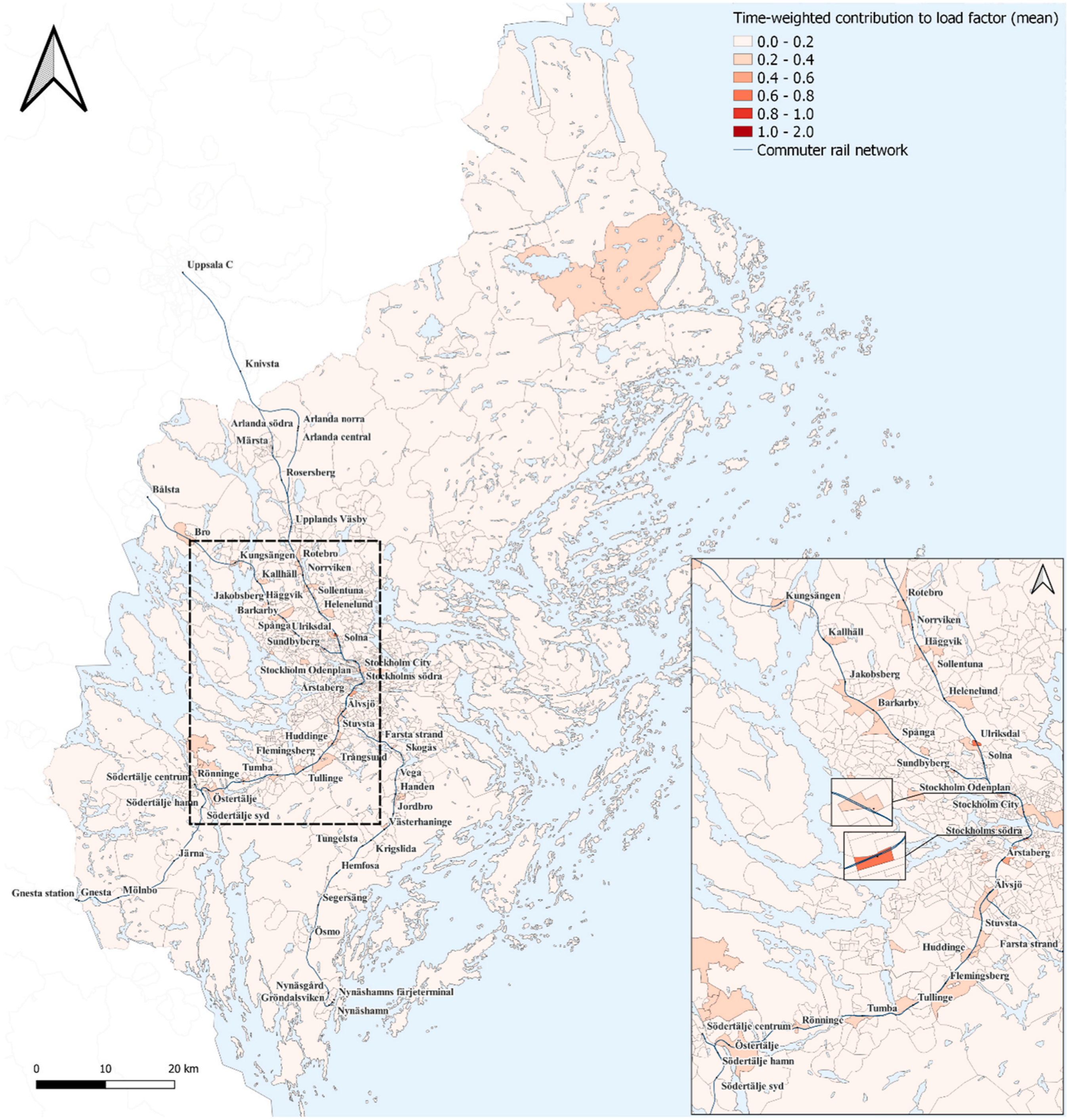

Fig. 7. Spatiotemporal aggregation of the time-weighted contribution to load factor (mean) during AM peak (07:00–09:00) – zones as destinations.

times of the day and days of the week.

4.2. Passengers traversing the inner city

Similar to the school students group, we aggregate the results of the time-weighted contribution to load factor metric $\overline{qt}_{g,z,t}$ per DeSo zone (see Eq. 3). Fig. 9 presents the mean of the time-weighted contribution to load factor metric, aggregated based on the origin and the destination of the non-inner city traversing journeys, for the AM peak. The results show that when the calculations of the metric are aggregated based on the origin of the journeys, significant crowding contributions (more than 40 % seating capacity utilization on average) are spotted both in Stockholm's suburbs (e.g., Huddinge in the south, Barkarby in the north) as well as zones located downstream, in the inner city (e.g., Stockholm södra). When results are aggregated based on the destinations, significant crowding contributions are observed in zones with a high concentration of businesses (e.g., Solna in the north), revealing the commuting pattern during the AM peak. Notably, in both aggregation types (origin and destination of passenger journeys), the zones that experience significant crowding contributions by passengers traversing the inner city (more than 40 % seating capacity utilization on average) are adjunct to the commuter rail network, thereby highlighting the significance of this mode for south-north connectivity.

Fig. 10 presents the mean of the maximum contribution to load

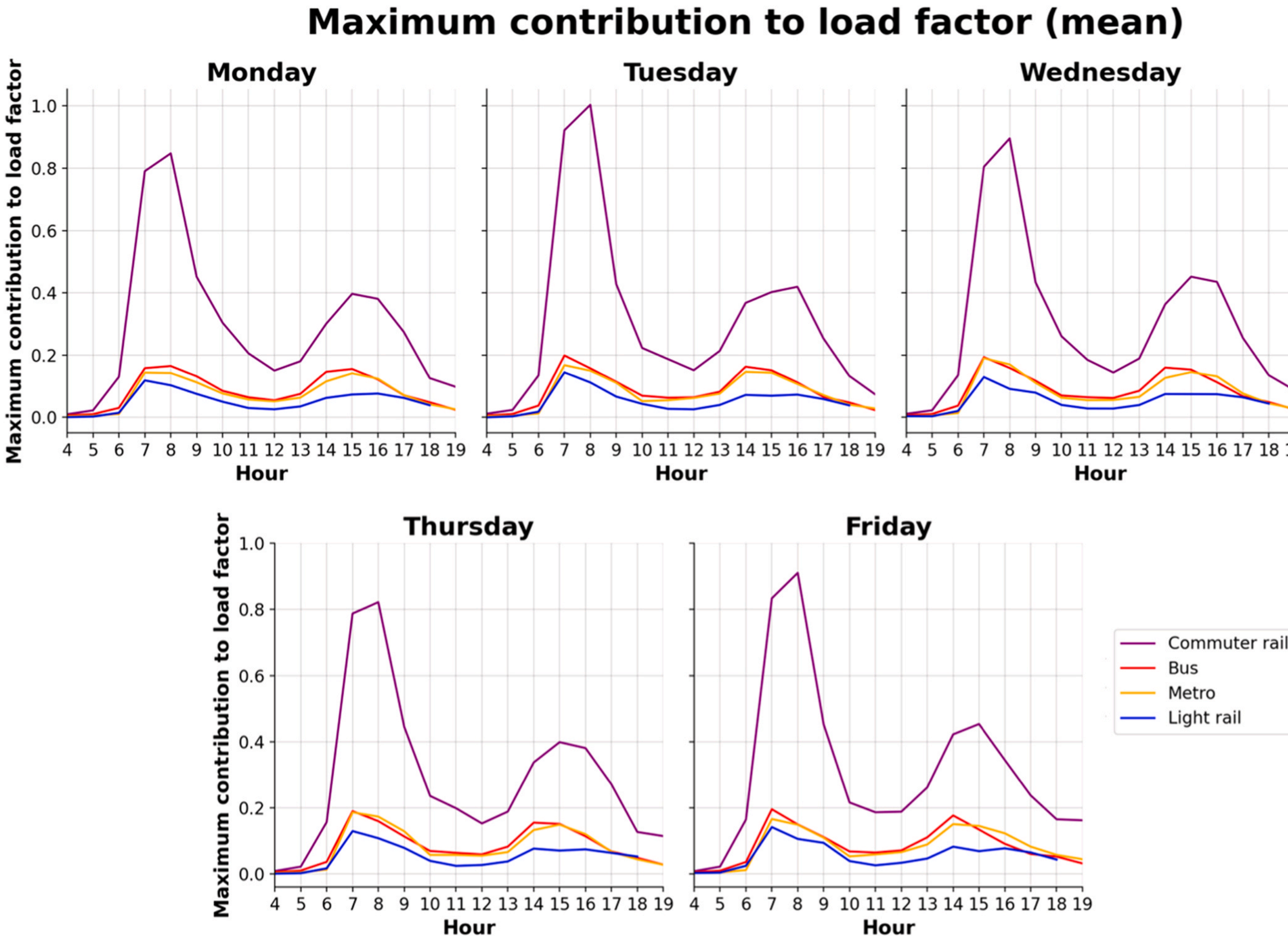

Fig. 8. Maximum contribution to load factor by students across time of day (mean).

factor metric ($\overline{f^{\max}}_{g,d,t,m}$) for different time of the day and each day of the week for the rest of the passenger journeys per travel mode (see Eq. 5). Passenger journeys that do not traverse the inner city experience the highest load factor contribution when traveling by commuter rail. Specifically, passengers traversing the inner city utilize, on the most affected network segment, 90 % and 80 % of the commuter rail seating capacity during the AM and the PM peak, respectively across all days. The $\overline{f^{\max}}_{g,d,t,m}$ is significantly lower, with less than 20 % seating capacity utilization, for the rest of the modes across all times and days of the week.

4.3. Comparison among the passenger groups

We synthesize our results for the different passenger groups in 4.1–4.2 by comparing the distribution of the crowding contribution inflicted by each of these groups on the remaining passengers. Fig. 11 presents the cumulative distribution function (CDF) of the proposed time-weighted contribution to load factor metric for the selected passenger groups during the AM and the PM peaks.

It is evident that the group of inner city traversing passengers has the highest maximum time-weighted contributions to load factor compared to the rest of the groups during AM and PM peaks. Overall, the groups of school students and inner-city traversing passengers have low crowding contributions for a significant portion of passengers not belonging to these groups for both peaks. There is a sharp increase in the number of passengers experiencing moderate crowding contributions (less than 40 % of the seating capacity utilized by the groups). Last, the number of passengers experiencing high crowding contributions (more than 100 % of the seating capacity utilized by the passenger groups) is lower.

5. Discussion and conclusion

Understanding the spatiotemporal dimension of crowding induced by passenger groups is crucial in the era of the ongoing urban agglomerations (The World Bank, 2023). To this end, the proposed method can assist public transportation authorities and operators in tailoring solutions towards on-board crowding reduction. In particular, it can serve as an auxiliary tool providing insight into the groups contributing most to on-board crowding and implications for relevant policymaking as described in the subsequent paragraphs. Smart card data facilitate investigating the travel patterns of individual passengers as well as a group of passengers. We propose a method and a set of metrics for capturing the crowding contributions of passengers. One of the merits of the proposed metrics is that they allow for a comparative assessment across time, space, groups, and networks. We apply the method and metrics for understanding and assessing the on-board crowding contributions of two passenger groups in the Stockholm Region, namely school students, and passengers traversing the inner city.

Our findings show that different groups can have different crowding contributions, given their differences in terms of size and spatiotemporal travel patterns. School students and passengers traversing the inner city have comparable contributions, 15 % during the AM peak and 11 % during the PM peak, despite the two groups constituting of very different shares of the passenger journeys in the respective time periods, 24 % and 9 %, respectively. In particular, the commuter rail network is heavily affected, with the maximum contribution to the load factor metric

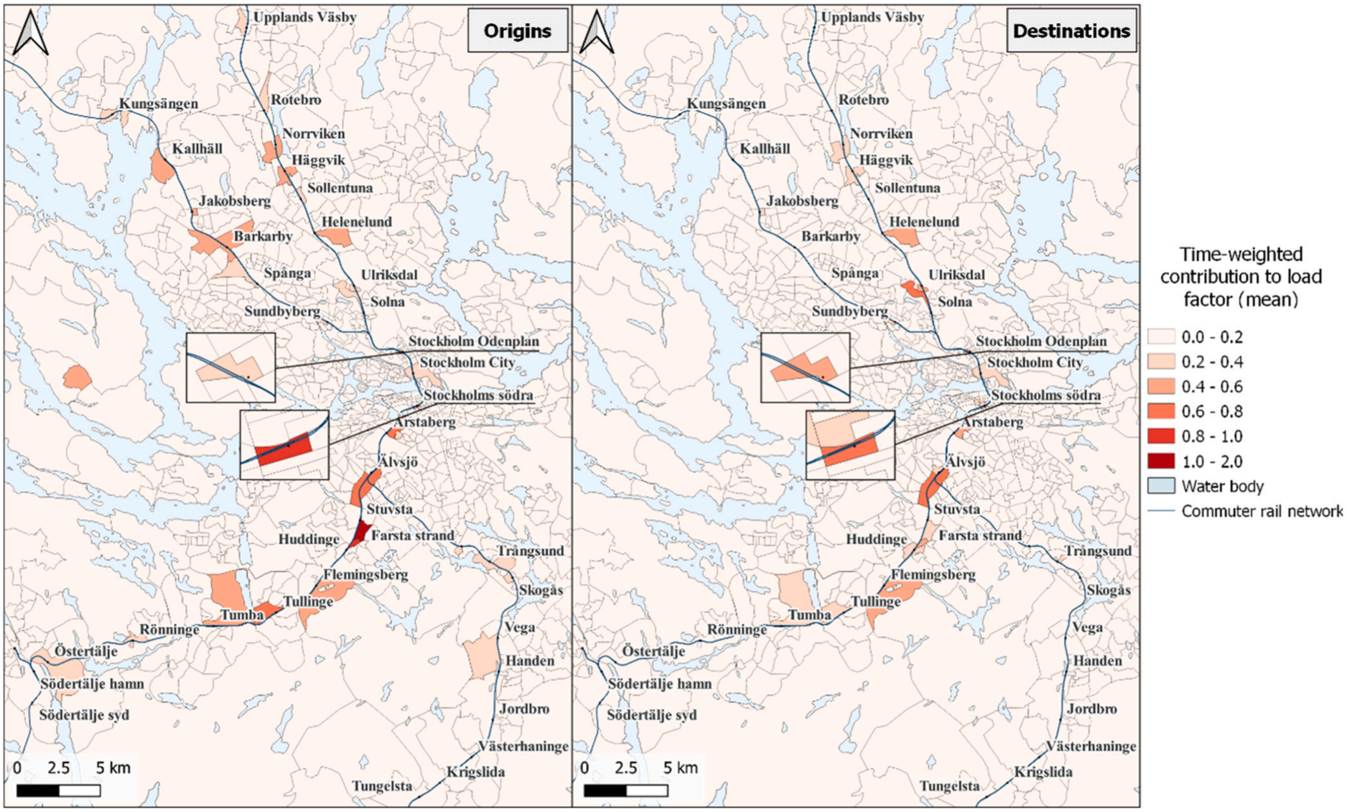

Fig. 9. Spatiotemporal aggregation of the time-weighted contribution to load factor (mean) during AM peak (06:00–08:00).

amounting, on average, to 72 % and 90 % of the commuter rail seating capacity for the school students and passengers traversing the inner-city groups, respectively.

The insights gained from this analysis can facilitate land use policy-making and demand management strategies, thus providing more efficient public transportation services. In the context of students, several initiatives can be undertaken towards better spread of the peak demand. Educational institutions can implement staggered starting times for their classes as a policy for reducing public transportation crowding induced by students (The University of British Columbia (UBC), 2002). Other interventions can target mode choice decisions by establishing school travel plans integrating more alternatives (e.g., active transport modes) and incentives (e.g., different pricing schemes, end-of-trip facilities such as lockers, showers, and safe bicycle parking). Provision of student housing on campus (in the context of university students) or considering proximity in the admission of students to schools (e.g., based on their home postal code) can further reduce the crowding contribution associated with students. Last, results can guide regional authorities in the placement of new schools and business developments so that overcrowding of already crowded network segments can be minimized.

The share of travelers that need to traverse the inner city in order to reach their destination depends on the underlying land-use configurations (Hu et al., 2016). The group of commuters is one of the groups mainly contributing to the on-board crowding. Promoting mixed land-use configurations (e.g., residential, business) can significantly decrease commuting distances (de Abreu e Silva et al., 2012) and, therefore, the public transportation crowding induced by commuters. In Stockholm, between 2004 and 2015 the average distance traveled by public transportation has remained stable at around 15 kilometers, and the travel time has increased by 5.4 % (44 minutes per commuting journey in 2015) (Bastian and Börjesson, 2017), reflecting the ongoing urban sprawl. Alternatively, new public transportation connections, which offer alternatives to the most heavily saturated corridors, are needed in order to relieve the congestion from those corridors, and improve network robustness (Jenelius and Cats, 2015).

The method and metrics proposed in this study can assist in assessing crowding contributions from a variety of passenger groups and compare them across periods and places. For example, the crowding contributions of different user profiles (e.g., tourists, regular commuters, attendees of special events), groups defined by different spatial criteria (e.g., passengers traveling within the inner city), or groups defined by a combination of criteria (e.g., airport travelers by combining ticket fare type and spatial criteria) can be compared across cities as well as the evolution thereof (e.g., during and after the pandemic crisis). For tourists, investigating their crowding contributions can be of added value, given their seasonal and uncertain travel patterns (Domènech and Gutiérrez, 2017). Supply provision can be challenging, especially in touristic cities and regions with a quite constant local passenger flow combined with tourist passenger flows, which tend to be irregular (Gutiérrez et al., 2020). Therefore, understanding their spatiotemporal crowding contributions can result in tailored policy-making to reduce tourist overcrowding (e.g., new fare structure, targeted public transportation investments, etc.). Furthermore, passengers with occasional patterns, such as the attendees of special events (e.g., concerts, festivals, etc.), are good examples of public transportation latent demand that can contribute to overcrowding (Papacharalampous et al., 2016). Special events can attract a significant number of attendees in very short time intervals. Despite their significance, the travel patterns and, therefore, the crowding contributions of special events' attendees have not yet been explored substantially. Applying the proposed method in this context can assist in relevant policy-making to alleviate overcrowding induced by special event attendees. Specifically, measures may include from frequency adjustments of specific public transportation lines and launching shuttle services to synergies between public and active transportation modes (e.g., walking, cycling), and alternate fare products for special events attendees.

In the context of urban planning, planners may apply the metrics to assess the crowding impact of additional travelers associated with urban

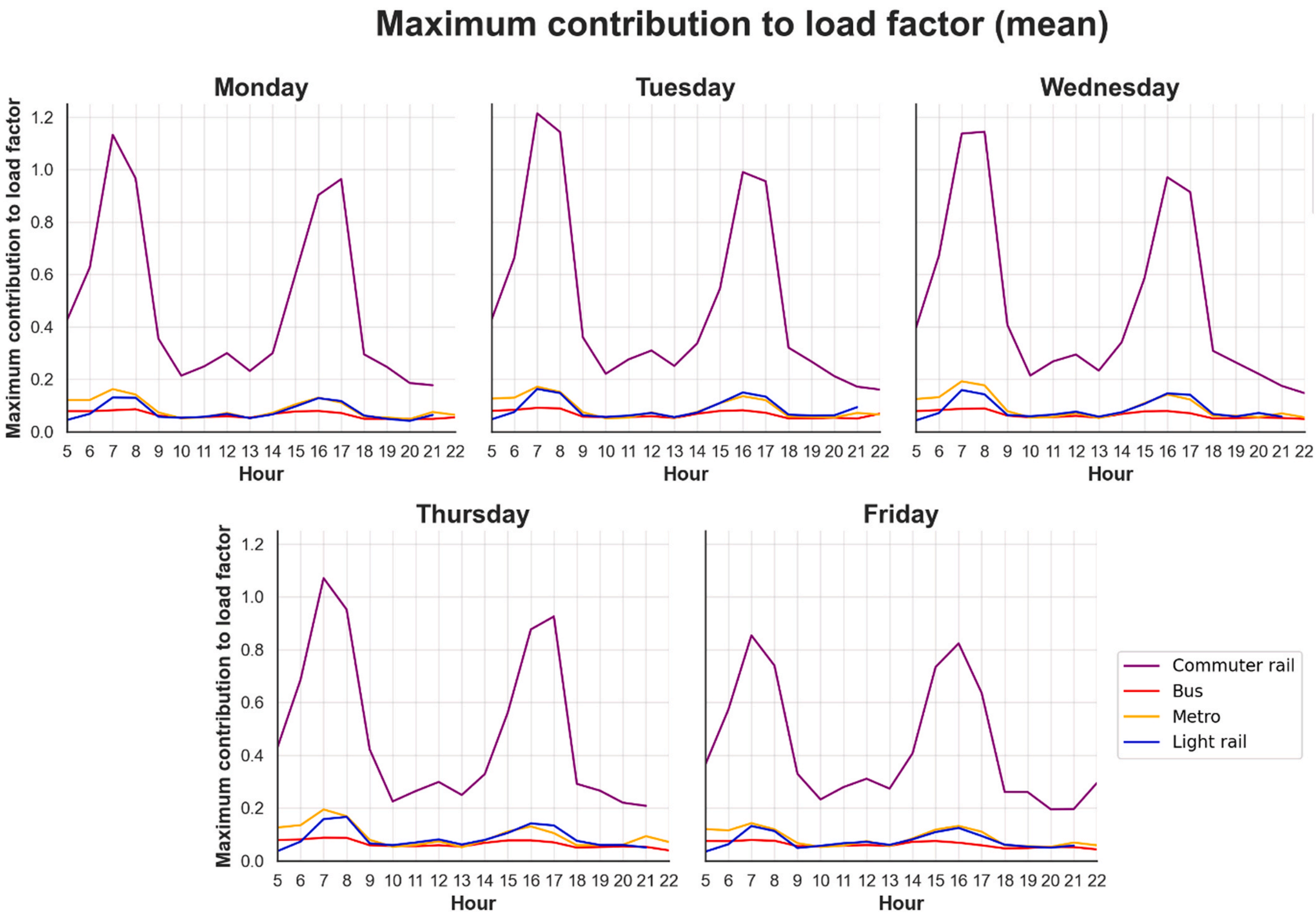

Fig. 10. Maximum contribution to load factor by inner city traversing journeys across time of day (mean).

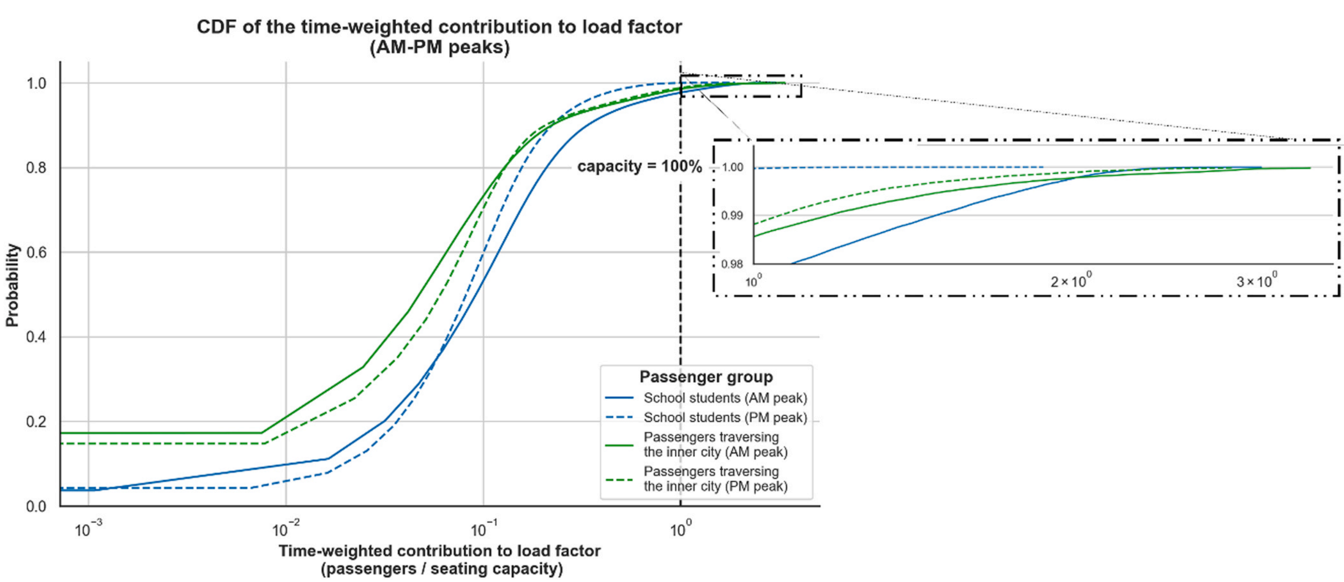

Fig. 11. CDF of the time-weighted contribution to load factor (x-axis on a logarithmic scale, data point: passenger journey).

developments, as demonstrated by Skoufas et al. (2024), who investigated the crowding contributions of one of Northern Europe's biggest ongoing urban developments, revealing critical network links experiencing high crowding contributions. Such results could be of value for designing attractive and efficient public transportation networks connecting to new development areas and assisting in strategic decision-making by integrating crowding contributions as an additional criterion. At the tactical level, measures such as introducing stop-skipping and short-turning service patterns and pricing incentives may be introduced, with the aim of alleviating crowding while catering to the needs of different passenger groups.

In future work, the proposed method can be generalized to quantify the crowding contributions by one group of travelers for any other specific group. For example, one could evaluate the crowding contribution of one group of school students to another group of school students, thereby enhancing its added value and potential useful insight for the competent authorities. Last, regarding the metrics as such, it would be interesting to explore the definition of a subjective crowding contribution metric so passengers' perceptions can be investigated. Such a metric would capture the varying effect of the non-focused passenger loads on the passenger journeys of the selected passenger group, therefore complementing the proposed objective crowding contribution metrics.

CRediT authorship contribution statement

Anastasios Skoufas: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Matej Cebecauer**: Writing – review & editing, Methodology, Conceptualization. **Wilco Burghout**: Writing – review & editing, Supervision. **Erik Jenelius**: Writing – review & editing, Supervision, Methodology, Conceptualization. **Oded Cats**: Writing – review & editing, Supervision, Project administration, Methodology, Funding acquisition, Conceptualization.

Declaration of Competing Interest

None

Acknowledgements

The authors would like to express their gratitude to the Transport Administration at Region Stockholm for providing the necessary data and funding for this study through the "CAPA-CITY: Identifying capacity gaps to support urban and regional development" research project (RS 2022-0210). Map data are copyrighted by OpenStreetMap contributors and available from: https://www.openstreetmap.org.

References

Abenoza, R.F., Cats, O., Susilo, Y.O., 2017. Travel satisfaction with public transport: determinants, user classes, regional disparities and their evolution. Transp. Res. Part A: Policy Pract. 95, 64–84. https://doi.org/10.1016/j.tra.2016.11.011.

de Abreu e Silva, J., Martinez, L.M., Goulias, K., 2012. Using a multi equation model to unravel the influence of land use patterns on travel behavior of workers in Lisbon. Transp. Lett.: Int. J. Transp. Res. 4, 193–209. https://doi.org/10.3328/TL.2012.04.04.193-209.

Basso, F., Frez, J., Hernández, H., Leiva, V., Pezoa, R., Varas, M., 2023. Crowding on public transport using smart card data during the COVID-19 pandemic: new methodology and case study in Chile. Sustain. Cities Soc. 96, 104712. https://doi.org/10.1016/j.scs.2023.104712.

Bastian, A., Börjesson, M., 2017. City a Driv. N. Mobil. Patterns, Cycl. Gend. Equal.: Travel Behav. Trends Stockh. 1985-2015 (9). 〈https://www.transportportal.se/swopec/CTS2017-9.pdf〉.

Börjesson, M., Rubensson, I., 2019. Satisfaction with crowding and other attributes in public transport. Transp. Policy 79, 213–222. https://doi.org/10.1016/j.tranpol.2019.05.010.

Cats, O., 2023. Identifying human mobility patterns using smart card data. Transp. Rev. 44 (1), 213–243. https://doi.org/10.1080/01441647.2023.2251688.

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

Cats, O., Ferranti, F., 2022b. Unravelling the spatial properties of individual mobility patterns using longitudinal travel data. J. Urban Mobil. 2, 100035. https://doi.org/10.1016/j.urbmob.2022.100035.

Cats, O., Rubensson, J., Cebecauer, M., Kholodov, Y., Vermeulen, A., Jenelius, E., & Yusak, S. 2019. *FairAccess - How fair is the fare? Estimating travel patterns and the impacts of fare schemes for different user groups in Stockholm based on smartcard data.*

Cats, O., Wang, Q., Zhao, Y., 2015. Identification and classification of public transport activity centres in Stockholm using passenger flows data. J. Transp. Geogr. 48, 10–22. https://doi.org/10.1016/j.jtrangeo.2015.08.005.

Cheng, Y.-H., 2010. Exploring passenger anxiety associated with train travel. Transportation 37 (6), 875–896. https://doi.org/10.1007/s11116-010-9267-z.

Cheng, Y., Ye, X., Fujiyama, T., 2020. Identifying crowding impact on departure time choice of commuters in urban rail transit. J. Adv. Transp. 2020 (1), 8850565. https://doi.org/10.1155/2020/8850565.

Diec, J., Coxon, S., & De Bono, A. 2010. Designing a public train station shelter to minimise anti-social behaviour and crime in Melbourne's metropolitan rail environment. *ATRF 2010: 33rd Australasian Transport Research Forum.*

Domènech, A., Gutiérrez, A., 2017. A GIS-Based evaluation of the effectiveness and spatial coverage of public transport networks in tourist destinations. ISPRS Int. J. Geo-Inf. 6 (3). https://doi.org/10.3390/ijgi6030083.

Evans, G.W., Wener, R.E., 2007. Crowding and personal space invasion on the train: please don't make me sit in the middle. J. Environ. Psychol. 27 (1), 90–94. https://doi.org/10.1016/j.jenvp.2006.10.002.

General Administration of Quality Supervision Inspection and Quarantine (AQSIQ). 2004. *Safety Specifications for Power Driving Vehicles Operating on Road. General Administration of Quality Supervision, Inspection and Quarantine of the People 's Republic of China.*

Goulet-Langlois, G., Koutsopoulos, H.N., Zhao, J., 2016. Inferring patterns in the multi-week activity sequences of public transport users. Transp. Res. Part C: Emerg. Technol. 64, 1–16. https://doi.org/10.1016/j.trc.2015.12.012.

Gripsrud, M., Hjorthol, R., 2012. Working on the train: from 'dead time' to productive and vital time. Transportation 39 (5), 941–956. https://doi.org/10.1007/s11116-012-9396-7.

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

Hörcher, D., Graham, D.J., Anderson, R.J., 2017b. Crowding cost estimation with large scale smart card and vehicle location data. Transp. Res. Part B: Methodol. 95, 105–125. https://doi.org/10.1016/j.trb.2016.10.015.

Hu, N., Legara, E.F., Lee, K.K., Hung, G.G., Monterola, C., 2016. Impacts of land use and amenities on public transport use, urban planning and design. Land Use Policy 57, 356–367. https://doi.org/10.1016/j.landusepol.2016.06.004.

Jara-Diaz, S.R., Gschwender, A., 2003. Towards a general microeconomic model for the operation of public transport. Transp. Rev. 23 (4), 453–469. https://doi.org/10.1080/0144164032000048922.

Jenelius, E., 2020. Personalized predictive public transport crowding information with automated data sources. Transp. Res. C: Emerg. Technol. 117, 102647. https://doi.org/10.1016/j.trc.2020.102647.

Jenelius, E., Cats, O., 2015. The value of new public transport links for network robustness and redundancy. Transp. A: Transp. Sci. 11 (9), 819–835. https://doi.org/10.1080/23249935.2015.1087232.

Katz, D., Rahman, M.M., 2010. Levels of overcrowding in bus system of Dhaka, Bangladesh. Transp. Res. Rec. 2143 (1), 85–91. https://doi.org/10.3141/2143-11.

Kholodov, Y., Jenelius, E., Cats, O., van Oort, N., Mouter, N., Cebecauer, M., Vermeulen, A., 2021. Public transport fare elasticities from smartcard data: evidence from a natural experiment. Transp. Policy 105, 35–43. https://doi.org/10.1016/j.tranpol.2021.03.001.

Kim, K., Hong, S.-P., Ko, S.-J., Kim, D., 2015. Does crowding affect the path choice of metro passengers? Transp. Res. Part A: Policy Pract. 77. https://doi.org/10.1016/j.tra.2015.04.023.

Kolkowski, L., Cats, O., Dixit, M., Verma, T., Jenelius, E., Cebecauer, M., Rubensson, I.J., 2023. Measuring activity-based social segregation using public transport smart card data. J. Transp. Geogr. 110, 103642. https://doi.org/10.1016/j.jtrangeo.2023.103642.

Lin, J.Y.-J., Jenelius, E., Cebecauer, M., Rubensson, I., Chen, C., 2023. The equity of public transport crowding exposure. J. Transp. Geogr. 110, 103631. https://doi.org/10.1016/j.jtrangeo.2023.103631.

Ministry of Housing and Urban-Rural Development of the People's Republic of China. 2013. *Code for Design of Metro*. 〈https://english.www.gov.cn/state_council/2014/09/09/content_281474986284089.htm〉.

Mohd Mahudin, N.D., Cox, T., Griffiths, A., 2012. Measuring rail passenger crowding: scale development and psychometric properties. Transp. Res. Part F: Traffic Psychol. Behav. 15 (1), 38–51. https://doi.org/10.1016/j.trf.2011.11.006.

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

National Academies of Sciences Engineering and Medicine, 2006. Using Archived AVL-APC Data to Improve Transit Performance and Management. The National Academies Press. https://doi.org/10.17226/13907

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

Sáez, D., Cortés, C.E., Milla, F., Núñez, A., Tirachini, A., Riquelme, M., 2012. Hybrid predictive control strategy for a public transport system with uncertain demand. Transportmetrica 8 (1), 61–86. https://doi.org/10.1080/18128601003615535.

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

Skoufas, A., Cebecauer, M., Burghout, W., Jenelius, E., 2024. Generating and evaluating route choice sets for large multimodal public transport networks: a case study for Stockholm Region (September). 2023 IEEE 26th Int. Conf. Intell. Transp. Syst. (ITSC) 2926–2931. https://doi.org/10.1109/itsc57777.2023.10422419.

Skoufas, A., Cebecauer, M., Burghout, W., Jenelius, E., & Cats, O.. 2024. Ex-post Assessment of Public Transport On-board Crowding Induced by New Urban Development. *103rd Transportation Reseach Board Annual Meeting 2024 (TRBAM 2024).* https://doi.org/10.13140/RG.2.2.12580.45443.

Statistiska Centralbyrån (SCB). 2017. *More common commuting to free school*. ⟨https://www.scb.se/hitta-statistik/artiklar/2017/Vanligare-pendla-till-friskola/⟩.
Stockholms Lokaltrafik (SL). 2022. *This is how the tickets for school youth apply during the holidays*. ⟨https://sl.se/biljetter/sortiment-och-regler/skolungdom/skolbiljetterna-under-lov-och-helger/⟩.
Sun, L., Axhausen, K.W., Lee, D., Huang, X., 2013. Understanding metropolitan patterns of daily encounters Supporting Information Appendix. Proc. Natl. Acad. Sci. 110 (34), 1–27.
The International Association of Public Transport (UITP). 2009. *Bus systems an efficient mode of transport. International Association of Public Transport.*
The World Bank 2023. *Urban Development / Overview*. ⟨https://www.worldbank.org/en/topic/urbandevelopment/overview#1⟩.
The University of British Columbia (UBC). *UBC Transportation Plan*. http://planning.ubc.ca/sites/planning.ubc.ca/files/documents/transportation/plans/UBC-Transportation-Plan-2014_Oct.pdf.
Tirachini, A., 2011. Bus dwell time: the effect of different fare collection systems, bus floor level and age of passengers. Transportmetrica 2011. https://doi.org/10.1080/18128602.2010.520277.
Tirachini, A., Hensher, D.A., Jara-Díaz, S.R., 2010. Restating modal investment priority with an improved model for public transport analysis. Transp. Res. Part E: Logist. Transp. Rev. 46 (6), 1148–1168. https://doi.org/10.1016/j.tre.2010.01.008.
Tirachini, A., Hensher, D.A., Rose, J.M., 2013. Crowding in public transport systems: effects on users, operation and implications for the estimation of demand. Transp. Res. Part A: Policy Pract. 53, 36–52. https://doi.org/10.1016/j.tra.2013.06.005.
Trafikanalys. 2011. *Commuting in Stockholm, Gothenburg and Malmö – a current state analysis*. ⟨https://www.trafa.se/globalassets/rapporter/summary-report/2011-2015/2011/summary-report-2011_3-commuting-in-stockholm-gothenburg-and-malmo—a-current-state-analysis.pdf⟩.
Transit Cooperative Research Program. 2003. *Transit Capacity and Quality of Service Manual*. Transportation Research Board (TRB).
Trépanier, M., Tranchant, N., Chapleau, R., 2007. Individual trip destination estimation in a transit smart card automated fare collection system. J. Intell. Transp. Syst.: Technol., Plan., Oper. 11 (1), 1–14. https://doi.org/10.1080/15472450601122256.
Wardman, M., Whelan, G., 2011. Twenty years of rail crowding valuation studies: evidence and lessons from british experience. Transp. Rev. 31 (3), 379–398. https://doi.org/10.1080/01441647.2010.519127.
Yap, M., Cats, O., van Arem, B., 2020. Crowding valuation in urban tram and bus transportation based on smart card data. Transp. A: Transp. Sci. 16 (1), 23–42. https://doi.org/10.1080/23249935.2018.1537319.
Yap, M., Wong, H., Cats, O., 2023. Public transport crowding valuation in a post-pandemic era. Transportation. https://doi.org/10.1007/s11116-023-10420-1.