Public Transport Experienced Service Reliability: Integrating Travel Time and Travel Conditions

Erik Jenelius

KTH Royal Institute of Technology, Sweden

Abstract

The paper proposes a generalization of public transport passenger-experienced service reliability, incorporating both travel times and travel conditions based on passengers' perceived journey time. Time is partitioned into waiting and transfer time as well as in-vehicle time under different travel conditions (crowding and seat availability), which may vary along a journey and between days. The experienced service reliability gap (ESRG) index is introduced, defined as the difference between an upper percentile (e.g., the 95th) and the median perceived journey time for a particular OD pair and departure time. The metric is evaluated by tracing virtual trips from origin to destination with journey times and travel conditions based on automated vehicle location (AVL) and automated passenger count (APC) data and seated status modelled probabilistically. A study of a high-frequency bus line in Stockholm, Sweden shows that travel conditions co-vary only weakly with nominal journey time, and the ESRG index display patterns across the day not evident in existing reliability measures, such as a wider and later afternoon peak. The ESRG displays significant variation between OD pairs along the line. Correlation with headway variability suggests that measures improving bus regularity have additional positive effects on experienced service reliability.

Keywords:
Public transport, Service reliability, Travel conditions, Perceived travel time, AVL data, APC data

1. Introduction

Public transport systems in many cities around the world are experiencing increasing congestion and crowding. Congestion tends to increase variability and uncertainty of travel times, in particular for buses and other services without exclusive right-of-way. Many studies have shown that travel time reliability is an important attribute for the satisfaction of public transport users (Bates et al., 2001; Friman et al., 2001; Beirão and Sarsfield Gabral, 2007; Cantwell et al., 2009). Uncertain arrival times imply a risk of being late to planned activities (Casello et al., 2009; Benezech and Coulombel, 2013), and uncertainty also has a negative value in itself for risk-averse travellers (Li and Hensher, 2013). Furthermore, irregular arrival times typically lead to uneven passenger loads across vehicles, which in turn increases the irregularity of arrival times and the frequency of bus bunching (Daganzo, 2009; Schmöcker et al., 2016; Andres and Nair, 2017; Wu et al., 2017).

Abkowitz (1978) defines public transport service reliability as "the invariability of service attributes which influence the decisions of travelers and transportation providers". Commonly used measures include schedule punctuality and on-time performance (for low-frequency services) and the coefficient of variation of vehicle headways (for high-frequency services). Although these metrics are

related to the waiting time of travellers, the metrics evaluate reliability primarily from the supplier's rather than passengers' perspective (van Oort, 2014; Diab et al., 2015).

In recent years, the increasing availability of automated vehicle location (AVL) and fare collection (AFC) data has enabled the development of more passenger-oriented reliability measures. Building on work by Furth and Muller (2006) and Chan (2007), Uniman et al. (2010) introduce the "reliability buffer time" (RBT) index, defined for a particular OD pair and time interval as the difference between an upper (e.g., the 95th) percentile travel time and the median travel time across days. The authors evaluate the RBT index for the London Underground based on AFC data from the Oyster smartcard system, where travel times are computed from validations on entry and exit to stations. Ehrlich (2010) evaluates the RBT metric for a selection of London bus routes based on AVL data. Wood (2015) proposes a set of design criteria for a passenger-focused reliability metric, and notes that the RBT metric meets most of the criteria.

In practice, most proposed and used public transport reliability metrics have focused quite narrowly on travel times, even though the definition of Abkowitz (1978) includes the reliability of the provided service in a broad sense. However, there are good reasons to expect that travellers value a reliable public transport service also in terms of travel conditions during the journey (Ceder, 2016). In particular, risk averse travellers will prefer two trips with identical travel conditions over two trips with different conditions, even if the average conditions are equal. It is well established that travellers experience in-vehicle travel time more negatively if the vehicle is crowded, in particular if it is not possible to get a seat (Cantwell et al., 2009; dell'Olio et al., 2011; Raveau et al., 2014; Kim et al., 2015). Furthermore, waiting and transfer time is typically perceived more negatively than nominal in-vehicle time (Abrantes and Wardman, 2011).

The aim of this paper is to develop a reliability metric of passenger-experienced public transport service, where service includes not only nominal journey time, but distinguishes between waiting and transfer time as well as in-vehicle time under different travel conditions (crowding and seat availability). The paper considers the perceived journey time of a trip, defined as the sum of waiting, in-vehicle and transfer time across all trip segments, where each time component is weighted by the time multiplier corresponding to the experienced conditions. The time multiplier may vary along a route, since crowding conditions and seat availability change as passengers board and alight the vehicle. The paper introduces the *experienced service reliability gap* (ESRG) index, defined as the difference between an upper percentile (here the 90th percentile is used) and the median perceived journey time across days for a particular OD pair and departure time. The metric integrates the reliability of journey times and the reliability of travel conditions.

A framework for the evaluation of experienced service reliability is developed that does not require automated fare collection (AFC) data, but uses AVL data for calculating journey time components, and automated passenger count (APC) data about boardings and alightings for inferring in-vehicle travel conditions. A journey by a virtual "probe traveller" is traced from origin to destination at the same start time for each day. Seat availability for a specific traveller is not directly observed from the data, but is modelled probabilistically based on reasonable assumptions about boarding and alighting processes consistent with APC data. The seat allocation model bears similarities to elements in the transit assignment models proposed by Schmöcker et al. (2009), Sumalee et al. (2009) and Hamdouch et al. (2011), with the important distinction that seat probabilities are here computed conditional on observed numbers of on-board, boarding and alighting passengers at each stop while origin-destination flows are not known.

The experienced service reliability gap index is applied and evaluated for a busy high-frequency inner city bus line in Stockholm, Sweden. Patterns across time of day and different origin-destination

pairs, as well as the contributions and interactions between journey time, in-vehicle crowding and seat availability, are investigated. Furthermore, the case study evaluates the relation between the ESRG and the RBT metrics, as well as the headway coefficient of variation (i.e., a traditional supplier-oriented reliability metric).

The remainder of the paper is organized as follows. Section 2 introduces the theoretical framwork and the proposed metric of experienced public transport service reliability, and Section 3 shows how the metric may be computed from AVL and APC data. Section 4 presents the Stockholm case study, with results given in Section 5. Section 6 concludes the paper.

2. Methodology

2.1. Journey time and reliability buffer time

Consider a public transport user who travels on a specific route between two stops repeatedly over a sequence of days. The (nominal) *journey time* T^{n} is the sum of waiting time T^{w} for the first bus with available capacity to arrive, total transfer time T^{tr} between journey segments, and total in-vehicle time T^{iv} on all journey segments. Here, walking time to the origin stop and from the destination stop (access and egress time) is not included in the journey time since it cannot be evaluated with AVL and APC data.

As all elements of the journey time vary unpredictably across days, observed values on a particular day are outcomes of stochastic variables. Conditional on the journey start time, the variability of the journey time across days captures the unreliability of the arrival time to the destination. The *reliability buffer time* (RBT) is defined for a particular OD pair and journey start time as the difference between an upper (e.g., the 95th) percentile journey time and the median journey time across days (Uniman et al., 2010),[1]

$$\mathrm{RBT} = P^{-1}_{T^{\mathrm{n}}}(x) - P^{-1}_{T^{\mathrm{n}}}(50), \tag{1}$$

where $P^{-1}(x)$ denotes the xth percentile. For travellers who commute 20 workdays per month, the 95th percentile implies that equal or longer journey times occur only once per month, while the 90th percentile corresponds to once every two weeks.

The RBT metric can be interpreted as the smallest buffer time in relation to the typical journey time needed for a traveller to arrive at the destination on time on x percent of days (Uniman et al., 2010). This interpretation is only approximate, however, since departing earlier (say, before instead of during the peak hour) may influence the journey time. In any case, the metric captures the spread of journey times between typical and severe days. The RBT may be aggregated to lines and networks by weighting the RBT of each OD pair by the travel demand.

Although the RBT is a passenger-oriented reliability metric, it does not consider the travel conditions during journeys as part of the service attributes. In the following, the service reliability evaluation is extended from nominal journey time to also include experienced travel conditions (waiting time, crowding level and seat availability) in terms of perceived journey time.

2.2. Perceived journey time and experienced service reliability gap

Travellers experience time differently depending on the circumstances under which the time is spent. The perceived journey time of a trip is the length of a trip under nominal travel conditions that

[1] Uniman et al. (2010) originally defined the RBT across a certain journey start time interval rather than a specific journey start time.

is perceived as equal in effort or cost. Several studies have estimated travellers' willingness to pay for shorter travel times under different crowding conditions, typically expressed as time multipliers to in-vehicle time under nominal conditions (Wardman and Whelan, 2011; Li and Hensher, 2011; Tirachini et al., 2017; Björklund and Swärdh, 2017; Haywood et al., 2017). Perceived journey time is used here as an indicator of passenger-experienced public transport service quality, which integrates nominal journey time and encountered travel conditions.

Perceived time is expressed as a positive multiple of nominal (clock-based) time. Let β^{w} and β^{tr} denote the time multipliers for waiting time and transfer time, respectively. Further, in-vehicle time can be partitioned into time under different crowding conditions, and into time spent seated and standing (Sumalee et al., 2009; Leurent et al., 2012). Assume that crowding conditions are discretized into M levels, and let T_m^{sit} and T_m^{std} denote the total in-vehicle time seated and standing, respectively, at crowding level $m = 1, \ldots, M$. Let β_m^{sit} and β_m^{std} denote the time multipliers for in-vehicle time seated and standing, respectively, at crowding level m. The perceived journey time is the total perceived time across all journey segments,

$$T^{\mathrm{p}} = \beta^{\mathrm{w}} T^{\mathrm{w}} + \beta^{\mathrm{tr}} T^{\mathrm{tr}} + \sum_{m=1}^{M} \left(\beta_m^{\mathrm{sit}} T_m^{\mathrm{sit}} + \beta_m^{\mathrm{std}} T_m^{\mathrm{std}} \right). \tag{2}$$

We observe that perceived journey time can be written as a multiple of nominal journey time,

$$T^{\mathrm{p}} = B T^{\mathrm{n}}, \tag{3}$$

where the *journey time multiplier* $B > 0$ is the average time multiplier over the total duration of the journey,

$$B = \frac{\beta^{\mathrm{w}} T^{\mathrm{w}} + \beta^{\mathrm{tr}} T^{\mathrm{tr}} + \sum_{m=1}^{M} \left(\beta_m^{\mathrm{sit}} T_m^{\mathrm{sit}} + \beta_m^{\mathrm{std}} T_m^{\mathrm{std}} \right)}{T^{\mathrm{w}} + T^{\mathrm{tr}} + \sum_{m=1}^{M} \left(T_m^{\mathrm{sit}} + T_m^{\mathrm{std}} \right)}. \tag{4}$$

The journey time multiplier B varies stochastically across days depending on the relative proportions of the waiting, transfer and in-vehicle time components. Thus, every journey with the same proportions between the different time components has the same journey time multiplier, even if the total journey times vary.

Equation (3) implies that variability of perceived journey time arises from two, possibly interrelated sources: nominal journey time, and experienced travel conditions. Unless there is a strong negative correlation between waiting time, transfer time or crowding conditions on one side, and total journey time on the other side, variability in travel conditions tends to increase the relative variability of perceived journey time compared to nominal journey time. In practice, crowding typically leads to longer boarding and alighting times as well as reduced service regularity and waiting time reliability. Thus, one may expect a positive correlation between severity of travel conditions and nominal journey time. This is investigated empirically in Section 4.2.

The distribution of perceived journey time across days captures the variability of experienced public transport service. A relevant indicator of experienced service reliability is the difference between the experienced service on a typical day and on a severe day. Thus, we define the *experienced service reliability gap* for a specific OD pair and journey start time as the span of perceived journey time between typical and severe days,

$$\mathrm{ESRG} = P_{T^{\mathrm{p}}}^{-1}(x) - P_{T^{\mathrm{p}}}^{-1}(50), \tag{5}$$

where x is a desired level of certainty (e.g., 95%). The ESRG metric is a generalization of the RBT metric based on perceived rather than nominal travel time. The term reliability gap is used rather than reliability buffer time since the metric does not share the same clear interpretation of a safety margin.

2.3. Evaluation with AVL and APC data

The experienced service reliability gap metric may be computed from automated vehicle location (AVL) and automated passenger count (APC) data. It is assumed here that AVL and APC data for all relevant lines, stops and vehicles from a sequence of days are available. AVL data contain the arrival and departure times of public transport vehicles for each stop in the studied network. Further, APC data contain the number of passengers boarding and alighting each vehicle at each stop, as well as the passenger load on every line segment. The number of boarding and alighting passengers, respectively, at stop k are denoted b_k and a_k, and the passenger load on the line segment between stops k and $k+1$ is denoted q_k.

AVL and APC data do not contain information about access and egress times and when passengers arrive to stops. Further, there is no information about passenger flows between stops and lines. Hence, the methodology must by necessity make certain assumptions when computing journey start times, in-vehicle travel conditions, transfer times etc.

Experienced travel conditions and journey times are computed by tracing a virtual journey from origin to destination at the same start time for each day. The simulated "probe traveller" does not occupy any space in the vehicles or at the stops, and hence does not influence the boardings, alightings and passenger loads obtained from APC data. The computational method is described in detail in the Appendix. The approach has some similarity to the platform-to-platform RBT (PPRBT) metric proposed for train services by Wood (2015). However, the PPRBT metric considers only nominal journey time and not travel conditions, and journey start times are obtained from AFC data.

The assumption of a fixed journey start time is a consequence of the lack of observed passenger arrival times to stops, but has some associated advantages. First, the variability of journey start times between days, which is an effect of passenger behavior, is removed from the evaluation of service reliability. Second, it allows the reliability metric to be used as a variable in travel behavior analysis, e.g. departure time choice models.

The probe traveller is assumed to board the first departing vehicle for which boarding is not denied, and waiting time is computed as the time difference between the journey start time and the departure time of the boarded vehicle. The calculation of transfer time is analogous and based on the difference between the arrival time and departure time of the two vehicles, assuming that the transfer between the two stops takes a minimum amount of time. Since passengers are only counted once they enter a vehicle, denied boarding can only be inferred in a limited way. Here, boarding is assumed to be denied if (i) the vehicle load after alightings exceeds the total (seated and standing) vehicle capacity, and (ii) no passengers board the bus according to APC data.

The seated status of the probe traveller is modelled probabilistically based on reasonable assumptions about boarding and alighting processes consistent with observed boardings, alightings and passenger loads. The perceived in-vehicle time is computed based on the expected seated status of the traveller along each inter-stop segment. Seat availability at boarding depends on (i) the existing passenger load on the boarded bus, and (ii) the number of passengers boarding the bus at the same stop ahead of the probe traveller. The model makes the same seat priority assumptions as Schmöcker et al. (2009) and Hörcher et al. (2017). Thus, sitting passengers are guaranteed a seat until alighting. Further, remaining standing passengers are given priority over boarding passengers when sitting passengers alight. A distinct difference from the previous studies, however, is that origin-destination passenger flows are not known here.

The number of passengers boarding the bus before the probe traveller at the same stop cannot be directly observed from APC data. However, the probabilistic properties of the number can be derived assuming a random passenger arrival process and FIFO queueing system at the stop. Under these

assumptions, the probability p^{ahd} that a passenger boards the same vehicle before the probe traveller is equal to the ratio between the time elapsed since the preceding bus departure and the headway to the preceding bus. The number of passengers boarding the vehicle ahead of the probe traveller is thus drawn among all boarding passengers b_1 according to a binomial distribution.

The probability that the traveller receives a seat is 0 if the passenger load before boardings, $q_1 - b_1$, exceeds the seat capacity s, and 1 if the passenger load after boardings is lower than capacity. In general, the probability is given by the cumulative distribution function of the binomial distribution,

$$\pi_1^{\mathrm{sit}} = \begin{cases} 0 & q_1 - b_1 \geq s \\ 1 & q_1 \leq s \\ \sum_{b=0}^{s-q_1+b_1} \binom{b_1}{b} (p^{\mathrm{ahd}})^b (1-p^{\mathrm{ahd}})^{b_1-b} & \text{else.} \end{cases} \tag{6}$$

In some settings a random queueing scheme, in which passengers' arrival times to the stop have no impact on the order of boarding the bus, may be more realistic. Such a boarding process can be modelled by a uniform distribution,

$$\pi_1^{\mathrm{sit}} = \frac{b_1}{s - q_1 + b_1}, \quad s < q_1 < s + b_1. \tag{7}$$

Inside the vehicle, seats that become available as passengers alight are filled by currently standing passengers. It is assumed that alighting passengers are randomly selected among all on-board passengers. Thus, the number of seated passengers alighting is drawn among all passengers according to a hypergeometric distribution. Further, the passengers getting a seat are randomly selected among all remaining standing passengers before new passengers board. Conditional on that the probe traveller is standing on segment $k-1$, the probability of getting a seat at stop k is 1 if the passenger load before boardings is lower than the seated capacity. Otherwise, the probability equals the ratio between the number of alighting seated passengers and the remaining number of standing passengers,

$$\pi_k^{\mathrm{sit}} = \begin{cases} 1 & q_{k-1} - a_k \leq s \\ \sum_{a=0}^{a_k} \min\left(\frac{a}{q_{k-1}-a_k-s+a}, 1\right) \frac{\binom{s}{a}\binom{q_{k-1}-s}{a_k-a}}{\binom{q_{k-1}}{a_k}} & \text{else,} \end{cases} \qquad k \geq 2. \tag{8}$$

The probability that the probe traveller is seated on line segment k is equal to the probability that the traveller gets a seat at stop k or at some preceding stop on the same line,

$$p_k^{\mathrm{sit}} = \sum_{k'=1}^{k} \pi_{k'}^{\mathrm{sit}} \prod_{k''=1}^{k'-1} (1 - \pi_{k''}^{\mathrm{sit}}), \quad k = 1, 2, \ldots \tag{9}$$

The expected total journey time seated and standing at crowding level m, which are used to compute the perceived journey time, are computed by summation over all line segments,

$$T_m^{\mathrm{sit}} = \sum_k p_k^{\mathrm{sit}} \mathcal{I}_{km} T_k^{\mathrm{iv}}, \quad T_m^{\mathrm{std}} = \sum_k (1 - p_k^{\mathrm{sit}}) \mathcal{I}_{km} T_k^{\mathrm{iv}}. \tag{10}$$

where $\mathcal{I}_{km}$ is 1 if crowding between stops k and $k+1$ is at level m and 0 otherwise, and T_k^{iv} is the in-vehicle travel time on segment k.

2.4. Aggregation

The ESRG may be aggregated across an interval of journey start times and across OD pairs. Note that ESRG is an absolute reliability metric that in itself tends to increase with the nominal travel time of the trip. Let $\mathrm{ESRG}_n(t)$ denote the experienced service reliability gap computed for journey start time t and OD pair n, and let $\lambda_n(t)$ be the expected passenger arrival rate to the origin stop for OD pair n at time t. Aggregation across time interval T and across all OD pairs in set N can be based on the demand-weighted mean,

$$\mathrm{ESRG}_{NT} = \frac{\sum_{n \in N} \int_{t \in T} \lambda_n(t) \mathrm{ESRG}_n(t) dt}{\sum_{n \in N} \int_{t \in T} \lambda_n(t) dt}. \tag{11}$$

which represents the average ESRG per traveller.

2.5. Possible model refinements

Certain assumptions of the methodology can be refined if OD demand information based on, e.g., AFC data or assignment models is available, such as the seat allocation model (cf. Schmöcker et al., 2009; Hamdouch et al., 2011; Hörcher et al., 2017). Further, some stops are characterized by many passengers transferring from other lines, in which case the assumption of random arrivals is less realistic. If OD demand information is available, the shares of boarding passengers transferring from different lines can be estimated, and their arrival times to the stop can be estimated from AVL data from these lines.

In gated systems such as metro networks, AFC data collected at the ticket gates can be used to infer travellers' arrival times to the stations and the probability of denied boarding (Zhu et al., 2017). This would allow the ESRG index to be computed based on observed journey start times across days as opposed to assuming the same start time every day.

In this paper, the considered experienced service quality attributes are journey waiting time, transfer time, and standing and seated in-vehicle time at various crowding levels. Other service attributes (e.g., driver's driving style, AC availability, quality of real-time information) may also be incorporated in the ESRG, given that two types of information are available: (i) disaggregate data about how the service varies between trips and days, and (ii) estimated time multipliers representing the importance of the service attribute.

The ESRG index may be evaluated separately for specific traveller groups (based on age, income, etc.) if estimated time multipliers are available. Instead of using average time multipliers for the population, the final ESRG index could then be calculated by aggregating the indices for individual user groups.

3. Case study

The case study analyses experienced service reliability along the high-frequency bus line 4 in Stockholm, Sweden, shown in Figure 1.[2] Service reliability is evaluated for all origin-destination pairs in both directions and for all journey start times between 7:30 to 19:00 in 5-minute steps (in total 139 distinct journey start times). The case study focuses on commuting trips and thus uses data from

[2] As described below, APC data are only available for a random sample of approximately 15% of all vehicles. Thus, the share of one-transfer journeys with APC data for both segments is around 2%, which leads to insufficient sample sizes even for a whole year of data. Thus, the study focuses on journeys without transfers.

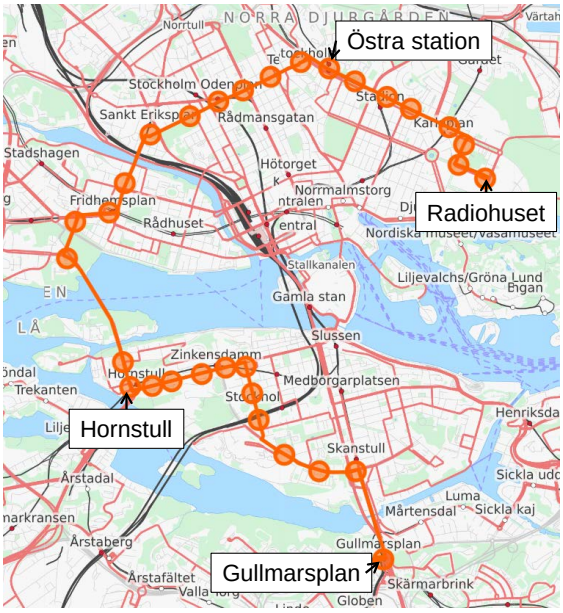

Figure 1: The route of bus line 4, Stockholm. Source: OpenStreetMap.

Monday–Friday and time multipliers representing commuters' valuations. Periods with lower service frequency, including summer, holidays and weekends, are excluded.

Line 4 is ca. 12.4 km long, and consists of 31 stops in one direction (north-south) and 30 stops in the other direction (south-north). The typical vehicle travel time from start to end in one direction is around 60 minutes. Parts of the route are equipped with dedicated bus lanes and/or transit signal priority. Between 7:05 and 18:56 the line is serviced with 4-6 minutes planned headway. The operations during this time period are regularity-based, i.e., each bus seeks to maintain equal headways to the previous and subsequent vehicles on the same line as opposed to following a fixed schedule.

Line 4 is the busiest bus line in Stockholm with around 60,000 boarding passengers per day, and in-vehicle crowding is common. It is rare that passengers take the line from beginning to end (for long trips, the metro or other options are typically faster). Rather, most passengers travel shorter distances and transfer to the metro or other bus lines; the average trip length is around 2.4 km. Figure 2 shows the average passenger load profile in the north-south direction for a trip from start to end at 16:30.

3.1. AVL and APC data

AVL data are available for all buses and all stops on line 4 from 2014 to 2016. APC data are available for the same time period and all stops, but only for about 10% of all buses. The sensors detecting boarding and alighting passengers are installed on a random sample of all buses in Stockholm, and are moved between vehicles at regular intervals. In this study, in-vehicle crowding conditions can thus be evaluated only for days where the first arriving bus is equipped with APC sensors. The set of days for which APC data are available varies depending on the considered origin stop and the journey start time, with an average of 28 days for 2014, 30 days for 2015 and 36 days for 2016. The number of days each year is relatively low for the analysis of travel service reliability, and aggregation over a sliding window (cf. Equation (11)) is generally used to highlight temporal and spatial patterns. However, since the vehicles equipped with APC sensors are selected at random, there is no systematic bias in the selection of days for the evaluation of service reliability.

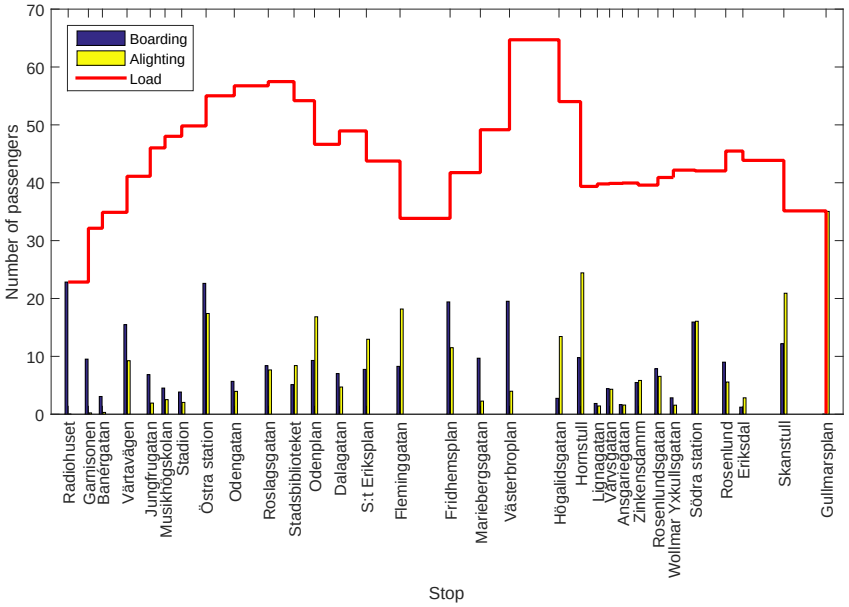

Figure 2: Average passenger load profile in the north-south direction. Journey from Radiohuset to Gullmarsplan at 16:30, Monday–Friday 2016. Inter-stop spacing indicates average travel time.

Table 1: In-vehicle time multipliers (Wardman and Whelan, 2011).

Crowding level m	Load factor (%)	β_m^{sit}	β_m^{std}
1	0–75	0.86	—
2	75–100	0.95	—
3	100–125	1.05	1.62
4	125–150	1.16	1.79
5	150–175	1.27	1.99
6	175–200	1.40	2.20
7	200–	1.55	2.44

3.2. Parameters

The seated and standing capacities of the buses are taken from the public transport planning guidelines for Stockholm (Trafikförvaltningen, 2016). According to these, the seated and total capacity of the articulated buses used on line 4 is $c^{\text{sit}} = 45$ and $c^{\text{tot}} = 120$ passengers, respectively.

Time multipliers for in-vehicle time at different crowding levels are adopted from the meta-study of Wardman and Whelan (2011). In particular, the multipliers for commuters, shown in Table 1, are used here.[3] For waiting, the time multiplier $\beta^{\text{w}} = 2.0$ is used (Wardman, 2004).

[3]These parameters, as with most crowding valuation studies in the literature, were estimated in the context of train crowding, but we assume here that they are reasonably valid also for bus crowding. An alternative would be to use crowding valuations based on standees per square meter (e.g. Björklund and Swärdh, 2017); however, information about standee surface area in the buses was not available for this study.

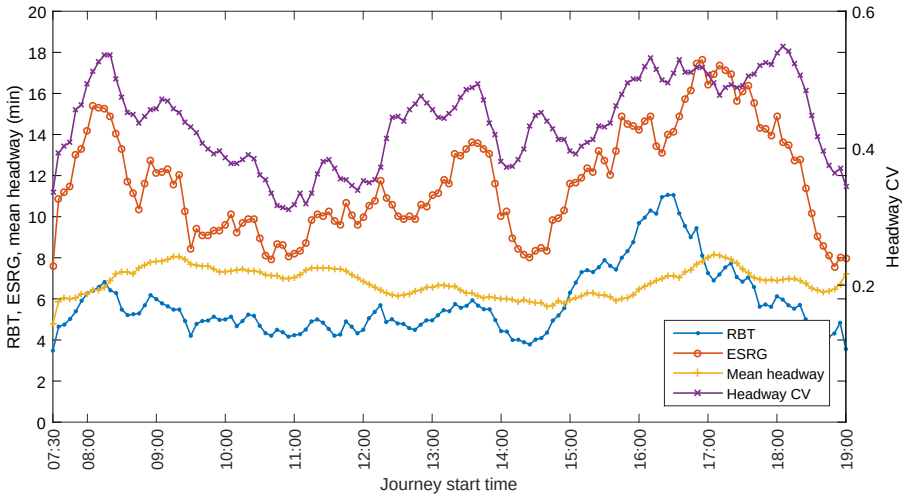

Figure 3: Reliability buffer time, experienced service reliability gap, mean headway, and headway coefficient of variation. Journey from Östra station to Hornstull, Monday–Friday 2016. Moving average with bandwidth 35 min.

4. Results

4.1. Variations across the day

A particular origin-destination pair, from Östra station to Hornstull, is chosen for analysis. The OD pair traverses a sufficient number of stops (13 stops including origin and destination) to capture fluctuations in travel conditions along the journey. Both origin and destination stops are located in connection to metro stations. The origin stop is close to the main campus of KTH Royal Institute of Technology, where travel demand tends to be high at the end of school and work days.

Figure 3 shows the reliability buffer time (RBT) and experienced service reliability gap (ESRG) across the day. Each data point represents an average over a sliding window of 35 min bandwidth that highlights the main temporal patterns. Nominal journey time varies generally around five min and at most ca 11 min between normal and severe days. Experienced journey time, meanwhile, varies generally around 10 min and up to 17 min. The difference between RBT and ESRG is particularly large during peak hours, when both congestion and in-vehicle crowding are high. It is notable that the peak in ESRG is wider and reaches its maximum around 30 min later than than the peak in RBT. This may indicate that the peak rush hour for private road users (who influence bus travel times through congestion and hence longer travel times) occurs slightly earlier than for public transport users.

Figure 3 also shows the mean actual headway and headway coefficient of variation at the journey start stop Östra station. The mean headway varies from about 6 minutes during early morning, early afternoon and evening to about 8 minutes during late morning and late afternoon. The coefficient of variation of vehicle headways is a traditional operator-oriented reliability metric. Large headway variability is expected to increase the variability of waiting time and in-vehicle crowding conditions. Headway variability displays a similar pattern across the day with the highest values in the morning and afternoon peaks. The wide afternoon peak may contribute to the wider peak in ESRG compared to RBT. Interestingly, headway variability remains high even until after 18:00 when nominal journey time variability has returned to typical levels. This suggests that there is a lag in the bus operations before the service can return to more regular headways.

To analyze the factors contributing to the reliability metrics, Figure 4 shows the median and 90th

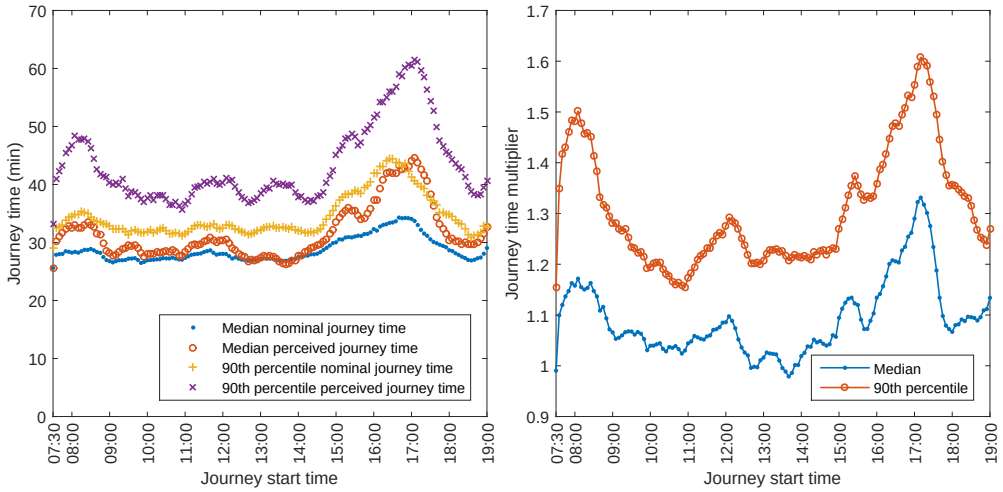

Figure 4: Left: Nominal and perceived journey time, median and 90th percentile across days. Right: Journey time multiplier, median and 90th percentile. Journey from Östra station to Hornstull, Monday–Friday 2016. Moving average with bandwidth 35 min.

percentile nominal and perceived journey time from Östra station to Hornstull in 2016 as functions of the journey start time, from 7:30 to 19:00. The median nominal journey time, representing the typical actual journey duration, is relatively stable around 28 minutes throughout the day, with a moderate afternoon peak slowly building from 14:00 to around 34 minutes at 17:00, and then dissipating until around 18:30. The 90th percentile nominal journey time, representing the longest journey time occurring at least once every two weeks, is at minimum five min longer than the median journey time at the same time of day, and has a sharper peak in the afternoon. The peak journey time is around 42 minutes and occurs at 16:30, ca. 30 min earlier than the peak on the typical day.

Perceived journey time shows greater within-day and between-days variability than nominal journey time. The median perceived journey time remains close to the median nominal journey time from late morning to early afternoon. Thus, travel conditions are close to nominal during this period. During the morning and particularly the afternoon peaks, perceived journey time is considerably higher than nominal journey time, at most around 45 minutes. The 90th percentile perceived journey time is consistently around 10–15 minutes longer than the median, and reaches up to over 60 minutes during the afternoon peak. Thus, perceived journey time may be twice as high on a severe day during peak hours than on a typical day during mid-day hours.

Figure 4 also shows the median and 90th percentile journey time multiplier for each journey start time. The journey time multiplier represents the combined impact of waiting time and in-vehicle time standing and seated at different crowding levels. On both typical and severe days, the time multiplier is the highest during the morning and afternnoon peak hours. Smaller peaks are visible around noon and just after 15:00, where the latter is likely due to many classes at KTH and other nearby schools ending at 15. On typical days, the time multiplier reaches at most around 1.3 in the afternoon peak, implying that perceived journey time is 30% longer than nominal journey time. The temporal pattern is similar for severe days, but the level is consistently higher and perceived journey time is up to 60% higher than nominal journey time in the afternoon peak.

The trips corresponding to the median and 90th percentile perceived journey times from Östra station to Hornstull during 2016, respectively, for each journey start time are studied in closer detail. Figure 5 shows the share of total journey time spent waiting and in-vehicle at different crowding levels

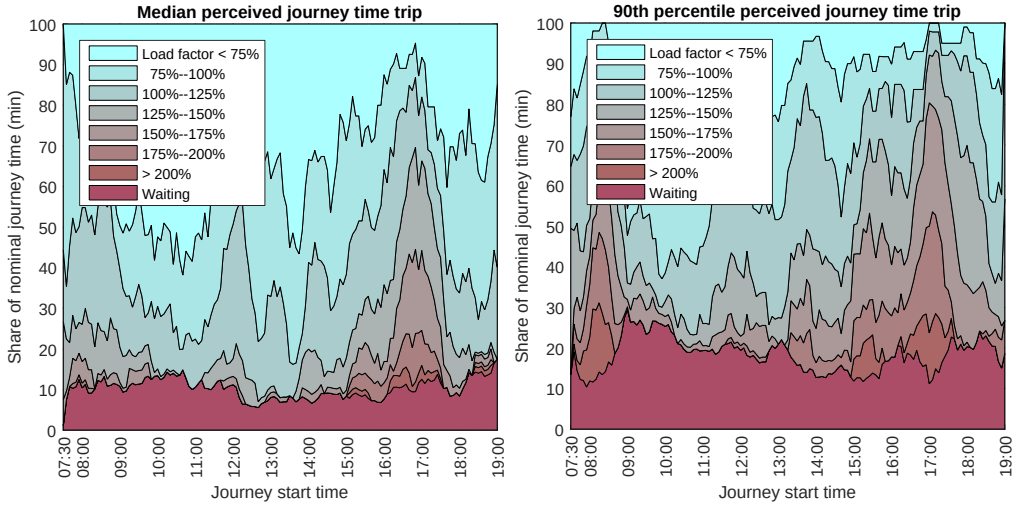

Figure 5: Share of nominal journey time at different crowding levels Left: Median perceived journey time trip. Right: 90th percentile perceived journey time trip. Journey from Östra station to Hornstull, Monday–Friday 2016. Moving average with bandwidth 35 min.

(Table 1) for both days. The severe day is generally characterized by larger proportions of journey time spent waiting for the bus (around 20%) than the typical day (around 10%), in particular during mid-day hours. On both typical and severe days, crowding is high during the morning peak, lower during mid-day hours, and at the highest levels during the afternoon peak. As expected, crowding levels are higher on the severe day, which manifests themselves in higher time multipliers of nominal time. Notably, there are peaks in crowding conditions around 14:00 and 15:00 in the afternoon on the severe day which are not as apparent on the typical day.

Figure 6 shows the share of total nominal journey time spent waiting, seated in-vehicle, and standing in-vehicle, respectively, during the same two trips. Not surprisingly, periods with high shares of in-vehicle standing closely correspond to periods with high crowding levels (cf. Figure 5). Peak hours imply that some proportion of journey time is spent standing on both the typical and the severe day, but the afternoon period where some standing is required is considerably wider on the severe day. Also in the morning peak and around noon, the share of journey time spent standing is higher on the severe day than on the typical day.

4.2. Relations between reliability metrics

Perceived journey time variability arises from variability in nominal journey time and in travel conditions, and total variability depends on the correlation between the two factors. Figure 7, left, shows the RBT against the standard deviation of the journey time multiplier for each journey start time. There is a positive relation between the two, although relatively weak ($R^2 = 0.241$). Thus, at times with higher travel time uncertainty there is also a somewhat higher uncertainty in travel conditions. The correlation may arise from dwell times which are longer when more passengers board and alight the vehicle, as well as from a more general connection across the day between on-road congestion, which causes larger travel time uncertainty, and public transport crowding. Still, much of the variability of travel conditions comes from other sources than journey time variability.

Figure 7, right, shows the RBT and ESRG indices against each other for each journey start time. As expected, there is a clear relation between the two ($R^2 = 0.457$). Still, a significant portion of

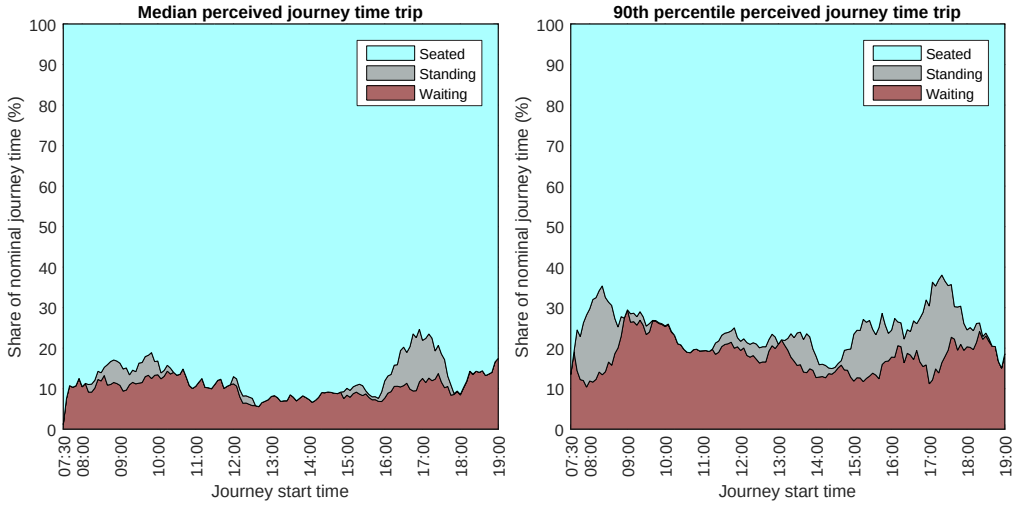

Figure 6: Share of nominal journey time waiting, standing and seated. Left: Median perceived journey time trip. Right: 90th percentile perceived journey time trip. Journey from Östra station to Hornstull, Monday–Friday 2016. Moving average with bandwidth 35 min.

the variability in ESRG comes from the variability in travel conditions which is not correlated with nominal journey time variability. In other words, the ESRG index captures aspects of service reliability that are not captured by the RBT index, either directly or indirectly.

Figure 8 shows the RBT and ESRG metrics against the headway coefficient of variation at the journey start station Östra station for each journey start time. The ESRG is more strongly related ($R^2 = 0.457$) than the RBT ($R^2 = 0.241$) to the headway coefficient of variation. This is not surprising since headway variability implies variability of waiting times[4], which are associated with a high time multiplier. Furthermore, headway variability implies variability in passenger loads and seat availability between buses.

4.3. Monitoring over years

An important application of service reliability metrics is the monitoring of performance over time. Figure 9 shows the RBT (left) and ESRG (right) indices across the day for each year from 2014 to 2016. The RBT metric displays a similar trend over the day for all three years, with the exception of the significant peak in the afternoon during 2016 which is absent during previous years. The ESRG metric displays no such marked deviation between years, but reliability is with only few exceptions higher during 2015 compared to 2014 and 2016.

To analyze the mechanisms behind the differences between years, Figure 10 shows the median and 90th percentile nominal and perceived journey time for each journey start time separately for each year. Comparison between years shows that both percentiles of the nominal journey time remained relatively constant during all years, except for an increase in the afternoon peak in 2016. Perceived journey times, meanwhile, increased more significantly in 2016 compared to the previous two years, mainly due to more severe travel conditions. Time multipliers are generally highest during 2016, in particular during peak hours and around noon, and lowest during 2015. In other words, travel conditions worsened

[4]Conditional on the journey start time, the squared waiting time coefficient of variation is proportional to the squared headway coefficient of variation. The expected waiting time is equal to half the expected headway.

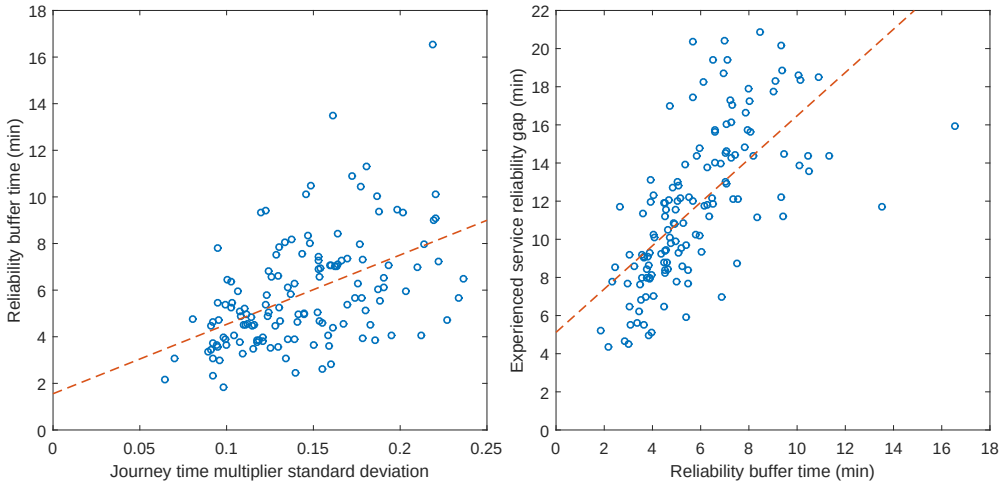

Figure 7: Left: RBT (min) against journey time multiplier standard deviation ($R^2 = 0.241$, slope t-stat $= 4.85$). Right: ESRG (min) against RBT (min) ($R^2 = 0.457$, slope t-stat $= 10.74$). Each point represents a five-minute interval from 7:30 to 19:00. Journey from Östra station to Hornstull, Monday–Friday 2016.

during 2016 on both typical and severe days, and experienced service reliability therefore remained at a relatively constant level throughout all three years.

4.4. Variations across OD pairs

Service reliability is analyzed for all origin-destination pairs along bus line 4.[5] Figure 11 shows reliability buffer time (left) and experienced service reliability gap (right) isocurves given journey start time 16:30 for every OD pair. Both reliability indices vary significantly along the line; ESRG ranges from around 8 minutes for some short journeys (e.g, start stop Östra station to end stop Roslagsgatan) to 26 minutes for some distant OD pairs (e.g., start stop Stadsbiblioteket to end stop Rosenlund). Service reliability levels are often quite different along the two line directions from the same origin stop, which reflects the variability of traffic and travel conditions along the line. Unreliability on one part of the line tends to propagate downstream and to the other direction of the line in later time intervals. Analysis of different journey start times, however, show that the spatial patterns are quite stable during the afternoon peak.

Figure 12 shows the median and 90th percentile journey time multipliers given journey start time 16:30 for every OD pair. Multipliers are largest for short journeys, which has two reasons. First, the shorter the journey, the larger share of the journey consists of waiting time on average. Second, the chance of getting a seat increases for each traversed stop where passengers alight, which means that the share of the journey spent standing is larger for shorter journeys on average.

The journey time multiplier varies significantly depending on the start stop of the journey. On typical days (Figure 12, left), multipliers are particularly low towards the beginning of the line, where headway variability and travel demand are relatively low, as well as around stop Fleminggatan about halfway along the line. Here many passengers alight (cf. Figure 2), which means that boarding at the stop leads to a high chance of being seated and low experienced crowding levels for short trips. On

[5]One stop (Lignagatan) exists only for the north-south line direction and is omitted from the figures for presentation clarity.

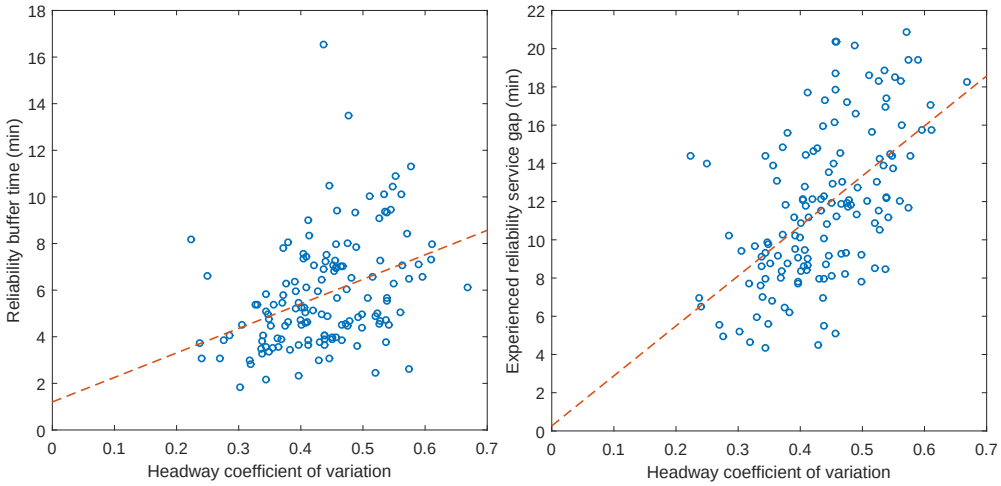

Figure 8: Coefficient of variation of headway against (left) RBT ($R^2 = 0.149$, slope t-stat $= 4.91$), and (right) ESRG ($R^2 = 0.302$, slope t-stat $= 7.70$). Each point represents a five-minute interval from 7:30 to 19:00. Journey from Östra station to Hornstull, Monday–Friday 2016.

severe days (Figure 12, right), perceived journey time is more than twice the nominal journey time for some OD relations. Even for long trips, perceived journey time may exceed nominal journey time by 40%. Like on typical days, travel conditions are least severe when boarding in the beginning of the line or at Fleminggatan in the middle. Similar patterns in travel conditions occur in both directions of the line.

Figure 13 shows the 90th percentile nominal and perceived journey time isocurves, respectively, for every OD pair given journey start time 16:30. Nominal journey times are relatively proportional to the lengths of different OD pairs, although some locations where the isocurves bend vertically can be identified. These locations coincide with well known bottlenecks where buses are often obstructed by conflicting traffic flows. Perceived journey times show more variation along the line. For some destinations, boarding at an upstream stop (e.g., Östra station, indicated with "Ö", in the south direction) can even lead to a lower perceived journey time than if boarding at a downstream stop at the same journey start time. This seemingly paradoxical phenomenon is mainly due to a higher probability of getting a seat when boarding upstream on the line. Analysis shows that the effect is less apparent on typical days.

5. Discussion and conclusion

The paper has proposed a generalization of passenger-oriented public transport service reliability evaluation, considering service quality in terms of travel conditions in addition to nominal travel time. Specifically, a travel time reliability metric with many attractive properties, the reliability buffer time (RBT), is extended to capture varying travel conditions based on passengers' perceived travel time. The paper introduces the experienced service reliability gap (ESRG) index, defined as the difference between an upper percentile (e.g., the 90th percentile used in the case study) and the median perceived journey time for a particular OD pair and departure time. Distinction is made between waiting and transfer time as well as in-vehicle time under different travel conditions (crowding and seat availability), which may vary along a journey and between days. By fixing the start time, the variability of experienced service due to passenger behavior is removed from the metric.

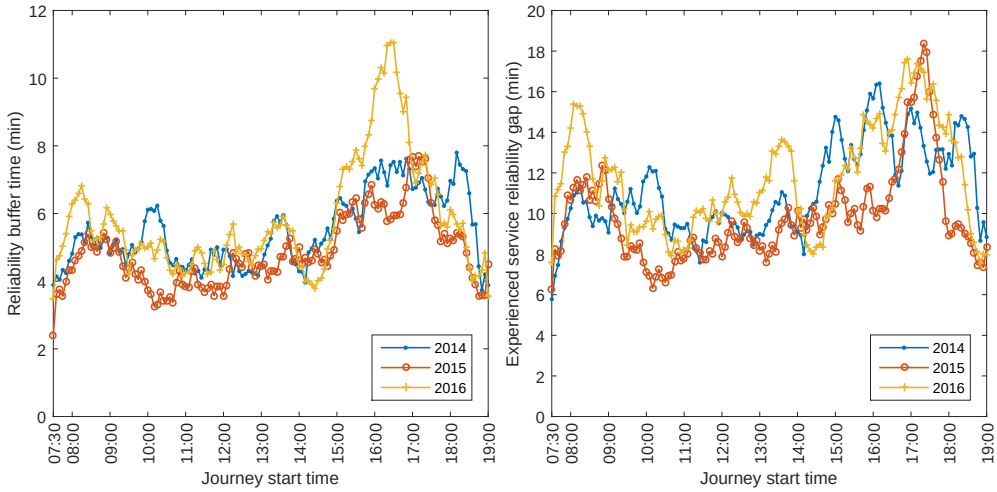

Figure 9: Time evolution 2014–2016. Reliability buffer time (left) and experienced service reliability gap (right). Journey from Östra station to Hornstull, Monday–Friday. Moving average with bandwidth 35 min.

The metric is evaluated by tracing virtual "probe traveller" journeys from origin to destination across multiple days, where journey times and travel conditions each day are consistent with AVL and APC data. Seated status is not directly observed from data and is treated probabilistically based on intuitive and easily adaptable models of passenger behaviour. This evaluation method represents a methodological contribution that is more general than the application to service reliability. For example, it may be used to compare passengers' nominal and perceived journey times in before-after evaluations of public transport investments, policies and planning schemes, etc.

The case study for a high-frequency bus line in Stockholm shows that perceived journey time is higher than nominal journey time particularly during afternoon peak, where in-vehicle crowding is significant and waiting times relatively long due to large headway variability. There is a weak but statistically significant positive relation between variability in nominal journey time (i.e., the RBT) and in travel conditions (represented by the journey time multiplier). In other words, the ESRG index captures aspects of service reliability not captured from nominal journey time alone. Peak ESRG occurs later than peak RBT in the afternoon, since in-vehicle crowding and waiting times remain high even after on-road congestion has cleared. Experienced service reliability also displays significant variation between OD pairs along the line.

The case study illustrates how experienced service reliability may be monitored over time for particular OD pairs. Evaluation over several years shows that travel conditions deteriorated in 2016 compared to previous years on both typical and severe days, so that ESRG remained at similar levels. The analysis may be extended to larger networks containing multiple lines as well as journeys involving transfers between lines. By suitable aggregation over OD pairs, passenger-experienced service reliability can be evaluated for lines, corridors and regions. A particularly promising application of the ESRG index is to evaluate the impacts of policy measures and investments, such as changes in service frequency or vehicle capacity, and mobility management actions such as dedicated bus lanes or transit signal priority.

The ESRG index is significantly correlated (more so than the RBT) with the headway coefficient of variation. This has direct policy implications, since headway variability can be regulated with various strategies, including bus holding (Daganzo, 2009), transit signal priority (Hounsell and Shrestha,

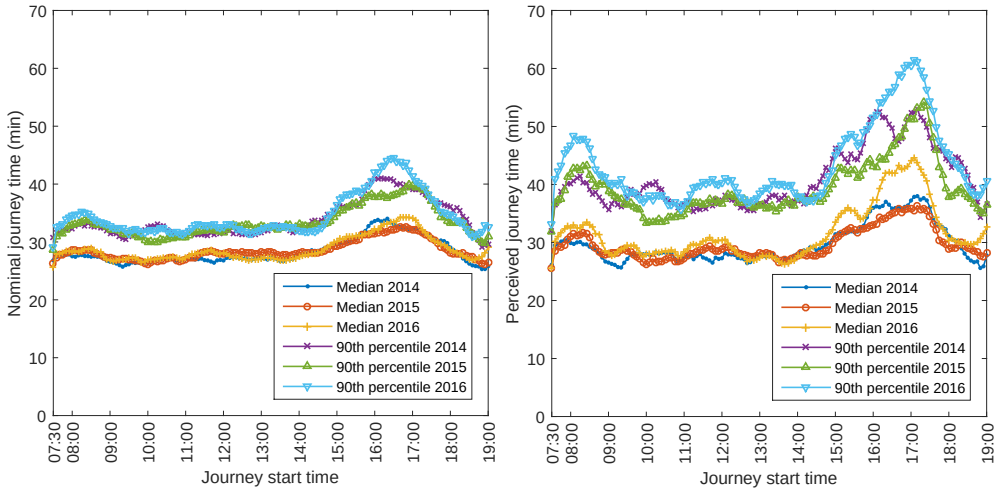

Figure 10: Time evolution 2014–2016. 90th percentile nominal (left) and experienced (right) journey time. Journey from Östra station to Hornstull, Monday–Friday. Moving average with bandwidth 35 min.

2012), and short-turning (Leffler et al., 2017). The results suggest that such actions have wider benefits for experienced service reliability than have previously been considered. Other types of actions may be considered to more directly target the reliability of in-vehicle travel conditions, for example real-time crowding information provision (Zhang et al., 2017).

A limitation of the proposed analysis is that not all relevant attributes can be observed from the considered automated data sources, including travellers' journey start times and ocurrences of denied boarding. In gated systems such as metro networks, AFC data collected at the ticket gates can be used to fill some of these gaps (e.g., Zhu et al., 2017; Hörcher et al., 2017). In bus networks where data are collected at boarding, these opportunities are more restricted; however, OD matrices inferred from AFC data could be used to improve, e.g., the probabilistic seat allocation model.

Journeys with transfers were not considered in this case study due to limited data availability, and is an area for further analysis. Another important topic for further investigation is to what extent travellers take the variability of journey time and travel conditions into account in their travel decisions, such as departure time and route choices. This may be studied by evaluating the ESRG index for multiple travel alternatives (e.g., routes and departure times) and identifying travellers' choices based on, e.g., AFC data.

Acknowledgement

The author would like to thank the Stockholm City Council Transport Administration for kindly providing the data. The work was funded in part by TRENoP Strategic Research Area. The author would also like to thank participants at the 7th INSTR in Sydney, 2018, and several anonymous referees for highly valuable comments and suggestions on an earlier version of the paper.

Appendix: ESRG evaluation with AVL and APC data

Consider a traveller whose journey starts at time t and consists of $L \geq 1$ journey segments on different lines, and $L-1$ intermediate transfers. The traveller boards line $l = 1, \ldots, L$ at stop k_l^{b} and alights at stop k_l^{a}. The same journey is repeated over multiple days. The arrival and departure

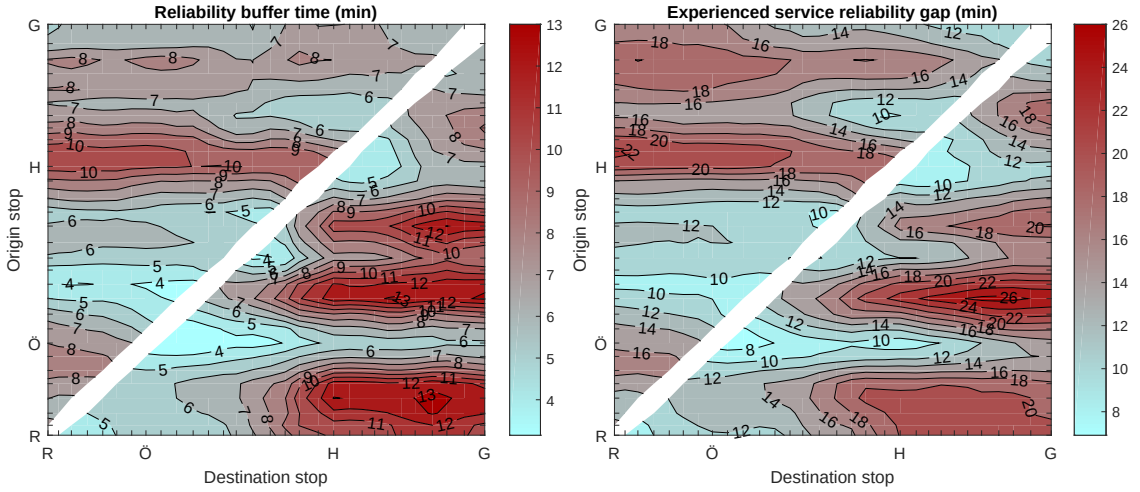

Figure 11: Reliability buffer time (left) and experienced service reliability gap (right) across OD pairs. Journey start time 16:30, Monday–Friday 2016. Moving average with bandwidth 3 stops. Inter-stop spacing indicates average travel time.

times, respectively, on day i of vehicle j from stop k along line l are denoted t^{arr}_{ijkl} and t^{dep}_{ijkl}. The number of boarding and alighting passengers, respectively, on day i and vehicle j at stop k along line l are denoted b_{ijkl} and a_{ijkl}, and the passenger load on the line segment between stops k and $k+1$ is denoted ℓ_{ijkl}.

Waiting time

The considered journey starts at stop k^{b}_1 on line 1. Each day the probe traveller boards the first non-full vehicle $j_{i,1}$ arriving to k^{b}_1 after the journey start time t, identified from AVL and APC data. A vehicle is considered to be full on arrival if (1) the passenger load after alightings exceeds total (seated and standing) vehicle capacity, and (2) no passengers board the bus. Waiting time T^{w}_i on day i is computed as the time difference between the journey start time t and the vehicle departure time $t^{\mathrm{dep}}_{ij_{i,1}k^{\mathrm{b}}_1,1}$,

$$T^{\mathrm{w}}_i = t^{\mathrm{dep}}_{ij_{i,1}k^{\mathrm{b}}_1,1} - t. \tag{12}$$

Transfer time

If $L \geq 2$, the considered journey contains a transfer from line $l-1$ at stop k^{a}_{l-1} to line l at stop k^{b}_l, $l = 2, \ldots, L$. The alighting and boarding stops may be identical if both lines serve the same stop. Assuming that the transfer between the two stops takes a minimum amount δ_l of time, the calculation of transfer time is analogous to the calculation of waiting time at the origin stop. Given that the probe traveller is on vehicle $j_{i,l-1}$ on line $l-1$, the traveller arrives to stop k^{a}_{l-1} at time $t^{\mathrm{arr}}_{ij_{i,l-1}k^{\mathrm{a}}_{l-1},l-1}$. Each day the traveller boards the first non-full vehicle j_{il} departing from k^{b}_l after $t^{\mathrm{arr}}_{ij_{il}k^{\mathrm{b}}_l l} + \delta_l$. Transfer time is computed as the time difference between the arrival time of line $l-1$ and the departure time of line l,

$$T^{\mathrm{tr}}_{il} = t^{\mathrm{dep}}_{ij_{il}k^{\mathrm{b}}_l l} - t^{\mathrm{arr}}_{ij_{i,l-1}k^{\mathrm{a}}_{l-1},l-1}, \quad l = 2, \ldots, L, \tag{13}$$

and total transfer time on day i is computed by summation over all transfers,

$$T^{\mathrm{tr}}_i = \sum_{l=2}^{L} T^{\mathrm{tr}}_{il}. \tag{14}$$

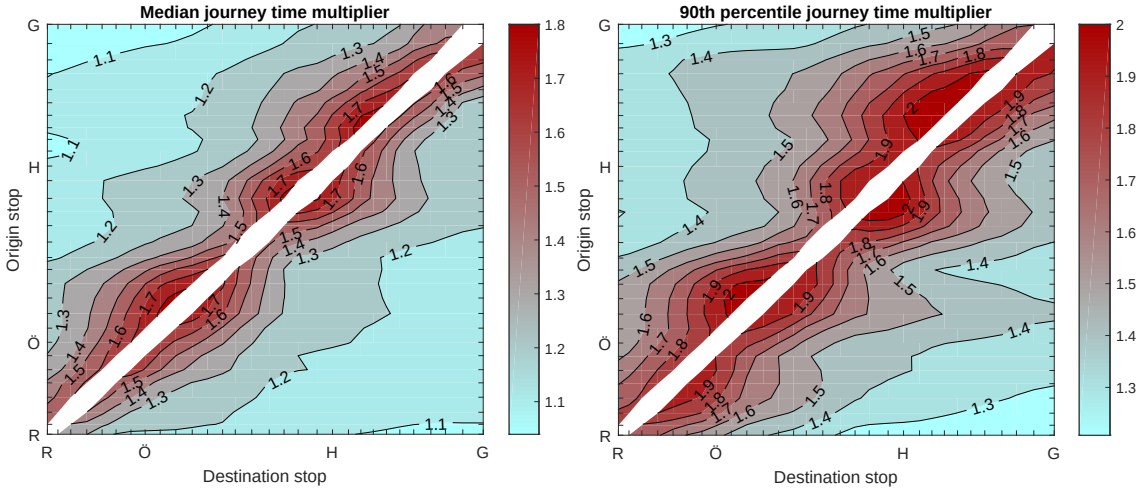

Figure 12: Journey time multiplier, median (left) and 90th percentile (right) across OD pairs. Journey start time 16:30, Monday–Friday 2016. Moving average with bandwidth 3 stops. Axes indicate stops Radiohuset (R, north end stop), Östra station (Ö), Hornstull (H), and Gullmarsplan (G, south end stop).

In-vehicle time

In-vehicle time between stops k and $k+1$ on line l and day i is computed from AVL data as the difference in vehicle departure times between the two stops,

$$T^{\mathrm{iv}}_{ikl} = t^{\mathrm{dep}}_{ij_{il},k+1,l} - t^{\mathrm{dep}}_{ij_{il}kl}, \quad k = k^{\mathrm{b}}_l, \ldots, k^{\mathrm{a}}_l - 2, \tag{15}$$

except for the final segment of the line where the in-vehicle time is the difference between the arrival time at the last stop k^a_l and the departure time from the preceding stop $k^a_l - 1$,

$$T^{\mathrm{iv}}_{i,k^{\mathrm{a}}_l-1,l} = t^{\mathrm{arr}}_{ij_{il}k^{\mathrm{a}}_l l} - t^{\mathrm{dep}}_{ij_{il},k^{\mathrm{a}}_l-1,l}. \tag{16}$$

The total journey in-vehicle time on day i is calculated by summation over all traversed inter-stop segments on all lines,

$$T^{\mathrm{iv}}_i = \sum_{l=1}^{L} \sum_{k=k^{\mathrm{b}}_l}^{k^{\mathrm{a}}_l-1} T^{\mathrm{iv}}_{ikl}. \tag{17}$$

In-vehicle crowding

The seated capacity of a vehicle, which may vary among lines and even among vehicles serving the same line, is denoted c^{sit}_{jl}. The load factor experienced by the probe traveller between stops k and $k+1$ on line l and day i is computed from APC data as the ratio between passenger load and vehicle capacity for the vehicle boarded by the probe traveller, $f_{ikl} = \ell_{ij_{il}kl}/c^{\mathrm{sit}}_{j_{il}l}$. The variable $\mathcal{I}_{iklm}$ indicates whether the experienced crowding conditions are at level m,

$$\mathcal{I}_{iklm} = \begin{cases} 1 & \phi_m \le f_{ikl} < \phi_{m+1}, \\ 0 & \text{otherwise}, \end{cases} \tag{18}$$

where the crowding thresholds ϕ_m are defined as in Section 2.2.

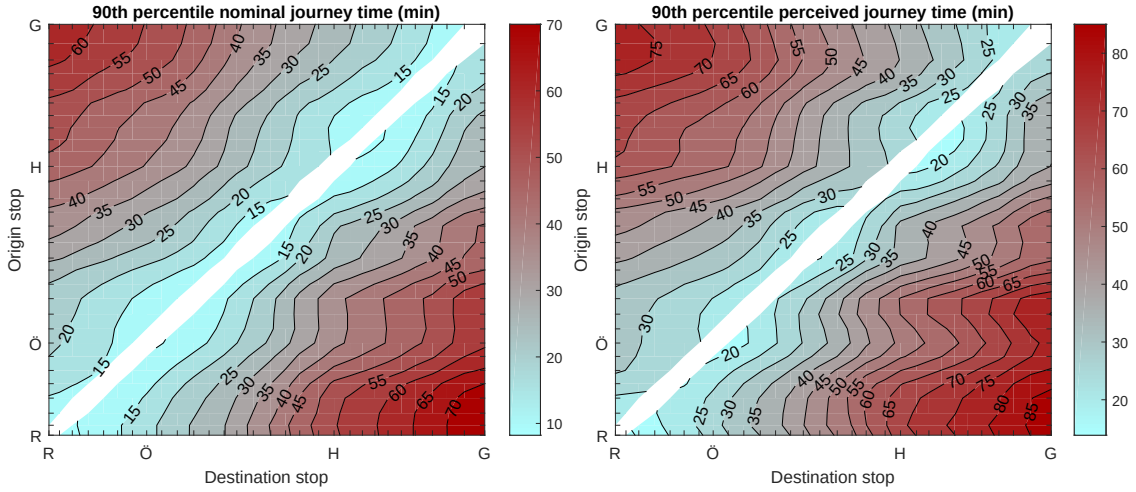

Figure 13: 90th percentile nominal (left) and perceived (right) journey time across OD pairs. Journey start time 16:30, Monday–Friday 2016. Moving average with bandwidth 3 stops. Axes indicate stops Radiohuset (R, north end stop), Östra station (Ö), Hornstull (H), and Gullmarsplan (G, south end stop).

Seat allocation

The number of passengers boarding the vehicle ahead of the probe traveller is drawn among all boarding passengers $b_{ij_{il}k_l^{\mathrm{b}}l}$ according to a Binomial distribution with probability parameter $p_{i,1}^{\mathrm{ahd}}$, which is equal to the ratio between the time elapsed since the preceding bus departure and the headway to the preceding bus, i.e., $p_{i,1}^{\mathrm{ahd}} = 1 - T_i^{\mathrm{w}}/h_{i,1}$ on the first journey segment and $p_{il}^{\mathrm{ahd}} = 1 - T_{il}^{\mathrm{tr}}/h_{il}$ on subsequent segments $l = 2, \ldots, L$. The headway is calculated as

$$h_{il} = t^{\mathrm{dep}}_{ij_{il}k_l^{\mathrm{b}}l} - t^{\mathrm{dep}}_{i,j_{il}-1,k_l^{\mathrm{b}}l}. \tag{19}$$

The probability that the traveller receives a seat is given by the cumulative distribution function of the binomial distribution,

$$\pi^{\mathrm{sit}}_{ik_l^{\mathrm{b}}l} = \begin{cases} 0 & q_{ij_{il}k_l^{\mathrm{b}}l} - b_{ij_{il}k_l^{\mathrm{b}}l} > s_{j_{il}l} \\ 1 & q_{ij_{il}k_l^{\mathrm{b}}l} \leq s_{j_{il}l} \\ \sum_{b=0}^{s_{j_{il}l} - q_{ij_{il}k_l^{\mathrm{b}}l} + b_{ij_{il}k_l^{\mathrm{b}}l}} \binom{b_{ij_{il}k_l^{\mathrm{b}}l}}{b} (p_{il}^{\mathrm{ahd}})^b (1 - p_{il}^{\mathrm{ahd}})^{b_{ij_{il}k_l^{\mathrm{b}}l} - b} & \text{else.} \end{cases} \tag{20}$$

Alighting passengers are randomly selected among all on-board passengers according to a hypergeometric distribution. The passengers getting a seat are randomly selected among all remaining standing passengers before new passengers board. Conditional on that the probe traveller is standing on segment $k-1$, the probability of getting a seat at stop k is

$$\pi^{\mathrm{sit}}_{ikl} = \begin{cases} 1 & q_{ij_{il},k-1,l} - a_{ij_{il}kl} \leq s_{j_{il}l} \\ \sum_{a=0}^{a_{ij_{il}kl}} \min\left(\frac{a}{\ell_{ij_{il},k-1,l} - a_{ij_{il}kl} - s_{j_{il}l} + a}, 1\right) \frac{\binom{s_{j_{il}l}}{a}\binom{q_{ij_{il},k-1,l} - s_{j_{il}l}}{a_{ij_{il}kl} - a}}{\binom{q_{ij_{il},k-1,l}}{a_{ij_{il}kl}}} & \text{else,} \end{cases} \quad k = k_l^{\mathrm{b}} + 1, \ldots, k_l^{\mathrm{a}} - 1. \tag{21}$$

The probability that the probe traveller is seated on line segment k is equal to the probability that the traveller gets a seat at stop k or at some preceding stop on the same line,

$$p^{\mathrm{sit}}_{ikl} = \sum_{k'=k^{\mathrm{b}}_l}^{k} \pi^{\mathrm{sit}}_{ik'l} \prod_{k''=k^{\mathrm{b}}_l}^{k'-1} (1 - \pi^{\mathrm{sit}}_{ik''l}). \tag{22}$$

Perceived journey time

The expected total journey time seated and standing, respectively, at crowding level m on day i is computed based on (17), (18) and (9) by summation over all lines and line segments,

$$\bar{T}^{\mathrm{sit}}_{im} = \sum_{l=1}^{L} \sum_{k=k^{\mathrm{b}}_l}^{k^{\mathrm{a}}_l-1} p^{\mathrm{sit}}_{ikl} \mathcal{I}_{iklm} T^{\mathrm{iv}}_{ikl}, \quad \bar{T}^{\mathrm{std}}_{im} = \sum_{l=1}^{L} \sum_{k=k^{\mathrm{b}}_l}^{k^{\mathrm{a}}_l-1} (1 - p^{\mathrm{sit}}_{ikl}) \mathcal{I}_{iklm} T^{\mathrm{iv}}_{ikl}. \tag{23}$$

The total perceived journey time is obtained as the sum of waiting time (12), transfer time (14), and in-vehicle time (23), each weighted by the corresponding time multipliers (Section 2.2),

$$T^{\mathrm{p}}_i = \beta^{\mathrm{w}} T^{\mathrm{w}}_i + \beta^{\mathrm{tr}} T^{\mathrm{tr}}_i + \sum_{m=1}^{M} \left(\beta^{\mathrm{sit}}_m \bar{T}^{\mathrm{sit}}_{im} + \beta^{\mathrm{std}}_m \bar{T}^{\mathrm{std}}_{im} \right). \tag{24}$$

The journey time multiplier for each day i is the average time multiplier over the duration of the journey,

$$B_i = \frac{\beta^{\mathrm{w}} T^{\mathrm{w}}_i + \beta^{\mathrm{tr}} T^{\mathrm{tr}}_i + \sum_{m=1}^{M} \left(\beta^{\mathrm{sit}}_m \bar{T}^{\mathrm{sit}}_{im} + \beta^{\mathrm{std}}_m \bar{T}^{\mathrm{std}}_{im} \right)}{T^{\mathrm{w}}_i + T^{\mathrm{tr}}_i + \sum_{m=1}^{M} \left(\bar{T}^{\mathrm{sit}}_{im} + \bar{T}^{\mathrm{std}}_{im} \right)}. \tag{25}$$

The experienced service reliability gap metric is computed from the percentiles of the empirical distribution of perceived journey time (24) across all considered days $i = 1, \ldots, I$. The reliability buffer time metric is computed in an analogous way from the distribution of nominal journey time.

References

Abkowitz, M., 1978. Transit service reliability. Tech. Rep. UMTA-MA-06-0049-78-1 Final Report., Urban Mass Transportation Administration.

Abrantes, P. A. L., Wardman, M. L., 2011. Meta-analysis of UK values of travel time: An update. Transportation Research Part A 45 (1), 1–17.

Andres, M., Nair, R., 2017. A predictive-control framework to address bus bunching. Transportation Research Part B: Methodological 104, 123–148.

Bates, J., Polak, J., Jones, P., Cook, A., 2001. The valuation of reliability for personal travel. Transportation Research Part E 37 (2–3), 191–229.

Beirão, G., Sarsfield Gabral, J. A., 2007. Understanding attitudes towards public transport and private car: A qualitative study. Transport Policy 14, 478–489.

Benezech, V., Coulombel, N., 2013. The value of service reliability. Transportation Research Part B 58, 1–15.

Björklund, G., Swärdh, J.-E., 2017. Estimating policy values for in-vehicle comfort and crowding reduction in local public transport. Transportation Research Part A: Policy and Practice 106, 453–472.

Cantwell, M., Caulfield, B., O'Mahony, M., 2009. Examining the factors that impact public transport commuting satisfaction. Journal of Public Transportation 12 (2), 1–21.

Casello, J. M., Nour, A., Hellinga, B., 2009. Quantifying impacts of transit reliability on user costs. Transportation Research Record 2112, 136–141.

Ceder, A., 2016. Public Transit Planning and Operation: Modeling, Practice and Behavior, 2nd Edition. Boca Raton: CRC Press.

Chan, J., 2007. Rail transit od matrix estimation and journey time reliability metrics using automated fare data. Master's thesis, Massachusetts Institute of Technology.

Daganzo, C. F., 2009. A headway-based approach to eliminate bus bunching: Systematic analysis and comparisons. Transportation Research Part B 43, 913–921.

dell'Olio, L., Ibeas, A., Cecin, P., 2011. The quality of service desired by public transport users. Transport Policy 18, 217–227.

Diab, E. I., Badami, M. G., El-Geneidy, A. M., 2015. Bus transit service reliability and improvement strategies: Integrating the perspectives of passengers and transit agencies in north america. Transport Reviews 35 (3), 292–328.

Ehrlich, J. E., 2010. Applications of automatic vehicle location systems towards improving service reliability and operations planning in london. Master's thesis, Massachusetts Institute of Technology.

Friman, M., Edvardsson, B., Gärling, T., 2001. Frequency of negative critical incidents and satisfaction with public transport services. I. Journal of Retailing and Consumer Services 8, 95–104.

Furth, P. G., Muller, T. H. J., 2006. Service reliability and hidden waiting time: Insights from automatic vehicle location data. Transportation Research Record 1955, 79–87.

Hamdouch, Y., Ho, H., Sumalee, A., Wang, G., 2011. Schedule-based transit assignment model with vehicle capacity and seat availability. Transportation Research Part B: Methodological 45 (10), 1805–1830.

Haywood, L., Koning, M., Monchambert, G., 2017. Crowding in public transport: Who cares and why? Transportation Research Part A 100, 215–227.

Hörcher, D., Graham, D. J., Anderson, R. J., 2017. Crowding cost estimation with large scale smart card and vehicle location data. Transportation Research Part B 95, 105–125.

Hounsell, N. B., Shrestha, B. P., 2012. A new approach for co-operative bus priority at traffic signals. IEEE Transactions on Intelligent Transportation Systems 13 (1), 6–14.

Kim, K. M., Hong, S.-P., Ko, S.-J., Kim, D., 2015. Does crowding affect the path choice of metro passengers? Transportation Research Part A 77, 292–304.

Leffler, D., Cats, O., Jenelius, E., Burghout, W., 2017. Real-time short-turning in high frequency bus services based on passenger cost. In: Models and Technologies for Intelligent Transportation Systems (MT-ITS), 2017 5th IEEE International Conference on. pp. 861–866.

Leurent, F., Benezech, V., Combes, F., 2012. A stochastic model of passenger generalized time along a transit line. Procedia - Social and Behavioral Sciences 54, 785–797.

Li, Z., Hensher, D. A., 2011. Crowding and public transport: a review of willingness to pay evidence and its relevance in project appraisal. Transport Policy 18, 880–887.

Li, Z., Hensher, D. A., 2013. Behavioural implications of preferences, risk attitudes and beliefs in modelling risky travel choice with travel time variability. Transportation 40, 505–523.

Rahmani, M., Jenelius, E., Koutsopoulos, H. N., 2015. Non-parametric estimation of route travel time distributions from low-frequency floating car data. Transportation Research Part C 58B, 343–362.

Raveau, S., Guo, Z., Munõz, J. C., Wilson, N. H. M., 2014. A behavioural comparison or route choice on metro networks: time, transfers, crowding, topology and socio-demographics. Transportation Research Part A 66, 185–195.

Schmöcker, J.-D., Fonzone, A., Shimamoto, H., Kurauchi, F., Bell, M. G. H., 2009. Frequency-based transit assignment considering seat capacities. Transportation Research Part B: Methodological 45, 392–408.

Schmöcker, J.-D., Sun, W., Fonzone, A., Liu, R., 2016. Bus bunching along a corridor served by two lines. Transportation Research Part B: Methodological 93, 300–317.

Sumalee, A., Tan, Z., Lam, W. H., 2009. Dynamic stochastic transit assignment with explicit seat allocation model. Transportation Research Part B: Methodological 2009 (8–9), 895–912.

Tirachini, A., Hurtubia, R., Dekker, T., Daziano, R. A., 2017. Estimation of crowding discomfort in public transport: Results from Santiago de Chile. Transportation Research Part A 103, 311–326.

Trafikförvaltningen, 2016. Riktlinjer Planering av kollektivtrafiken i Stockholms län. Stockholms Läns Landsting, sL-S-419761, in Swedish.

Uniman, D. L., Attanucci, J., Mishalani, R. G., Wilson, N. H. M., 2010. Service reliability measurement using automated fare card data: Application to the London underground. Transportation Research Record 2143, 92–99.

van Oort, N., 2014. Incorporating service reliability in public transport design and performance requirements: International survey results and recommendations. Research in Transportation Economics 48, 92–100.

Wardman, M., 2004. Public transport values of time. Transport Policy 11 (4), 363–377.

Wardman, M., Whelan, G., 2011. Twenty years of rail crowding valuation studies: Evidence and lessons from British experience. Transport Reviews 31 (3), 379–398.

Wood, D. A., 2015. A framework for measuring passenger-experienced transit reliability using automated data. Master's thesis, Massachusetts Institute of Technology.

Wu, W., Liu, R., Jin, W., 2017. Modelling bus bunching and holding control with vehicle overtaking and distributed passenger boarding behaviour. Transportation Research Part B 104, 175–197.

Zhang, Y., Jenelius, E., Kottenhoff, K., 2017. Impact of real-time crowding information: A Stockholm metro case study. Public Transport 9 (3), 483–499.

Zhu, Y., Koutsopoulos, H. N., Wilson, N. H., 2017. A probabilistic passenger-to-train assignment model based on automated data. Transportation Research Part B: Methodological 104, 522–542.