



## Passenger arrival and waiting time distributions dependent on train service frequency and station characteristics: A smart card data analysis

Ingvardson, Jesper Bláfoss; Nielsen, Otto Anker; Raveau, Sebastián; Nielsen, Bo Friis

*Published in:*  
Transportation Research. Part C: Emerging Technologies

*Link to article, DOI:*  
[10.1016/j.trc.2018.03.006](https://doi.org/10.1016/j.trc.2018.03.006)

*Publication date:*  
2018

*Document Version*  
Peer reviewed version

[Link back to DTU Orbit](#)

*Citation (APA):*  
Ingvardson, J. B., Nielsen, O. A., Raveau, S., & Nielsen, B. F. (2018). Passenger arrival and waiting time distributions dependent on train service frequency and station characteristics: A smart card data analysis. *Transportation Research. Part C: Emerging Technologies*, 90, 292-306. <https://doi.org/10.1016/j.trc.2018.03.006>

---

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

**Passenger arrival and waiting time distributions dependent on train service frequency and station characteristics: A smart card data analysis**

## **Abstract**

Waiting time at public transport stops is perceived by passengers to be more onerous than actual in-vehicle time, hence it largely influences the attractiveness and use of public transport. Transport models assume traditionally that average waiting times are half the service headway by assuming random passenger arrivals. However, research agree that two distinct passenger behaviour types exist: one group arrives randomly, whereas another group actively tries to minimise their waiting time by arriving in a timely manner at the scheduled departure time. This study proposes a general framework for estimating passenger waiting times which incorporates the arrival patterns of these two groups explicitly, namely by using a mixture distribution consisting of a uniform and a beta distribution. The framework is empirically validated using a large-scale automatic fare collection system from the Greater Copenhagen Area covering metro, suburban, and regional train stations thereby spanning service frequencies from 2-60 minutes. It was shown that the proposed mixture distribution is superior to other distributions proposed in the literature. This can improve waiting time estimations in public transport models. The results show that even at 5-minute headways 43% of passengers arrive in a timely manner to stations when timetables are available. The results bear important policy implications in terms of providing actual timetables, even at high service frequencies, in order for passengers to be able to minimise their waiting times.

## **Keywords**

Automated fare collection data; Smart card; waiting time; public transport; frequency-based timetables; mixture distributions.

## **Highlights**

- Passenger waiting times are modelled as a uniform and beta mixture distribution
- Validated framework that can improve waiting time estimations in transport models
- Results show that many passengers arrive in a timely manner to stations even at short headways
- Results highlight the importance of published timetables

# 1 Introduction

Waiting time at public transport stops is perceived by passengers to be more onerous than in-vehicle time (Nielsen, 2000; Fan et al., 2016; Fosgerau et al., 2007). Reducing waiting time is therefore of great importance when designing public transport systems. For en-route transfers this can be achieved by optimising public transport timetables in order to ensure short transfers (Parbo et al., 2014). However, this only affects transferring passengers which in the Greater Copenhagen Area corresponds to 45% (Christiansen, 2015). For all passengers it is also important to consider the waiting time experienced pre-route at the departure stop so that passengers can actively reduce their waiting time.

In general, when timetables are available to passengers, two distinct types of travel behaviour are observed when arriving to a departure stop: i) one group arrives randomly, and ii) a second group will try to minimise the waiting time by arriving in a timely manner at the scheduled departure time (Csikos and Currie, 2008; Frumin and Zhao, 2012; Jolliffe and Hutchinson, 1975; Luethi et al., 2007). The shares of the two groups are influenced by service characteristics such as headway and reliability as well as other factors, such as time of day (Csikos and Currie, 2008; Frumin and Zhao, 2012; Luethi et al., 2007; Nygaard and Tørset, 2016). As headway or reliability decreases, the share of passengers arriving randomly increases as the potential benefit of reduced waiting time is reduced (Bowman and Turnquist, 1981). Capturing such behaviour accurately in transport models is important for estimating impacts of public transport investments. However, most traditional public transport assignment models assume all passengers to arrive randomly to the stop, hence assuming the average waiting time to be half the headway (Fu et al., 2012; Nielsen, 2000; Nökel and Wekeck, 2009; Schmöcker et al., 2011; Szeto et al., 2013, 2011). Therefore such models might overestimate waiting times as more passengers will time their arrival at the station leading to lower actual waiting times (Csikos and Currie, 2008).

This paper contributes to the existing literature by analysing a large-scale empirical data set of passenger arrivals and waiting times at train stations in the Greater Copenhagen Area. The contribution is three-fold.

Firstly, this study proposes a general methodology to model passenger waiting times at public transport stations by explicitly taking into account passengers arriving randomly and non-randomly. The methodology is an extension of the approach proposed in Luethi et al. (2007) where the arrival patterns of passengers were modelled as a mixture of a uniform and a Johnson SB distribution, thus taking into account random and non-random passenger arrivals, respectively. In this present study the method is further developed by proposing a general mixture of a uniform and a beta distribution which can be fitted to specific service frequencies by adjusting the share of uniform passenger arrivals as well as the parameters for the beta distribution. The simple and general formulation makes it easily adoptable in public assignment models.

Secondly, data used for this study is based on a large-scale automated fare collection (AFC) system containing more than 1.5 million trips covering all modes of public transport in the Greater

Copenhagen Area during September and October 2014. This allows for extending the work of Frumin and Zhao (2012), which deployed a large-scale AFC dataset from the London Overground to analyse passenger arrivals at boarding stations at headways in the range 7-30 minutes. This present study covers stations with headways ranging from 2 minutes on the metro to 60 minutes on regional train lines. In addition, it includes modes with traditional published timetables (suburban and regional trains) and frequency-based timetables where actual departure times at stations are not published (metro). This makes it possible to compare results across a wide range of service frequencies and timetable types.

Thirdly, this study takes into account the effects of multiple station characteristics on the arrival patterns and waiting times of passengers. This data is joined onto databases including information on station layouts, station amenities and land use types surrounding the stations. This makes it possible to estimate the importance of such characteristics on passenger waiting times.

The paper is organised with a review of the existing literature on analysing passenger waiting times in public transport in section 2. The methodology and data used in this study is described in section 3 while the results are presented and discussed in section 4. In section 5 conclusions are drawn while policy implications are highlighted in section 6.

## **2 Literature review**

### **2.1 Estimation of waiting times**

Some of the earliest studies focusing on passenger arrival patterns analysed the relationship between waiting time and headway by a simple linear relationship. O’Flaherty and Mangan (1970) found that average passenger waiting time,  $W$ , could be related to average bus headway,  $h$  (measured in minutes) by the simple linear relationship  $W = 1.79 + 0.14h$  during an evening peak period in Leeds, UK. Seddon and Day (1974) improved the simple model by adding the influence of random bus arrivals. In their study the relationship was found to be  $W = 2.34 + 0.26h$  (measured in minutes) for stops in Manchester during both peak and off-peak hours. Hence, evidence of non-random arrivals were found as random arrivals would have implied  $W = 0.5h$ . Jolliffe and Hutchinson (1975) further improved the estimation of passenger waiting times by analysing the influence of day-to-day variability of bus arrivals. The study also proposed a three-fold categorisation of passengers based on their behaviour: i) those who arrive to minimise their waiting time, ii) those who arrive randomly, and iii) those whose arrival coincides with the bus, i.e. by running to catch it. Based on an analysis of ten bus stops with varying headways at 6-31 minutes in London they found that actual passenger waiting times were 30% less than if passengers arrived randomly. Bowman and Turnquist (1981) extended the estimation of passenger waiting times by modelling explicitly the arrival distribution of passengers timing their arrival at stops by using a decision model of arrival time choice to the bus stop. By using this model the study found that passengers are more sensitive to reliability than to scheduled headway when deciding between arriving randomly or timing their arrival.

Later studies investigated aggregate passenger arrival patterns with the purpose of estimating the threshold for when passengers arrive randomly. Fan and Machemehl (2002) found that 10 minute headway was the limit between random and non-random passenger arrivals based on 2,491 observations of bus passengers from Austin, Texas. In a later study, the same authors found that 11 minutes was the transition point between practically random arrivals to less random arrivals, and that all passengers timed their arrival at headways of over 38 minutes (Fan and Machemehl, 2009). Similar results were found in a study on bus passenger arrivals in Trondheim, Norway, where passengers were found to time their arrival at headways of 10 minutes (Nygaard and Tørset, 2016). These studies only analysed arrival patterns of bus passengers. Luethi et al. (2007) analysed passengers across public transport modes, i.e. bus, tram, and commuter rail based on data from Zürich, Switzerland. The study estimated the threshold for random arrivals even lower at 5 minutes for which a significant group of passengers timed their arrival. This study also found that the share of passengers arriving non-randomly was higher at 5-minute headways than at 6-minute headways, hence highlighting the importance of timetables that are easy to remember.

In a more recent study, Frumin and Zhao (2012) investigated passenger arrival patterns using a large-scale AFC dataset from the London Overground rail network finding that a notable share of passengers arrive in a timely manner at all analysed headways (7-30 minutes). The study distinguished between two types of passenger behaviour, namely timetable-dependent passengers arriving according to the timetable and passengers arriving randomly. Hence, instead of reporting a transition headway for when passengers arrive randomly, the study found that the share of passengers arriving randomly varied across the network and time-of-travel as more passengers arrived randomly at longer headways and at services with known reliability problems. The explicit two-fold arrival behaviour of passengers were subsequently used to evaluate service quality in terms of actual realised (and excess) journey times (Zhao et al., 2013), and to evaluate the effects of a timetable change on a rail line of the London Overground network (Frumin et al., 2013).

Several studies investigated the properties of the arrival patterns using statistical distributions. Luethi et al. (2007) proposed an advanced framework for modelling passenger arrivals using a mixture of uniform and Johnson SB distributions, hence explicitly taking into account the two-fold representation of the arrival behaviour of passengers. Several later studies have used a simpler approach. An analysis of transferring passengers were conducted in Beijing showing that the lognormal and gamma distributions had the best fit for direct and non-direct passengers, respectively (Guo et al., 2011). And, at Beijing bus stops Gong et al. (2016) found that gamma and lognormal distributions best fitted passenger waiting times during evening peak hours. Nygaard and Tørset (2016) analysed bus stops with 10-20 minute headways in Trondheim, Norway. They concluded that passenger arrivals were non-uniform without giving further details on alternative descriptions. While focusing explicitly on passenger arrival patterns, Frumin and Zhao (2012) and Zhao et al. (2013) did not focus explicitly on the statistical properties of the passenger arrival distributions.

The influence of travel and station characteristics on waiting times has also been the focus of recent research. Fan and Machemehl (2009) investigated the influence of stop location, gender and travel period, but did not find any effect on waiting times. However, the study found that passengers with car as access mode wait shorter than others. Currie and Csikos (2007) found time-of-day effects as random passenger arrivals were more common during off-peak hours as compared to travellers during peak hours. Similar results were found in Luethi et al. (2007) where travellers during morning and afternoon commute periods experienced shorter waiting times than those travelling mid-day. The study also found lower average waiting time for passengers at stations where the perceived service reliability was high, hence highlighting the importance of reliable operations. However, each station only had observations for one time period, hence the effect could be related to the station rather than the time period. The study suggested further research on the topic including the effects of time-of-day, route reliability, travel purpose, station location (in network), station environment, and previous activity (work, school, etc.). In a more recent study Fan et al. (2016) found significant effects of station amenities and perceived safety on the perceived waiting time using data from 36 light rail, commuter rail and bus stops in the Minneapolis-Saint Paul metropolitan area. If stations had limited amenities in terms of shelters and benches the waiting time was perceived 30% longer, and for women waiting in insecure places for more than 10 min the wait was perceived as much longer.

While these studies have investigated waiting times in detail across various study areas with different public transport structures including the influence of service frequency and station characteristics most studies rely on manual data collection making sample sizes for specific combinations of mode, time-of-day and service frequency relatively small. Only few studies deployed large-scale AFC data, but did not focus explicitly on the properties of the waiting time distributions. Furthermore, most studies relied on data covering only one mode of public transport, most often buses or single rail lines.

## **2.2 Implementation in transport models**

The first studies on public transport assignment models implemented passenger arrival patterns at departure stops simply by the so-called half-headway approach (Clerq, 1972; Dial, 1967). This approach relies on three important assumptions: i) deterministic transit vehicle headways hence assuming perfect service regularity, ii) passengers can board the first arriving vehicle, and iii) random arrival of passengers at stops (Fan and Machemehl, 2002). The first two assumptions are related to the operations of the public transport system which are not the main focus of this paper. A description of these assumptions can be found in Gentile et al. (2016). The third assumption is related to the travel behaviour of passengers which in this case are assumed to not consider the timetable of the public transport service. Hence, the average of the waiting time is estimated to be half the headway. The simplicity made the approach very popular resulting in wide usage within assignment models (Ceder and Marguier, 1985; Hess et al., 2004; Liu et al., 2010). This includes most frequency-based transit assignment models (Fu et al., 2012; Nökel and Wekeck, 2009; Schmöcker et al., 2011; Szeto et al., 2013, 2011).

However, as previous research states random passenger arrivals can only be reasonably assumed at short headways. Hence, such models might overestimate average waiting times as more passengers will time their arrival at the station (Csikos and Currie, 2008). Therefore, other transit assignment models assume half headway for short frequencies with a maximum waiting time for longer headways, e.g. the schedule-based model suggested in Nielsen (2000), thus resolving the issue slightly. In other schedule-based models the waiting time is modelled implicitly as part of the departure time choice of passengers based on the timetable of the public transport system (Nielsen, 2004; Nielsen & Frederiksen, 2006; Gentile et al., 2016; Nuzzolo et al., 2001). However, actual passenger arrivals at stops, and hence actual passenger waiting times, are not affected by the timetable due to high service frequency (Nuzzolo et al., 2015, 2012, 2001). By this, both model types do not accurately take into account the actual passenger arrival patterns and waiting time distributions, thus highlighting the need for a simple approach to estimate passenger waiting time patterns.

### 3 Methodology

The framework developed in this study was to analyse passenger waiting time distributions using AFC systems. Such data are being used more and more for various public transport planning purposes (Pelletier et al., 2011). Specifically, this analysis was tailored to investigate explicitly the ratio between passengers arriving randomly and those arriving non-randomly. This can be modelled by a mixture distribution holding two distributions which was first introduced in Luethi et al. (2007).

Passengers who do not consider the timetable are assumed to arrive according to a Poisson arrival process. Hence, they arrive randomly to stations and their waiting times follow the uniform distribution as also suggested in several studies (Fu et al., 2012; Luethi et al., 2007; Nielsen, 2000; Nökel and Wekeck, 2009; Schmöcker et al., 2011). The other group of passengers who do consult timetables will arrive with varying arrival intensity as this group of passengers will try to minimise their waiting time by arriving close to the departure time. The waiting time distribution will therefore not follow a uniform distribution. Instead this study proposes that the waiting time distribution of this group of passengers can be described by a beta distribution. This was done because the beta distribution possesses three important characteristics;

1. It is bounded on a defined interval which fits with the operational characteristics of public transport where passengers' waiting times are bounded by the service headway.
2. The beta distribution can handle the specific characteristics of passengers knowing the timetable. Namely, that they are assumed to add a buffer time to their arrival time at stations in order to not miss the departure (Fonzone et al., 2015; Frumin and Zhao, 2012). As the buffer time varies across passengers and service frequency, the arrival intensity is assumed to be highest a few minutes before the departure time and be decreasing towards the time of the next departure, cf. Figure 1. The shape parameters of the beta components ( $\alpha$  and  $\beta$ ) can explicitly model this as they define the skewness which might also vary across service frequency.

3. The uniform distribution is a special case of the beta distribution. By this there is a strong link between the distribution of the random passenger arrivals and those arriving according to the timetable.

Hence, with  $\Gamma(z)$  denoting the gamma function defined as  $\Gamma(z) = \int_0^{\infty} x^{z-1} e^{-x} dx$ , the probability density function (PDF) of the mixture distribution,  $f(x)$ , proposed by this study can be written as:

$$f(x) = \zeta \cdot \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha) \cdot \Gamma(\beta)} \cdot x^{\alpha-1} \cdot (1-x)^{\beta-1} + (1-\zeta) \quad (1)$$

The  $\zeta$  denotes the proportion of the beta component whereas  $(1-\zeta)$  denotes the proportion of the uniform component. Figure 1 shows the PDF with the y-axis denoting the density and the x-axis denoting the normalised waiting time (i.e. share of full headway).

[Place Figure 1 about here]

Maximum likelihood estimation was used to fit the waiting times calculated from the AFC data to the specified mixture distribution framework using the R statistical software package (R Core Team, 2017). This was done for all headways in the data (2/3/4/6/10 minutes for metro, 5/10/20 minutes for suburban trains, and 20/30/60 for regional and intercity trains). By this, a mixture distribution for each headway was estimated including the share of observations contained in the uniform and beta parts, respectively. Finally, the fit was compared to that of a mixture distribution containing one part uniform and one part Johnson  $S_B$  distribution as proposed by Luethi et al. (2007) to evaluate the methods.

### 3.1 Case study: Greater Copenhagen Area

The study area covers the commuting area<sup>1</sup> of Copenhagen in the eastern part of Denmark in which approximately 2 million people live. The study utilised the Danish *Rejsekort* (English: Travel card) which is a tap-in-tap-out AFC system. This holds information on when and where passengers check into and out from the public transport system as well as intermediate transfer locations. The system covers all modes of public transport in Denmark, i.e. buses, metro and railways, and is used for more than 100 million trips annually (Rejsekort A/S, 2017). However, not all passengers use it for all trips as monthly passes are not currently implemented in the card. Hence, most commuting trips are not included in the data. For this study a data sample covering 1,767,858 public transport trips for the months of September and October 2014 was utilised corresponding to 12% of all public transport trips during that period.

The AFC data was linked to timetable data for suburban trains (S-trains) and regional trains in order to calculate the waiting times for passengers. For metro trains no public timetable is available due to the system running at high service frequency, i.e. every four minutes during rush hour, every six minutes during other hours, and every 10-20 minutes during the night. As the two metro lines run partly in parallel service frequency is doubled within the city centre. Hence, for the metro a synthetic timetable

---

<sup>1</sup> This also includes towns located throughout Zealand from which extensive commuting patterns towards Copenhagen, e.g. Kalundborg, Korsør, and Vordingborg.

based on the time-of-day-dependent service frequency was produced in order to calculate waiting times for metro passengers.

The effects of various travel and station characteristics on waiting time distributions were analysed using different data sources, cf. Table 1. Station characteristics were taken from a study on transfer attributes (Dyrberg and Christensen, 2015). This included whether stations were located underground or had platform shelters to protect from weather, and a subjective measure of station layout ranking stations in terms of wayfinding. The inclusion of this measure was not straight-forward, but was included to analyse whether stations with simple layout and easy access to platforms from the street level influenced passenger arrival behaviour. In addition, detailed land use data comprising 44 categories of land use types were tested and eventually aggregated into only two types, i.e. leisure and commute. By this, land use types such as offices, industry and residential areas were categorised as being related to commuting travel whereas land uses such as nature, recreational areas, shopping and cultural institutions were categorised as leisure. For each station the land uses within 500 meters of the station were identified since the Danish travel survey suggest that the mode share of public transport is notably higher for trips originating and terminating within 500 meters of a station (Christiansen, 2015). The land use type with the largest share was assigned to that station. However, if the land uses surrounding a station were mixed, i.e. either no share larger than 30% or several shares larger than 30%, the land use was specified as mixed. Lastly, results from a passenger survey conducted during the years 2013-2015 was included to test the influence of perceived safety.

[Place Table 1 about here]

### **3.2 Data preparation**

An algorithm was developed to connect passenger arrivals to train departures on the origin station corresponding to the procedure proposed in Frumin and Zhao (2012), but with some differences. It was designed to accommodate all passengers, namely those travelling directly from origin station to destination station using only one mode of transport as well as passengers transferring midway. Each passenger was assigned on the first train that travelled between i) the station of check-in, and ii) the station of check-out (or transfer point). The waiting time was then calculated as the time from check-in to the scheduled departure time of the assigned train.

Simultaneously, the algorithm calculated the number of scheduled departures between the station-pairs within the next 180 minutes. For the analysis of waiting time distributions it was required that headways were constant, hence headways for the next three departures were required to be fixed, e.g. exactly 20 minutes. The procedure did require a number of assumptions;

*Firstly*, the waiting times were calculated based on the time of check-in at the check-in stands. By this, it is assumed that passengers tap in when they arrive to the station instead of waiting for the train to arrive before checking in. This was assumed to be reasonable as most often there is no financial incentive to wait for the train before checking in. Also, all stations had tap-in stands at the platforms

which made it possible for passengers to check in when they arrived. Most stations do not have tap-in stands at the waiting hall, but underground metro stations have stands at the intermediate level. Hence, for these stations there could be some walking time from tap-in to arrival at the platform. However, the intermediate level is within visible range of the platform, i.e. short walk on escalators, hence this assumption was assumed to be reasonable also for these stations. The assumption of check-in was addressed specifically by controlling the timestamp for the next tap-out at the destination station (or tap-in if transferring) to ensure that the passenger actually travelled on the assigned train. Finally, the validity of this first assumption was tested by use of manually collected passenger arrival data.

An important element to consider from this first assumption is the provision of real-time information to passengers before arriving at the check-in stands. In Copenhagen the widely used online travel planner is updated with real-time information on delays, hence for passengers planning their trips real-time information is easily available before and during the trip. This makes it possible for passengers to change plans in case of delays if multiple travel alternatives are present for their trip. But, if passengers try to arrive in a timely manner to the station that implies that passengers are performing their route choice prior to arriving at the station. As some stations in the network have multiple lines the route choice might be affected when arriving to the station, both in case of delays or even just based on the actual arrival time at the station. Specifically, passengers might choose other lines or routes based on real-time information, either from their smart-phone or from information shown at the station. As most stations in the Copenhagen area have information shown to passengers physically close to platforms the influence on actual waiting time is probably small. However, it might affect the route choice on stations where passengers have multiple options, e.g. stations served by multiple lines. Hence, a separate analysis was performed to test this assumption specifically (refer to section 4.2).

*Secondly*, buses were not included because check-in happens inside the vehicles making it impossible to estimate the arrival time of the passengers at the bus stop.

*Thirdly*, only the first trip leg<sup>2</sup> of each trip was included in the analysis. This was done to ensure that waiting times were not influenced by possible prior public transport modes as it is only the first trip leg that is affected by the access mode, e.g. walk or bicycle. By this, it is ensured that arrival times were not influenced by transferring passengers arriving simultaneously in larger groups.

*Fourthly*, the main analyses only included trips performed on weekdays. However, a separate analysis was performed to test possible differences between weekdays and weekends as well as possible time-of-day effects.

*Lastly*, as the metro does not have an official published timetable this study adopted a pseudo-timetable based on the local service frequency which ranged from 2-6 minutes during day hours and 7-20 minutes during night hours. This assumption was important for calculating passenger waiting times

---

<sup>2</sup> A trip is defined based on having one single travel purpose. One trip might contain more trip legs, e.g. in public transport which includes access and egress and possibly several modes due to transfers.

and was assumed to be reasonable when taking into account that passengers are not able to look up a timetable for the metro trains as they can only know the service frequency at their preferred station. Hence, they are hypothesised to arrive randomly, and the analysis of whether this is true is not affected by the use of a pseudo-timetable.

By the end of these procedures the sample included 701,252 trip legs for weekdays and 308,091 for Saturday and Sundays.

### 3.3 Effects of realised timetables

The initial calculation of passenger waiting times based on the scheduled timetable revealed problems in the dataset that needed to be corrected, cf. Figure 2 (left). As expected the largest arrival rate of passengers happens just prior to a departure. This points towards a trend of passengers trying to minimise their waiting time, but still adding a buffer time in order to not miss the departure, which was also observed in Frumin and Zhao (2012). The arrival rate then decreases steadily until the next departure. However, in the first minutes after the prior departure the arrival rate increases as also observed for London Overground passengers (Frumin and Zhao, 2012). This is due to three reasons.

1. The timetable data for S-trains was based on arrival times rather than departure times. This meant that platform dwell times were not taken into account even though they range from 10-30 seconds depending on the station. Hence, S-train departure times were corrected by adding the official station-specific minimum dwell times to the timetable arrival times.
2. Passengers being on the verge of catching a train might run for it. This leads to a larger intensity of passengers arriving just before and after the scheduled departure time. For the passengers arriving just late according to the timetable they will not catch the train and have to wait for the next departure hence experiencing a waiting time of almost the headway. However, if the train is delayed these people might slow down when arriving at the station because they can see that the train is delayed. Hence, these passengers will check in after the scheduled departure time, but still catch the delayed train thereby showing up in the tail-end of the distribution.
3. There might be some passengers that check in just prior to entering the train which in the case with delayed trains will be just after the original departure time.

By taking actual delays into account it was possible to verify that many of the trip legs from the tail of the distribution were indeed passengers who actually were able to catch the previous delayed train. Official delay data from the Danish railway manager was used, but the procedure also required implementation of delay correction factors to accurately estimate the actual departure times from the station platforms (Richter et al., 2013). By removing these trip legs using the realised timetable of delayed trains the distribution looked more as expected, cf. Figure 2 (right). These trip legs could have been included with updated, realised waiting times. However, in order to not conflict the general assumption of passengers timing their arrival to the *scheduled* timetable they were removed. Hence, only scheduled departure times were used for the calculation of waiting times. As delay data was not available for all

data this procedure resulted in a smaller sample of 617,996 observations for weekdays and 277,958 for weekends.

[Place figure 2 about here]

### 3.4 Validation

The waiting times deducted from the AFC system were validated using manually collected data of passenger arrivals at Bernstorffsvej S-train station. A total of 182 observations were collected during a morning peak period on August 11, 2016 from 7:30 to 8:45. During the data collection efforts were made to not include passengers arriving to the station by bus. The manually collected data was compared to 440 observations from the AFC data, cf. Figure 3. Unfortunately, data were not available for the exact same period. Instead data from similar morning peak periods during September and October 2014 were used.

[Place Figure 3 about here]

The validation was performed using the Kolmogorov-Smirnov statistical test which showed that it cannot be rejected that the distributions are identical. This finding also validates the assumption that passengers do not generally wait until the time of departure before checking in using their Rejsekort. If that was the case the distributions of waiting times for the AFC data would be more positively skewed than that of the manual data which is not the case. Although not statistically significant, it can be seen that the largest deviation is concentrated around the middle, i.e. at half the headway. However, no obvious reasons for this were identified, and the deviation is still small when taking into account the relative small sample sizes used for the validation. Furthermore, when considering the full range there seems to be no systematic deviations.

## 4 Results

### 4.1 Main findings

The full dataset of 617,996 weekday trip legs were analysed to estimate the fit of the two mixture distributions, i.e. the mixture of uniform and beta proposed by this present study, and the mixture of uniform and Johnson proposed by Luethi et al. (2007). Table 2 shows the results of the fit for each headway of the schedule-based public transport services. The fit of each mixture distribution to the data was evaluated using the two-sample Kolmogorov-Smirnov test statistic,  $D$ , and the corresponding p-value, whereas the Akaike information criterion (AIC) was used to compare the fit of the two mixture distributions with a lower number signifying better fit (Akaike, 1974). The results are illustrated specifically for 20-minute headways in Figure 4.

[Place Table 2 about here]

[Place Figure 4 about here]

The results from the 20-minute headway illustrated in Figure 4 visually suggest a better fit from the beta-mixture distribution as compared to the Johnson-mixture distribution while the traditional simple uniform distribution has a much worse fit. When analysing the fits across all service headways this tendency is clear. The assumption of a Johnson-mixture distribution is rejected at the 99% confidence level at all service headways except for 5-minute headways. This can be compared to the fit of the beta-mixture which can only be rejected for 10-minute headways. The beta-mixture fit is visualised specifically in Figure 5 together with the remaining service headways.

[Place Figure 5 about here]

As the headway between train departures increases the share of random passenger arrivals decreases. Hence, at short five minute headway 57% of train passengers arrive randomly which decreases to 7% at 60 minute headway. At 10 minute headway the share is 48% suggesting that approximately half will try to time their arrival at stations. These results are consistent with previous studies which found that passengers will start timing their arrival if headways are larger than 5-10 minutes (Jolliffe and Hutchinson, 1975; Luethi et al., 2007; O’Flaherty and Mangan, 1970; Seddon and Day, 1974).

[Place Table 3 about here]

Table 3 shows the shares of the beta component ( $\zeta$ ) and the uniform component ( $1 - \zeta$ ) of the mixture distribution including the shape parameters of the beta component ( $\alpha$  and  $\beta$ ). As headway increases,  $\alpha$  decreased consistently from 0.41 at 5 minute headways to 0.14 at 60 minute headways. Similarly,  $\beta$  increased consistently from 2.85 at 5 minute headways to 11.2 at 60 minute headways, hence suggesting increasing positively skewness as headway increases.

[Place Figure 6 about here]

For metro passengers (i.e. frequency-based timetable) the estimated mixture distributions did not include a significant beta component for any of the analysed service headways. Instead, the best fit was a uniform waiting time distribution, hence suggesting random passenger arrivals, cf. the results overview in Figure 5. This suggests that passengers are unable to time their arrival at public transport services that use frequency-based timetables where actual departure times are not published. This can be seen specifically when comparing the waiting time distributions of metro passengers at 6 minute headways with those of S-train passengers at 5 minute headways, cf. Figure 6. Despite the similar headways the waiting time distribution of metro passengers is estimated as fully uniform, whereas that of S-train passengers is 57% uniform and 43% beta. The results are pronounced at all headways of the metro where there is no significant beta-component for a mixture distribution resulting in the uniform component being estimated at 100%. However, it should be noted that statistical tests for fully random arrivals, i.e. fully uniform waiting time distributions, were rejected at the 95% confidence level for all but 10-minute headways.

The results imply that average waiting times for passengers are higher when timetables are not published, i.e. in the example case 49% of headway for metro passengers as compared to 46% for S-train passengers, cf. Figure 6. In these cases the average waiting time is approximately half the headway.

However, as headway increases for scheduled public transport the average waiting time decreases to 35%, 26% and 16% of the headway for 20, 30 and 60 minute headway, respectively. In other words, average waiting times are reduced by 8% and 15% at short headways of 5- and 10-minutes, and by 30%, 47% and 67% at longer headways of 20-, 30- and 60-minutes, respectively. Thus, the findings generally highlight the importance of using actual timetables. The findings are larger in magnitude than those observed for passengers in the London Overground network where the actual average waiting time was 11.2 minutes at 30-minute headways as compared to 7.9 minutes for passengers in the Copenhagen rail network (Frumin and Zhao, 2012). Hence, this suggests that passengers in the Copenhagen network are more sensitive to waiting times, more aware of public transport timetables, and/or it could be related to a generally more reliable service.

## **4.2 Influence of route choice**

The stations in the public transport network vary in terms of number of lines served. As some stations are served by several lines going to the same destination, it is possible for passengers to adapt their route choice dependent on their arrival time at the station, whereas arrivals to specific lines might follow different patterns. For the data in Copenhagen two main cases were identified;

1. Trip legs from S-train stations served by multiple S-lines versus those served by one line
2. Trip legs from stations served by multiple train types, e.g. regional trains, suburban trains and metro

The first case compares the arrival pattern of passengers at the suburban train network, namely at stations which have only one line with that of stations with multiple lines serving the station. As headway needs to be similar in order to perform a consistent comparison this analysis only compares Lyngby station, which is served by both a slow and a fast train, with the stations on the S-train Ring line which only has one service type stopping at all stations. As can be seen in Figure 7 the arrival distributions were very similar, and statistical tests revealed that the hypothesis of similar arrival behaviour could not be rejected for these cases at 5-minute headway.

[Place Figure 7 about here]

The second case covers trip legs where the passenger can choose between multiple train types, namely between metro, regional, and S-trains. These cases cover services with widely separation between train and metro platforms, requiring long walking distances and use of escalators, with no visual connection between the platforms, and with no joint information systems on departures. In addition, the service frequency for metro services is very high, hence reducing the potential benefit of an adapted route choice. However, as these cases relates to multiple public transport modes with varying service headways at all operating hours the actual service headway experienced by the passengers is uneven. As constant headway is required for the analyses these cases could not be compared.

## **4.3 Influence of time of travel**

The effects of time of travel shown in Table 4 indicate that more passengers arrive randomly outside of peak hours than those traveling during peak hours, consistent with findings in other studies (Frumin and Zhao, 2012; Luethi et al., 2007)

[Place Table 4 about here]

This suggests that travellers during peak hours are frequent travellers who know the timetable better and are more eager to time their arrival to their preferred departure. This is in contrast to travellers outside peak hours and during weekends which seem to have less focus on minimising their waiting time. This difference in travel behaviour leads to increased waiting times for passengers outside peak hours as also seen in Figure 8 for 20-minute and 30-minute headways whereas the difference is minimal at 60-minute headways.

[Place Figure 8 about here]

#### **4.4 Influence of station characteristics**

The characteristics listed in Table 1 were analysed to investigate their influence on passenger waiting time distributions. Of the parameters tested notable effects were found for station layout in terms of ease of wayfinding whereas no effects were found for characteristics such as availability of various types of shelters, perceived safety level or being located underground or in close proximity to specific land use types.

[Place Table 5 about here]

The results showed in Table 5 suggests that waiting times are generally longer when station layouts are more complicated and confusing. Hence, at headways of 20, 30 and 60 minutes passengers arrive earlier at the station platform if stations are perceived as more confusing. This could suggest that passengers require a larger access time buffer for such stations in order not to risk missing their departure. As many of the stations with complicated layouts also have many passengers this result could also be related to congestion. At 5 and 10 minute headways no significant differences were found suggesting less incentive for buffer times at shorter headways.

#### **4.5 Discussion and study limitations**

While the methodology proposed by this study proved suitable across various service frequencies and station characteristics, still some assumptions could be addressed in future studies.

Firstly, the fundamental assumption of this study was that the actual waiting time of passengers can be calculated as the time from check-in to the time of the first departing train that also stopped on the destination (or transfer) station of the passenger.

This require i) that passengers check in immediately on arrival, ii) that there is a very short distance from check-in to boarding the public transport vehicle, and iii) that the passengers do not change plans after arriving at the station, e.g. changes to another route due to delays. Several measures were taken to address the implications of these assumptions. As the check in stands are located on

station platforms, the assumption of check-in on arrival is reasonable. This was also seen from the validation in section 3.4 which showed that there was not significantly more tap-ins immediately before departures, hence there was no evidence of this assumption being violated. Similarly, it was checked that the designated train did not arrive prior to the tap-in of the passenger, hence eliminating the risk of passengers being assigned to a train that they did not catch. Finally, the possible effects of changes to route choice decisions were evaluated in section 4.2. This analysis was also performed to test the influence of provision of (real-time) information at the station. This might influence the behaviour of passengers before arriving to the station, but also after arriving where passengers might choose another travel alternative than originally planned due to delays. The results shown in section 4.2 revealed no significant differences between waiting time distributions for stations served by suburban skip-stop services and normal suburban train services.

In continuation of this, it should be emphasised that the method used scheduled train departure times to calculate passenger waiting times. This was chosen deliberately as passengers were hypothesised to use the scheduled timetables if planning their arrival at the station. However, results from Frumin and Zhao (2012) suggested that passengers arrive more randomly on lines with lower reliability. The results from this present study showed a generally larger percentage of timetable-dependent passengers, which could be related to the relatively high punctuality of especially suburban trains (93.7%) and regional trains (82.6%) in the Copenhagen area 82.6% (DSB, 2018).<sup>3</sup> However, future studies could investigate further the impacts of reliability on passenger arrival patterns.

Secondly, the present study did not include bus trips. This was not possible as the arrival time of bus passengers at the stop cannot be determined due to the check-in hardware being located inside the bus. However, recent research efforts have enabled the use of new technology such as capturing waiting passengers using Bluetooth or Wi-Fi detection from smartphones (Shlayan et al., 2016).

Thirdly, this study adopted a pseudo-timetable for metro trains as no published timetable is available for passengers. This included the correct service frequency, but no accurate departure times. However, as the results showed that passenger arrivals were almost entirely random independent of service frequency, using an actual planned timetable would similarly result in almost entirely uniform waiting time distributions. Hence, the general conclusions would not change notably.

Fourthly, the approach in this paper suggests that the waiting time distributions at each station is a mixture of a uniform and a beta distribution. Methodologically, it can be argued that the full distribution of each headway is a weighted sum of mixture distributions due to aggregating waiting times over multiple stations. This entails that a mixture distribution should be estimated for *each* station. However, that would require extensive data requirements making it infeasible for practical use. In this study we therefore validated the *aggregate* distribution making the approach easily adaptable for use in transport models.

---

<sup>3</sup> A passenger is considered punctual if arriving at the destination within 3 minutes of the scheduled arrival time.

Finally, a path for future research on the topic could be how other travel characteristics influence waiting time distributions and the share percentages of passengers arriving randomly or in a timely manner. Specifically, this could include the influence of weather or seasons which have been highlighted to have an effect on ridership (Arana et al., 2014; Zhou et al., 2017). In Copenhagen this might have an effect as the weather conditions changes notably during the year, and even during the same day. It could therefore be an interesting extension to analyse whether passenger arrival behaviour at stations is affected by local weather conditions. Unfortunately it was exorbitant expensive to acquire local weather data from the Danish Meteorological Agency to carry out this analysis.

## **5 Conclusions**

The paper presents an approach to model passenger waiting time as a mixture distribution to explicitly take into account the arrival behaviour of public transport passengers at stations. Specifically, arrival patterns of passengers can be grouped into two groups: i) one group arriving randomly, e.g. due to not knowing the timetable, and ii) another group arriving according to a beta distribution, e.g. due to timing their arrival at the station in order to minimise the waiting time at the station. The approach was tested using a large-scale Automatic Fare Collection (AFC) system from the Greater Copenhagen Area revealing better fit than prior arrival models proposed in the literature. The measurements were also validated against manually collected data to ensure that the passenger arrival times matched the check-in time, which was a fundamental assumption used in the calculation of waiting times. The data source used ensured a large sample covering trip legs in metros, suburban, regional and intercity trains across headways of 2-60 minutes.

The results showed that the share of passengers arriving randomly decreases as the headway increases. Even at short headways a large share of passengers are trying to minimise their waiting time by timing their arrival time at stations. At 5- and 10-minute headways this share was estimated at 43% and 52%, respectively, increasing with higher service frequencies. These results were only evident for train stations with public schedule-based timetables whereas arrival patterns at metro stations with frequency-based timetables were random independent of headway. This finding points towards the importance of publishing real timetables to passengers even at short headways typically seen in high-frequency metro services. However, if severe regularity issues are observed published timetables will be less important.

In addition, the study revealed more timed arrival patterns at rush hours as compared to other time of day and weekends. The results of station characteristics showed that stations with confusing station layout led to increased passenger waiting times. This was probably due to the need for passengers to add an extra time buffer to their access time when arriving to the station. Hence, it is of great importance to ensure easy accessibility to stations in order to reduce passenger waiting times and thus to ensure attractive public transport systems.

## **6 Implications**

The study has two important implications for policy and practice. Firstly, the findings highlight the importance of providing passengers real timetables whenever possible. This makes it possible for passengers to be able to time their arrival at the station in order to minimise their waiting time. Even at high service frequency this study found empirical evidence that many passengers actively time their arrival to the train departure. By not providing exact timetables passengers are forced to arrive randomly, thus prolonging their waiting time. Taking into account that passengers value waiting time higher than other time components further highlights the importance.

Secondly, the proposed framework of modelling passenger waiting times points towards new alternative ways of incorporating passenger behaviour in transport assignment models. Instead of assuming fixed hidden waiting times based on scheduled service frequency transport models can be improved by explicitly model the actual passenger arrival patterns at stations. The present study found that a mixture distribution consisting of one part uniform and one part beta distribution fitted actual arrival patterns well, and that the actual distribution parameters depend on service frequency.

## **Acknowledgements**

The authors would like to thank the Danish Metro Company (Metroselskabet I/S) for financing the Ph.D.-study making this work possible, and the IPTOP (Integrated Public Transport Optimisation and Planning) research project for help financing this study and negotiating access to the AFC and realised timetable data. Furthermore, the authors would like to thank research assistant Mads Paulsen for help with coding and structuring the AFC and timetable data, research assistant Morten Eltved for providing observed passenger arrival times at Bernstorffsvej station used for the validation, and PhD-student Fabrizio Cerreto for providing correction factors for the train data.

## **References**

- Akaike, H., 1974. A New Look at the Statistical Model Identification. *IEEE Trans. Automat. Contr.* 19, 716–723.
- Arana, P., Cabezudo, S., Peñalba, M., 2014. Influence of weather conditions on transit ridership: A statistical study using data from Smartcards. *Transp. Res. Part A Policy Pract.* 59, 1–12. <https://doi.org/10.1016/j.tra.2013.10.019>
- Bowman, L.A., Turnquist, M.A., 1981. Service frequency, schedule reliability and passenger wait times at transit stops. *Transp. Res. Part A Gen.* 15, 465–471. [https://doi.org/10.1016/0191-2607\(81\)90114-X](https://doi.org/10.1016/0191-2607(81)90114-X)
- Ceder, A., Marguier, P.H.J., 1985. Passenger waiting time at transit stops. *Traffic Eng. Control* 26, 327–329.
- Christiansen, H., 2015. Documentation of the Danish National Travel Survey. Kgs. Lyngby, Denmark.
- Clerq, F.P., 1972. A Public Transport Assignment Method. *Verkeerstechniek*, Netherlands 23.
- Csikos, D., Currie, G., 2008. Investigating Consistency in Transit Passenger Arrivals. *Transp. Res. Rec.*

- 2042, 12–19. <https://doi.org/10.3141/2042-02>
- Currie, G., Csikos, D.R., 2007. The Impacts of Transit Reliability on Wait Time: Insights from Automated Fare Collection System Data, in: Transportation Research Board 86th Annual Meeting. p. No. 07-0544.
- Dial, R.B., 1967. Transit Pathfinder Algorithm. *Highw. Res. Rec.*
- DSB, 2018. Punctuality for Long-distance and Regional trains and S-trains (in Danish) [WWW Document]. URL <https://www.dsb.dk/om-dsb/virksomheden/rettidighed/> (accessed 1.11.18).
- Dyrberg, M.B., Christensen, C.B., 2015. Transfers in public transport route choice models. Technical University of Denmark.
- Fan, W. (David), Machemehl, R.B., 2009. Do Transit Users Just Wait for Buses or Wait with Strategies? *Transp. Res. Rec. J. Transp. Res. Board* 2111, 169–176. <https://doi.org/10.3141/2111-19>
- Fan, W., Machemehl, R.B., 2002. Characterizing Bus Transit Passenger Waiting Times. 2nd Mater. Spec. Conf. Can. Soc. Civ. Eng. 7, 1–10.
- Fan, Y., Guthrie, A., Levinson, D., 2016. Waiting time perceptions at transit stops and stations: Effects of basic amenities, gender, and security. *Transp. Res. Part A Policy Pract.* 88, 251–264. <https://doi.org/10.1016/j.tra.2016.04.012>
- Fonzone, A., Schmöcker, J.D., Liu, R., 2015. A Model of Bus Bunching under Reliability-based Passenger Arrival Patterns. *Transp. Res. Part C Emerg. Technol.* 7, 276–299. <https://doi.org/10.1016/j.trpro.2015.06.015>
- Fosgerau, M., Hjort, K., Lyk-Jensen, S.V., 2007. The Danish Value of Time Study. Kgs. Lyngby, Denmark.
- Frumin, M., Zhao, J., 2012. Analyzing Passenger Incidence Behavior in Heterogeneous Transit Services Using Smartcard Data and Schedule-Based Assignment. *Transp. Res. Rec. J. Transp. Res. Board* 2274, 52–60. <https://doi.org/10.3141/2274-05>
- Frumin, M., Zhao, J., Wilson, N., Zhao, Z., 2013. Automatic Data for Applied Railway Management. *Transp. Res. Rec. J. Transp. Res. Board* 2353, 47–56. <https://doi.org/10.3141/2353-05>
- Fu, Q., Liu, R., Hess, S., 2012. A review on transit assignment modelling approaches to congested networks: a new perspective. *Procedia -Social Behav. Sci.* 54, 1145–1155. <https://doi.org/10.1016/j.sbspro.2012.09.829>
- Gentile, G., Florian, M., Hamdouch, Y., Cats, O., Nuzzolo, A., 2016. The Theory of Transit Assignment: Basic Modelling Frameworks, in: Gentile, G., Noekel, K. (Eds.), *Modelling Public Transport Passenger Flows in the Era of Intelligent Transport Systems*. Springer, pp. 287–386. [https://doi.org/10.1007/978-3-319-25082-3\\_6](https://doi.org/10.1007/978-3-319-25082-3_6)
- Gong, H., Chen, X., Yu, L., Wu, L., 2016. An application-oriented model of passenger waiting time based on bus departure time intervals. *Transp. Plan. Technol.* 39, 424–437. <https://doi.org/10.1080/03081060.2016.1160583>
- Guo, S., Yu, L., Chen, X., Zhang, Y., 2011. Modelling waiting time for passengers transferring from

- rail to buses. *Transp. Plan. Technol.* 34, 795–809. <https://doi.org/10.1080/03081060.2011.613589>
- Hess, D.B., Brown, J., Shoup, D., 2004. Waiting for the Bus. *J. Public Transp.* 7, 67–84. <https://doi.org/10.5038/2375-0901.7.4.4>
- Jolliffe, J.K., Hutchinson, T.P., 1975. A Behavioural Explanation of the Association Between Bus and Passenger Arrivals at a Bus Stop. *Transp. Sci.* <https://doi.org/10.1287/trsc.9.3.248>
- Liu, Y., Bunker, J., Ferreira, L., 2010. Transit users' route-choice modelling in transit assignment: A review. *Transp. Rev.* 30, 753–769. <https://doi.org/10.1080/01441641003744261>
- Luethi, M., Weidmann, U.A., Nash, A., 2007. Passenger Arrival Rates at Public Transport Stations. *Transp. Res. Board 86th Annu. Meet. No. 07-063*.
- Nielsen, O.A., 2000. A stochastic transit assignment model considering differences in passengers utility functions. *Transp. Res. Part B Methodol.* 34, 377–402.
- Nuzzolo, A., Crisalli, U., Rosati, L., 2012. A schedule-based assignment model with explicit capacity constraints for congested transit networks. *Transp. Res. Part C Emerg. Technol.* 20, 16–33. <https://doi.org/10.1016/j.trc.2011.02.007>
- Nuzzolo, A., Crisalli, U., Rosati, L., Comi, A., 2015. DYBUS2: A Real-Time Mesoscopic Transit Modeling Framework, in: *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*. pp. 303–308. <https://doi.org/10.1109/ITSC.2015.59>
- Nuzzolo, A., Russo, F., Crisalli, U., 2001. A Doubly Dynamic Schedule-based Assignment Model for Transit Networks. *Transp. Sci.* 35, 268–285. <https://doi.org/10.1287/trsc.35.3.268.10149>
- Nygaard, M.F., Tørset, T., 2016. Waiting Time Strategy for Public Transport Passengers, in: *Proceedings from the Annual Transport Conference at Aalborg University*. pp. 1–14.
- Nökel, K., Wekeck, S., 2009. Boarding and Alighting in Frequency-Based Transit Assignment. *Transp. Res. Rec. J. Transp. Res. Board* 60–67. <https://doi.org/10.3141/2111-08>
- O'Flaherty, C.A., Mangan, D.O., 1970. Bus Passenger Waiting Times in Central Areas. *Traffic Eng. Control* 11, 419–421.
- Parbo, J., Nielsen, O.A., Prato, C.G., 2014. User perspectives in public transport timetable optimisation. *Transp. Res. Part C Emerg. Technol.* 48, 269–284. <https://doi.org/10.1016/j.trc.2014.09.005>
- Pelletier, M.-P., Trépanier, M., Morency, C., 2011. Smart card data use in public transit: A literature review. *Transp. Res. Part C* 19, 557–568. <https://doi.org/10.1016/j.trc.2010.12.003>
- R Core Team, 2017. *R: A Language and environment for statistical computing*. Vienna, Austria.
- Rejsekort A/S, 2017. Travel card in numbers (in Danish) [WWW Document]. URL <http://www.rejsekort.dk/om-rejsekort/rejsekort-i-tal.aspx> (accessed 4.11.17).
- Richter, T., Landex, A., Andersen, J.L.E., 2013. Precise and accurate train run data: Approximation of actual arrival and departure times, in: *WCRR (World Congress Railway Research)*. International Association of Railways, Sydney, Australia, pp. 1–10.
- Schmöcker, J.D., Fonzone, A., Shimamoto, H., Kurauchi, F., Bell, M.G.H., 2011. Frequency-based transit assignment considering seat capacities. *Transp. Res. Part B Methodol.* 45, 392–408.

<https://doi.org/10.1016/j.trb.2010.07.002>

- Seddon, P.A., Day, M.P., 1974. Bus Passenger Waiting Times in Greater Manchester. *Traffic Eng. Control* 15, 442–445.
- Shlayan, N., Kurkcu, A., Ozbay, K., 2016. Exploring pedestrian bluetooth and WiFi detection at public transportation terminals. *IEEE Conf. Intell. Transp. Syst. Proceedings, ITSC* 229–234. <https://doi.org/10.1109/ITSC.2016.7795559>
- Szeto, W.Y., Jiang, Y., Wong, K.I., Solayappan, M., 2013. Reliability-based stochastic transit assignment with capacity constraints: Formulation and solution method. *Transp. Res. Part C Emerg. Technol.* 35, 286–304. <https://doi.org/10.1016/j.trc.2011.09.001>
- Szeto, W.Y., Solayappan, M., Jiang, Y., 2011. Reliability-based transit assignment for congested stochastic transit networks. *Comput. Civ. Infrastruct. Eng.* 26, 311–326. <https://doi.org/10.1111/j.1467-8667.2010.00680.x>
- Zhao, J., Frumin, M., Wilson, N., Zhao, Z., 2013. Unified estimator for excess journey time under heterogeneous passenger incidence behavior using smartcard data. *Transp. Res. Part C Emerg. Technol.* 34, 70–88. <https://doi.org/10.1016/j.trc.2013.05.009>
- Zhou, M., Wang, D., Li, Q., Yue, Y., Tu, W., Cao, R., 2017. Impacts of weather on public transport ridership: Results from mining data from different sources. *Transp. Res. Part C Emerg. Technol.* 75, 17–29. <https://doi.org/10.1016/j.trc.2016.12.001>

## List of figures

Figure 1; Illustration of the density function of a mixture distribution with  $\zeta=0.67$ , and the shape parameters of the beta component  $\alpha=1.5$  and  $\beta=4$ .

Figure 2; Distribution of passenger waiting times before and after data cleaning. Left: Full data for 20-minute headways. Right: After data cleaning by taking into account realised timetable and delayed trains as well as corrected dwell times for suburban trains.

Figure 3; Comparison of manually collected data and Rejsekort data for passenger arrivals at Bernstorffsvej station on August 11, 2016.

Figure 4; Comparison of the fit of various distributions to passenger waiting times.

Figure 5; Mixture of uniform and beta distributions of passenger waiting times for 5-60 minutes headways, and results overview (lower right).

Figure 6; Comparison between effects of traditional timetables and frequency-based timetables on passenger waiting time distributions.

Figure 7; Comparison of waiting time distributions for passengers at stations with choice options at origin station.

Figure 8; Difference between passenger waiting time distributions for passengers travelling during peak hours and off-peak hours for headways of 20 minutes (left) and 30 minutes (right).

## List of tables

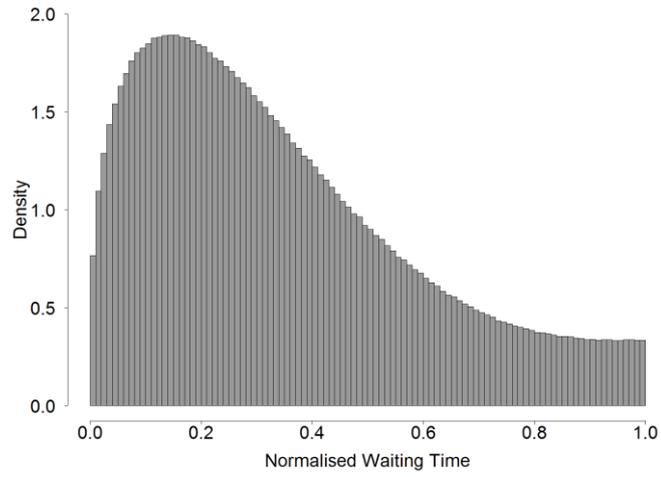
Table 1; Station characteristics included in the analyses (\* based on Dyrberg and Christensen (2015))

Table 2; Statistical tests of the fit of beta-mixture and Johnson-mixture to passenger waiting times for schedule-based public transport services (suburban and regional trains).

Table 3; Parameter values for the beta-component of the mixture distributions at the evaluated headways for schedule-based public transport services.

Table 4; The percentage of passengers arriving randomly as function of time of travel. Average passenger waiting times in parenthesis [minutes]. \*\* <200 observations; \* <500 observations!

Table 5; The percentage of passengers arriving randomly as function of station layouts. Average passenger waiting times in parenthesis [minutes].



*Figure 1; Illustration of the density function of a mixture distribution with  $\zeta=0.67$ , and the shape parameters of the beta component  $\alpha=1.5$  and  $\beta=4$ .*

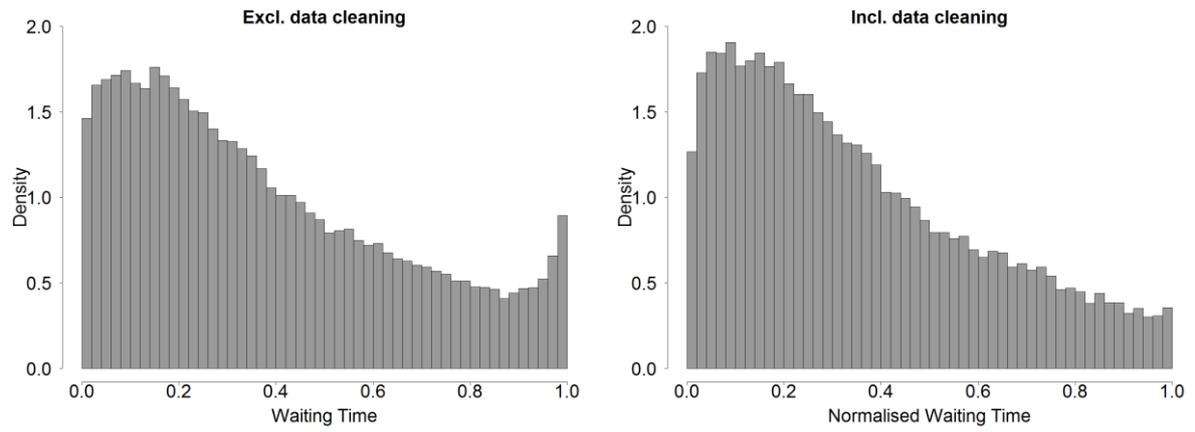


Figure 2; *Distribution of passenger waiting times before and after data cleaning. Left: Full data for 20-minute headways. Right: After data cleaning by taking into account realised timetable and delayed trains as well as corrected dwell times for suburban trains.*

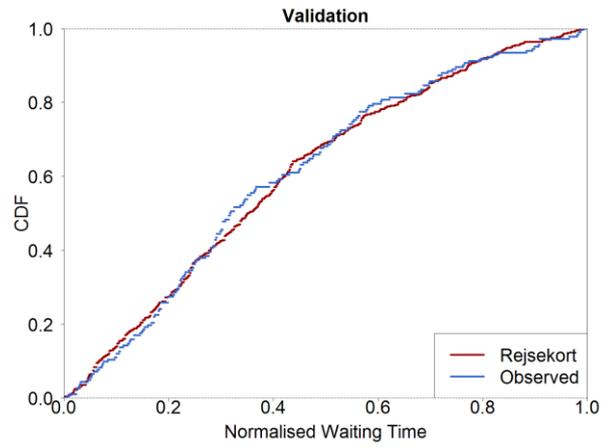


Figure 3; Comparison of manually collected data and Rejsekort data for passenger arrivals at Bernstorffsvej station on August 11, 2016.

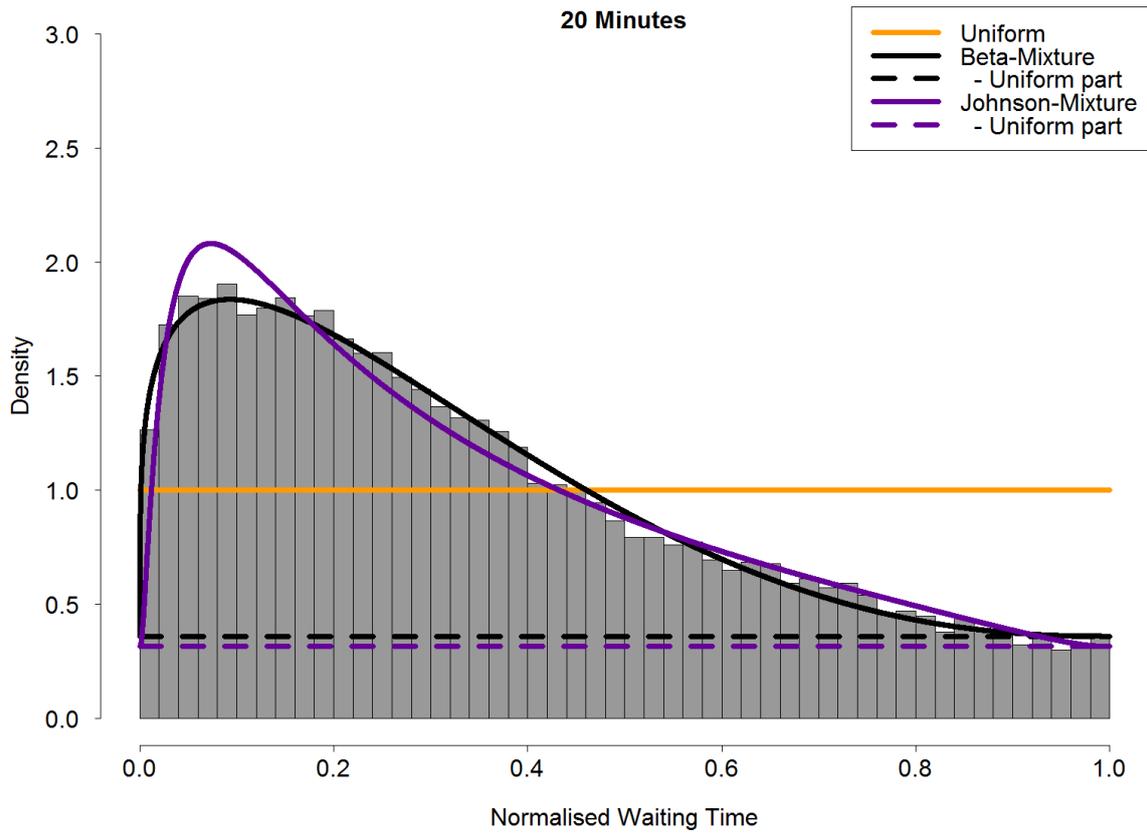


Figure 4; Comparison of the fit of various distributions to passenger waiting times.

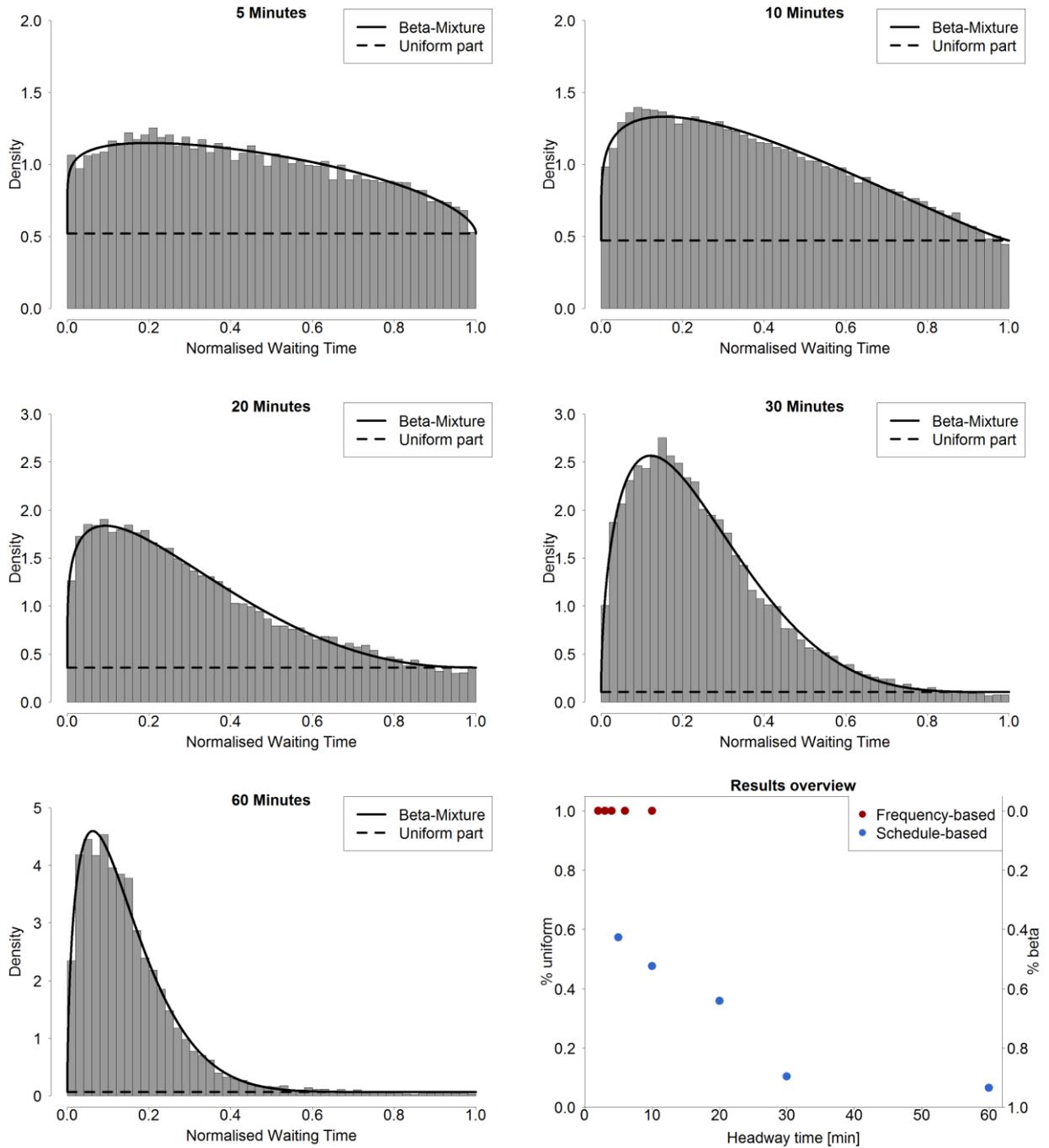


Figure 5; Mixture of uniform and beta distributions of passenger waiting times for 5-60 minutes headways, and results overview (lower right).

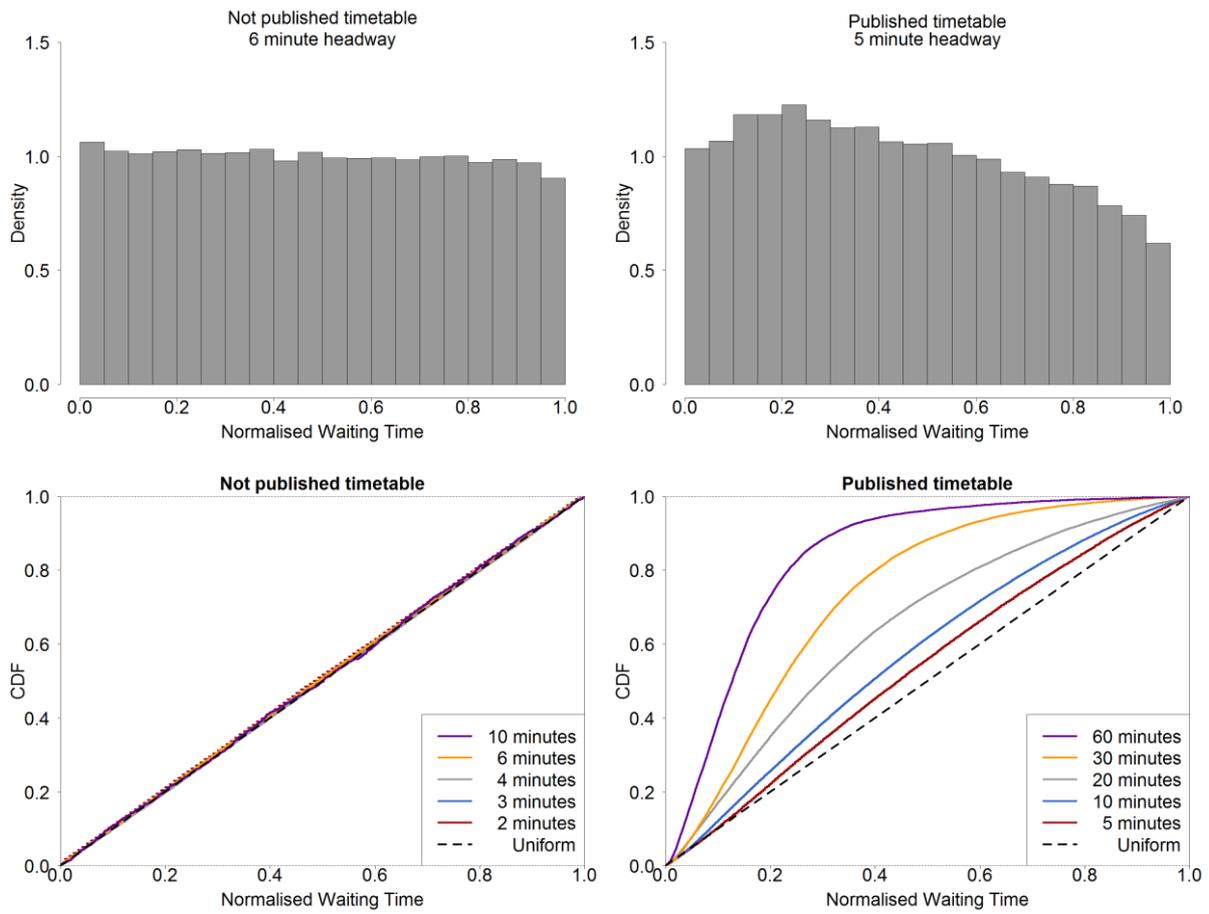


Figure 6; Comparison between effects of traditional timetables and frequency-based timetables on passenger waiting time distributions.

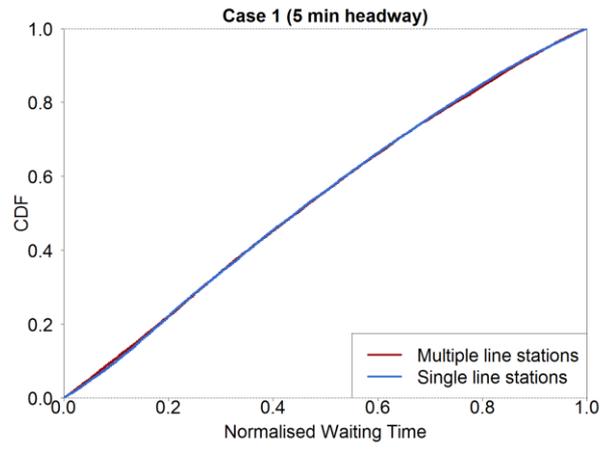


Figure 7; Comparison of waiting time distributions for passengers at stations with choice options at origin station.

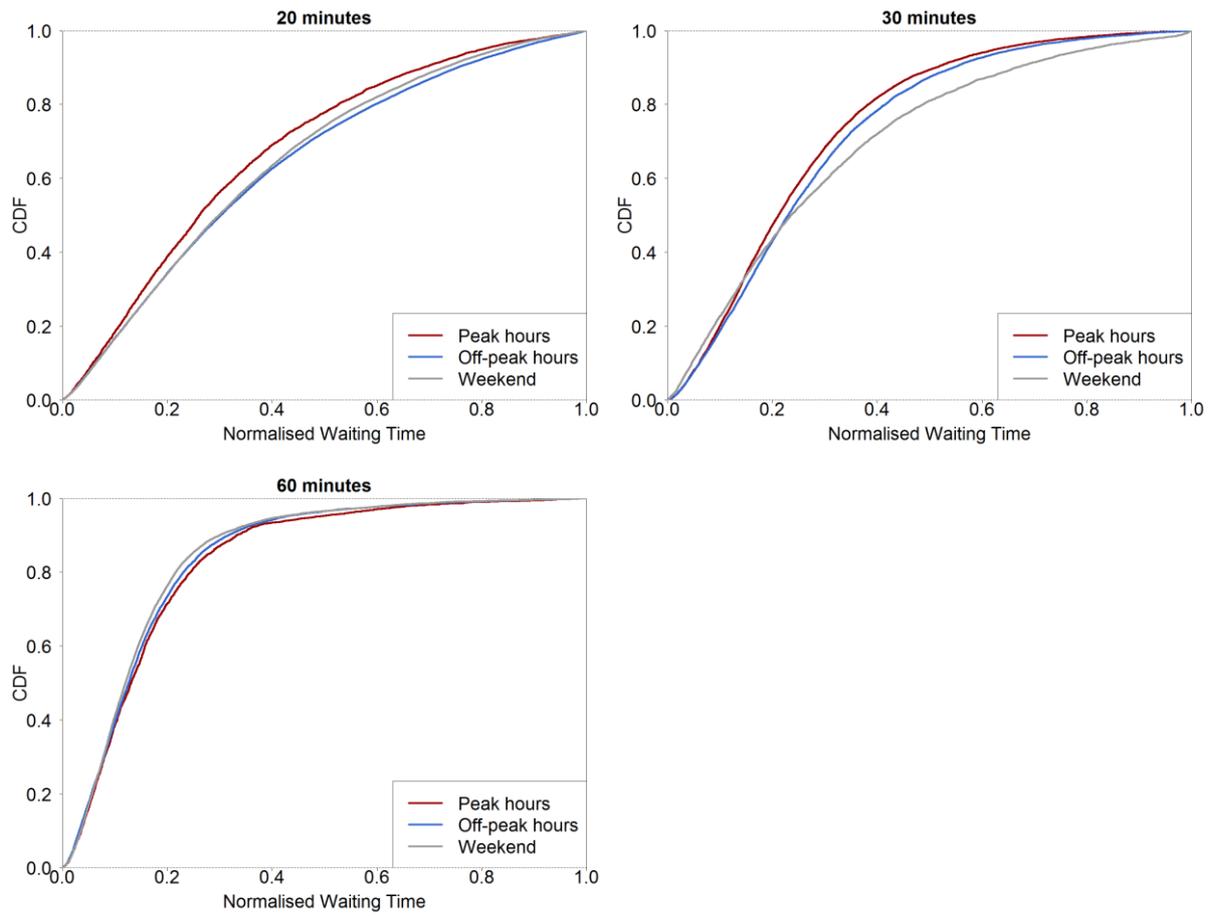


Figure 8; Difference between passenger waiting time distributions for passengers travelling during peak hours and off-peak hours for headways of 20 minutes (left) and 30 minutes (right).

Parameter	Value used in comparisons
Shelters*	No
	Yes
Underground*	No
	Yes
Station layout / Ease of wayfinding*	Easy
	Difficult
Land use <sup>4</sup>	Leisure-oriented
	Commute-oriented
Time-of-travel	Weekday, Peak (6-9, 15-18)
	Weekday, Off-peak (9-15, 18-0)
	Weekday, Night (00-06)
	Weekend, Day (06-00)
	Weekend, Night (00-06)

*Table 1; Station characteristics included in the analyses (\* based on Dyrberg and Christensen (2015))*

<sup>4</sup> This data is based on the Danish HSK land use data, which is a database containing detailed disaggregate information on the land uses in the Greater Copenhagen area.

Headway	No. obs	Beta-mixture			Johnson-mixture		
		D	p-Value	AIC	D	p-Value	AIC
5	39,140	0.0064	0.08442	-949.70	0.0079	0.01566	-888.34
10	184,917	0.0051	0.00015	-14,917.55	0.0084	<0.00001	-14,523.44
20	48,670	0.0063	0.03940	-13,843.48	0.0144	<0.00001	-13,444.79
30	24,589	0.0098	0.01845	-20,701.18	0.0265	<0.00001	-20,109.29
60	12,702	0.0096	0.19566	-22,259.63	0.0382	<0.00001	-21,715.79

*Table 2; Statistical tests of the fit of beta-mixture and Johnson-mixture to passenger waiting times for schedule-based public transport services (suburban and regional trains).*

Headway	Beta-share	Uniform-share	Beta-component	
	$\zeta$	$1 - \zeta$	$\alpha$	$\beta$
5	0.43	0.57	0.41	2.85
10	0.52	0.48	0.36	3.39
20	0.64	0.36	0.27	4.57
30	0.90	0.10	0.24	6.52
60	0.93	0.07	0.14	11.20

*Table 3; Parameter values for the beta-component of the mixture distributions at the evaluated headways for schedule-based public transport services.*

Headway	Weekday All hours	Weekday Peak hours	Weekday Off-Peak hours	Weekday Night hours	Weekend Day hours	Weekend Night hours
5	57% (2.3)	41% (2.3)	67% (2.3)	-	-	-
10	48% (4.3)	45% (4.2)	50% (4.3)	60% (4.4)*	50% (4.3)	65% (4.1)**
20	36% (7.1)	26% (6.4)	37% (7.2)	30% (6.3)**	28% (7.0)	47% (7.6)*
30	10% (7.9)	10% (7.6)	11% (8.2)	6% (6.3)**	12% (8.4)	31% (9.2)
60	7% (9.8)	8% (10.2)	6% (9.7)	4% (8.0)**	7% (9.2)	8% (9.7)*

*Table 4: The percentage of passengers arriving randomly as function of time of travel. Average passenger waiting times in parenthesis [minutes]. \*\* <200 observations; \* <500 observations!*

Headway	All stations	Easy wayfinding	Difficult wayfinding
5	57% (2.3)	38% (2.3)	63% (2.3)
10	48% (4.3)	42% (4.2)	50% (4.3)
20	36% (7.0)	31% (6.5)	38% (7.3)
30	10% (7.9)	7% (6.7)	11% (8.3)
60	7% (9.8)	5% (7.1)	7% (9.8)

*Table 5: The percentage of passengers arriving randomly as function of station layouts. Average passenger waiting times in parenthesis [minutes].*