The spread of passengers on platforms and dwell times for commuter trains: A case study using automatic passenger count data

Ruben A. Kuipers[a,b,*], Carl-William Palmqvist[a,b]
[a] Department of Technology and Society, Lund University, Lund, Sweden
[b] K2 Swedish Knowledge Centre for Public Transport, Lund, Sweden
ruben.kuipers@tft.lth.se, carl-william.palmqvist@tft.lth.se

Extended abstract submitted for presentation at the 11th Triennial Symposium on Transportation Analysis conference (TRISTAN XI)
June 19-25, 2022, Mauritius Island

March 22, 2022

Keywords: Dwell time, passengers, concentrated boarding, delays, railway

1 INTRODUCTION

An increasing pressure to reduce greenhouse gas emissions calls for changes in our daily patterns. One of these changes is how we travel. Transport accounts for around 28% of the emissions in Europe and a modal shift towards public transport is considered to be one of the ways to reduce greenhouse gas emissions (Blayac & Stéphan, 2021). To achieve this modal shift it is important to increase the attractiveness of trains as a mode of transport. Punctuality, travel time reliability, and travel time predictability are found to be the most important indicators of the attractiveness of trains from the perspective of users (van Loon et al., 2011). Improvements of these key indicators can thus be seen as an important step to help achieve a modal shift away from private vehicles and towards train travel. Improvements here also increase the efficiency and robustness of railway operations, allowing for a higher frequency of trains to be operated and better use of existing railway capacity. One way to do make improvements to these key indicators is to reduce dwell time delays at stations. Dwell times are considered to be one of the main constraints in terms of rail capacity, and have as much of an impact on capacity as maximum running speeds of trains (Harris, 2005). Several factors which influence dwell times have been identified in the past, such as rolling stock design (Thoreau et al., 2016), friction between boarding and alighting passengers (Seriani et al., 2019), and the volume of passengers (Palmqvist et al., 2020) for example. To add to the understanding of dwell times, the study we present here focuses on the relation between the spread of passengers on a station platform before boarding a train and dwell times. The spread of passengers, also known as concentrated boarding (Oliveira et al., 2019), has been shown to have a large impact on dwell times.

Measures to help spread out passengers more evenly between the available doors during the boarding procedure have been suggested in the past, ranging from platform markings signalling that people should spread out across to the platform as mentioned by Oliveira et al. (2019), to providing real-time information on the onboard crowding levels (Zhang et al., 2017). When implementing such measures it is not only important to understand the effectiveness of the measures itself but it is also important to understand the context in which these measures can have a beneficial effect on dwell times. The study we present here focuses on the latter and aims to study the relation between the spread of passengers between the available doors and dwell times on a network-wide level to help understand where it is relevant for interventions to be made.

2 Method

2.1 Data availability and preparation

For our study, we make use of real-world data from the commuter trains in the region of Scania in Southern Sweden, collected during 2017 and 2018. In total there are 99 commuter trains in use on the network which each consists of four carriages. Each train has a total of 240 seats available and five doors on either side of the train. Individual trainsets can be combined to increase capacity, increasing the available seats and doors. The majority of the trains in operation during the studied period consisted of one vehicle (67%), and the second most common are trains consisting of two vehicles (32%), having eight carriages, ten doors, and 480 seats. Most of the commuter trains in use are equipped with an automatic passenger count system. These systems register the volume of both boarding and alighting passengers by making use of infrared beams at each door of a train. The onboard system also provides the actual arrival and departure times in a magnitude of seconds.

Data preparation consisted of excluding stops where secondary activities take place and introducing a lower bound for the volume of boarding passengers. Delays at stops where secondary activities, such as crew changes, take place can be due to reasons other than passengers. To avoid including such stops, we limited ourselves to the shortest scheduled dwell time at each station, as extra dwell time is scheduled at stops with secondary activities. In practice this means we limit ourselves to stops with either 60 or 120 seconds of scheduled dwell times. A lower bound for the number of boarding passengers to avoid stops where the number of boarding passengers is such that concentrated boarding will always occur. The value for this lower bound was set at five boarding passengers per train, following a criterion previously used by Christoforou et al. (2020). Once these steps were completed, we were left with information on *741,008 station stops* for further analyses.

2.2 Logistic regression

To analyse the influence of the spread of passengers on the probability of dwell times exceeding the scheduled time we perform a logistic regression analysis using the statistical software package R. The logistic regression is similar to multiple linear regression, with the difference being that the response variable is binomial, and the result relates to the conditional probability that an outcome occurs based on a set of explanatory variables (Sommet & Morselli, 2017; Sperandei, 2014). Using the probability of a dwell time delay occurring is more robust to outliers compared to the size of dwell time delays. The basic model for the log-odds takes for the form of Equation 1 (Sperandei, 2014), and the probabilities of an outcome occurring based on these log-odds can be determined by making use of Equation 2 (Sommet & Morselli, 2017).

$$\log\left(\frac{\pi}{1-\pi}\right) = \beta_0+\beta_1X_1+\beta_2X_2+.....\beta_mX_m \qquad (1)$$

$$\text{Probability} = \frac{\exp\left(\beta_0+\beta_1X_1+\beta_2X_2+.....\beta_mX_m\right.}{1+\exp\left(\beta_0+\beta_1X_1+\beta_2X_2+.....\beta_mX_m\right.} \qquad (2)$$

The variables under consideration for this study and their respective units are shown in Table 1. The volume of boarding passengers was normalized based on the range between the minimum and maximum values. The ratio between boarding and alighting passengers is included as it provides an indication for the type of station in terms of travel patterns. The spread of passengers is determined based on the proportion of passengers boarding through the busiest door. To be able to compare trains with one and two carriages, we normalized these values in the same way as the volume of boarding passengers. We consider the spread of passengers to be even (0) when all boarding passengers are evenly spread between each door and an uneven spread (1) indicates a situation where all passengers board through a single door. During certain hours of the day, some stations will show a large number of boarding passengers, passengers commuting from a place, and vice versa with passengers commuting to a place. Dummy variables for the time of the day, scheduled dwell time, and arrival status were included in addition to the variables related to passengers. In addition to this, interaction terms were included for the number of boarding passengers and the spread of passengers, as well as the scheduled dwell time and time of the day. Onboard congestion is not included here since passengers can move freely through the train, meaning that the door they use for boarding does not determine their seating position. We can, therefore, not link the number of boarding passengers per door to the actual onboard crowding in that section of the train.

Collinearity for the included variables was tested using the VIF index which showed that there was no problematic multicollinearity present in the model for the variables under consideration. Testing for linearity between the continuous independent variables and the log-odds of the dependent variable revealed that this assumption was not met for the ratio between boarding and alighting passengers. To capture this non-linear relationship and improve our model fit, we included a polynomial term for this variable in our model.

Table 1: Overview of the variables used in the logistic regression model.

	Variable	**Unit**
Y	Probability of dwell time delay	
X1	Arrival status	On time ; late ; early
X2	Scheduled dwell time	60s ; 120s
X3	Time of the day	Peak ; off-peak
X4	Volume of boarding passengers on a train level	Interval 0 : 1
X5	Spread of passengers between the available doors	Interval 0 : 1
X6	Ratio of boarding versus alighting passengers on a door level	Interval 0 : 1

3 Results

To understand the relation between the spread of passengers between doors and dwell times we conducted a numerical example based on the logistic model. The results of this are shown in Table 2. We make use of values at the 25th, 50th, and 75th percentiles for both passenger volume and the ratio of boarding passengers as input for this example. We do so to highlight cases that are common to occur, rather than focussing on the maximum values for extreme cases. For sake of clarity, we only show the probabilities of delays for stops where a train arrives on time, during peak hours. The results show that spread of passengers has a stronger effect on the probability of a dwell time being delayed when passenger volumes are higher. Comparing the change in probabilities between the 25th and 75th percentile of passenger volumes for an *even* (panel B) and *uneven* (panel C) spread of passengers we see a larger increase in the probabilities of delays for an uneven (an average increase of 0.18) compared to an even (an average increase of 0.02) spread of passengers. The results also show that the probability of dwell time delays is lower when more passengers board than alight. This effect is in a similar order of magnitude for both an even and uneven spread of passengers.

Table 2 Predicted probabilities of a dwell time delay occurring at stops with a scheduled dwell time of 60 seconds (left) and 120 seconds (*right*).

	Spread of passengers = even	**Spread of passengers = median**	**Spread of passengers = uneven**
Panel A (Passenger volume= 25th percentile)			
Ratio of boarding = 25th percentile	0.57 / *0.20*	0.58 / *0.21*	0.62 / *0.24*
Ratio of boarding = median	0.53 / *0.18*	0.54 / *0.18*	0.58 / *0.21*
Ratio of boarding = 75th percentile	0.51 / *0.17*	0.52 / *0.17*	0.56 / *0.19*
Panel B (Passenger volume= median)			
Ratio of boarding = 25th percentile	0.58 / *0.21*	0.60 / *0.23*	0.67 / *0.28*
Ratio of boarding = median	0.53 / *0.18*	0.56 / *0.20*	0.63 / *0.25*
Ratio of boarding = 75th percentile	0.51 / *0.17*	0.54 / *0.18*	0.62 / *0.23*
Panel C (Passenger volume= 75th percentile)			
Ratio of boarding = 25th percentile	0.59 / *0.21*	0.60 / *0.23*	0.79 / *0.42*
Ratio of boarding = median	0.54 / *0.19*	0.56 / *0.20*	0.76 / *0.38*
Ratio of boarding = 75th percentile	0.53 / *0.17*	0.54 / *0.18*	0.75 / *0.36*

4 Discussion and conclusion

The study presented here focuses on the relation between the spread of passengers between the available doors and dwell time delays, for commuter trains. The study aims to help understand the context in which measures aimed at spreading out passengers would be most beneficial. Based on our findings we can state that the degree of concentrated boarding has the strongest effect on dwell time delays at stops with larger volumes of boarding passengers. The relative change in probabilities at such stops is larger compared to instances with a lower volume of boarding passengers. In addition to this, we find that this effect is relatively constant across different levels for the ratio between boarding and alighting passengers. This indicates that the type of station in terms of travel patterns, i.e. passengers mainly travelling to or from a station, is less relevant in the light of trying to improve dwell time punctuality by spreading out passengers more evenly. Based on these results we can say that measures to more evenly spread out passengers are most beneficial at stations with a larger expected volume of boarding passengers. Applying measures at such stations can help to reduce the probability of trains incurring a dwell time delay and lead to an overall improvement in travel time reliability, predictability, and punctuality. A limitation of our study, however, is the use of probabilities rather than delay sizes. Although this approach is more robust, this does mean that both small and large delays are shown in the same way. To overcome this, future studies can include different thresholds for dwell time delays, thus showing the probabilities for a wider range of delays.

5 References

Blayac, T., & Stéphan, M. (2021). Are retrospective rail punctuality indicators useful? Evidence from users perceptions. *Transportation Research Part A: Policy and Practice*, *146*, 193–213. https://doi.org/10.1016/j.tra.2021.01.013

Harris, N. G. (2005). Train boarding and alighting rates at high passenger loads. *Journal of Advanced Transportation*, *40*(3), 249–263. https://doi.org/10.1002/atr.5670400302

Oliveira, L. C., Fox, C., Birrell, S., & Cain, R. (2019). Analysing passengers' behaviours when boarding trains to improve rail infrastructure and technology. *Robotics and Computer-Integrated Manufacturing*, *57*, 282–291. https://doi.org/10.1016/j.rcim.2018.12.008

Palmqvist, C. W., Tomii, N., & Ochiai, Y. (2020). Explaining dwell time delays with passenger counts for some commuter trains in Stockholm and Tokyo. *Journal of Rail Transport Planning and Management*, *14*, 100189. https://doi.org/10.1016/j.jrtpm.2020.100189

Pritchard, J. (2018). Providing Improved Crowding Information to Provide Benefits for Rail Passengers and Operators. In N. A. Stanton (Ed.), *Advances in Human Aspects of Transportation* (Vol. 597, pp. 973–984). Springer International Publishing. https://doi.org/10.1007/978-3-319-60441-1

Seriani, S., Fernandez, R., Luangboriboon, N., & Fujiyama, T. (2019). Exploring the Effect of Boarding and Alighting Ratio on Passengers' Behaviour at Metro Stations by Laboratory Experiments. *Journal of Advanced Transportation*, *2019*. https://doi.org/10.1155/2019/6530897

Sommet, N., & Morselli, D. (2017). Keep Calm and Learn Multilevel Logistic Modeling: A Simplified Three-Step Procedure Using Stata, R, Mplus, and SPSS. *International Review of Social Psychology*, *30*(1), 203–218. https://doi.org/10.5334/irsp.90

Sperandei, S. (2014). Understanding logistic regression analysis. *Biochemia Medica*, 12–18. https://doi.org/10.11613/BM.2014.003

Thoreau, R., Holloway, C., Bansal, G., Gharatya, K., Roan, T. R., & Tyler, N. (2016). Train design features affecting boarding and alighting of passengers. *Journal of Advanced Transportation*, *50*(8), 2077–2088. https://doi.org/10.1002/atr.1446

van Loon, R., Rietveld, P., & Brons, M. (2011). Travel-time reliability impacts on railway passenger demand: A revealed preference analysis. *Journal of Transport Geography*, *19*(4), 917–925. https://doi.org/10.1016/j.jtrangeo.2010.11.009

Zhang, Y., Jenelius, E., & Kottenhoff, K. (2017). Impact of real-time crowding information: A Stockholm metro pilot study. *Public Transport*, *9*(3), 483–499. https://doi.org/10.1007/s12469-016-0150-y