PASSENGER DATA COMPLEXITY IN TRAM STOP DWELL TIME MODELLING

Maja Ahac[1], Leonard Cvitić[2], Igor Majstorović[1], Saša Ahac[1], Silvio Bašić[1]
[1] *University of Zagreb, Faculty of Civil Engineering, Croatia*
[2] *Neborex d.o.o., Croatia*

Abstract

The stop dwell time can be modelled by using the volumes of boarders and alighters, and it is a common conclusion that the use of additional information on the number and width of doors, number of seats, and number of through standees in model creation improves its estimation of stop dwell time. However, such an approach demands detailed knowledge and/or assumptions on passenger distribution both inside the vehicle and on the stop platform, which makes the model creation and its application more challenging. The research presented in this paper is focused on the passenger input data requirements for the creation of tram stop dwell time prediction models. It is based on passenger and tram dwell time data collected at an island tram stop in Zagreb. The data acquisition included the field recording of the trams in operation during five working days, laboratory processing of 70 hours of collected video data, and creation of a synthesized database of observed and measured data. Three different multiple linear regression models for tram dwell time prediction were created, with the following independent variables: (1) the volume of boarders and alighters and a type of passenger flow transiting through the busiest tram doors, (2) the volume of boarders and alighters transiting through the busiest tram doors, and (3) the total volume of boarders and alighters per tram. The cross-validation of the model showed that passenger input data simplification has a minor effect on the model's goodness of fit, and a mild effect on it's accuracy and precision, which could be adequately addressed by the application of a larger operating margin.

Keywords: island stop, passenger volume, passenger flow, alighters, boarders

1 Introduction

The stop dwell time is defined as the difference between the public transport (PT) vehicle departure and arrival times. The arrival time is defined as the time when the vehicle changes its state from moving to standing still, and vice versa for the departure time. There are at least five processes between the arrival time and the departure time: door unlocking, door opening, passengers alighting and boarding, door closing, and vehicle dispatching [1]. The stop dwell time prediction is a major issue in urban PT vehicle travel speed modelling, used for the definition of travel time and timetable creation. This is due to numerous stop dwell time influential factors, variable both in space and time, like passenger flow, vehicle and stop design characteristics, and traffic organization. This problem is especially pronounced on high-frequency and high-ridership PT systems, with long routes, and consecutive stops that are relatively close to one another [2], such as tram systems.

DOI: https://doi.org/10.5592/CO/CETRA.2022.1427

According to [3], models for the prediction of urban PT travel time can be grouped as (1) models based on historical data, i.e., on the observation that dwell times are repeatable between days, at the same time of the day, and the same day of the week, or (2) statistical models based on regression analyses which use several identified influential factors as independent variables and make a prediction based on their statistic distributions and correlations. Historical data models rely on average values in previous days as a prediction factor and therefore are reliable only when the traffic patterns are relatively constant. On the other hand, the precision of the regression models depends on all the variables that need to be recognized and incorporated into the model. Therefore, research of different sources of model's uncertainty, especially regarding the passenger traffic impact on PT stop dwell time [4] is ongoing. As passengers are not distributed uniformly on the platform [5] or in the vehicle [6], most models define stop dwell time as the maximum time needed at one vehicle door for passengers to alight and board (passenger flow time) plus additional fixed time consisting of doors opening and closing time and an operating margin [7]. Operating margin considers all factors that are variable both in space and time, i.e., variations in passenger volumes, and deviations from the timetable. This is the extra time added to a line's headway to allow for irregular operation and ensure that one vehicle does not delay the following one. According to [8], it is suggested that a range from 10 s to 30 s should be considered for the operating margin for tram stops. It should be lower the higher the frequencies of the vehicles are. When capacity is not an issue, 25 s or more is recommended. The operating margin can be reduced to 20 s or even 15 s if it is necessary to provide sufficient service to meet the estimated demand. The passenger flow time (the time passengers need to board and/or alight) depends on the number of passengers transiting through the busiest doors, the type of their flow, and tram type defining the number of channels per door and floor height [7, 8].
Previous research has shown that the use of additional information on the number and width of doors, number of seats, and number of through standees in model creation improves its estimation of stop dwell time [9]. However, such an approach demands detailed knowledge and/or assumptions on passenger distribution both inside the vehicle and on the stop platform, which makes both the model creation and its application more challenging. The research presented in this paper is focused on the influence of the passenger input data complexity on the accuracy and precision of statistical multiple linear regression (MLR) models for tram stop dwell time prediction. The investigation is based on passenger and tram dwell time data collected at an island tram stop in the City of Zagreb and is performed to answer the question of whether the tram stop dwell time can be adequately estimated only by the total number of boarders and alighters per stopped tram.

2 Data acquisition and sample creation

The tram network in the City of Zagreb consists of a total of 58 km of double, 1.000 mm gauge, tracks (excluding the tracks in two tram depots). The network contains 18 tram turnarounds, nine of which are PT terminals, and 240 single and 18 double tram stops. 50 % of the tracks are located inside the street carriageway, adjacent to the sidewalks which are used as platforms for 40 % of the stops. Along city avenues, tracks are laid in separate central corridors, and stop platforms are constructed as elevated islands, usually far side larger street intersections. Tram transport is organized through 15 daytime (4 a.m.–12 a.m.) and four nighttime (12 a.m.–4 a.m.) lines in a total length of 216.5 km. 266 trams of 6 different construction types annually transport more than 190 million passengers [10].
During investigation location scouting it was decided to research an island stop platform where stop operations will not be affected by individual car traffic, and there will be no interference between tram passengers and pedestrians. The chosen tram stop is located on the southern edge of the city centre. As it is situated near several higher education institutions

it is used mostly by younger passengers, randomly arriving during the day, so distinctive passenger boarding or alighting peak volume hours were not expected. The passengers approach the platform on the far end of the stop sign via the crosswalk. The distance between the stop and the nearest downstream signal is 85 m, so it was presumed that the tram dwell time won't be affected by it.
The measurements at the selected stop included the field recording of the trams in operation, laboratory processing of collected video data, and creation of a synthesized database of observed and measured data. Measurements were conducted in October 2020 during five working days, Monday to Friday. It should be emphasized that the research was conducted during COVID-19 restrictions in the form of limiting the number of passengers in the vehicle to 40 % of the maximum vehicle occupancy, and a ban on purchasing the tickets from the driver.
The recording of the tram vehicles operation and passenger exchange on the stop was performed by Miovision Scout devices, mounted on a stop sign. The recording on each day lasted from 6 a.m. to 8 p.m., during which a total of 70 h of video material was recorded. Laboratory data processing included the video material analysis and the collection of data on the tram vehicle type, arrival and departure time at the tram stop, and the number of passengers boarding ($P_{D, B}$) and alighting ($P_{D, A}$) the tram through a specific door (D). The collected data were combined into a single database. The database analysis showed that out of 935 recorded trams that served the stop, 66 % of them were the high-floor, four-doors TMK301 type tram vehicles with the door opening and closing time of 3 s. To ensure the sample representativeness, it was decided to continue the investigation by considering only the TMK301 tram data. Furthermore, two-step data filtering was performed. In the first step, the recordings with no passenger exchange ($P_{D, B} = P_{D, A} = 0$) were eliminated from the sample. In the second step, the dwell time of each TMK301 tram ($t_{d, o}$) was calculated as the difference between recorded tram departure and arrival time. The mean and standard deviation of the dwell time sample was used as a cut-off for identifying outliers: dwell time values more than three standard deviations away from the sample mean were excluded from further analysis. The resulting changes in the data sample size and five-number descriptive statistics for observed tram stop dwell time are given in Table 1.

Table 1 Observed dwell time ($t_{d, o}$) statistics [11]

$t_{d, o}$ statistics	Raw data	Filter 1 applied	Filter 2 applied
Sample size	595	580	571
Average [s]	12.8	12.8	12.5
Standard deviation [s]	4.9	4.9	3.9
Minimum [s]	4	5	5
1st Quarter [s]	10	10	10
Median [s]	12	12	12
3rd Quarter [s]	15	15	15
Maximum [s]	60	60	26

The filtered sample contains data on 4,762 passegers. The ratio between boarding and alighting passengers is 40:60. The hourly distribution of the counted passengers and trams, defined after the filtering process and presented in Figure 1 (left), shows that the alighters volume (P_A) is evenly distributed between 7 a.m. and 4 p.m., boarders (P_B) are more frequent in the afternoon period, and tram frequency is lower between 11 a.m. and 2 p.m. As it can be seen from Figure 1 (right), the latter has a direct influence on an increase in the average value of observed dwell time ($t_{d, o}$) during these hours.

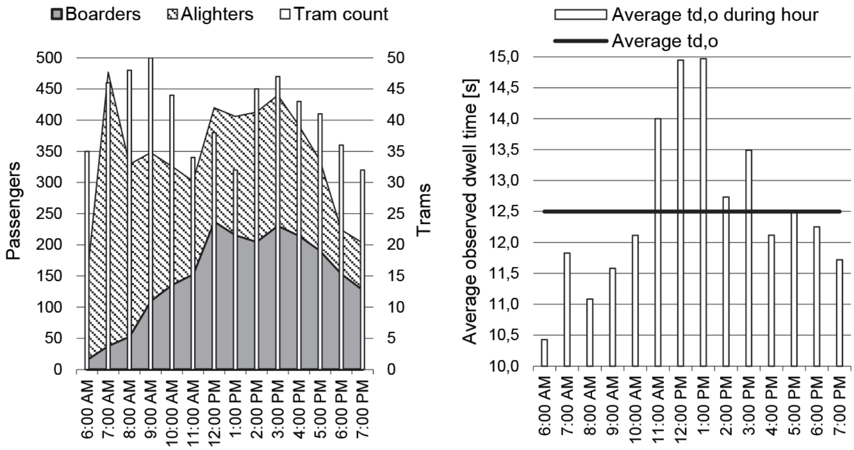

Figure 1 Passenger and tram volumes (left) and average observed dwell time values (right)

The filtered tram stop data was divided into training and test sample (in ratio 80:20) to cross-validate the LRM models. The test sample was created by excluding every fifth tram stopping observation from the total sample. The ratio of boarding and alighting passengers in both training and test sample remained 40:60.

3 Stop dwell time modelling and model validation

Three different MLR models were created by using the training sample passenger data as independent variables, and observed tram dwell time as a dependent variable. The models differ according to the input passenger data complexity.

The first model (MLR-DF) uses the training sample volumes of alighters and boarders transiting through the busiest tram doors ($P_{D, A}$ and $P_{D, B}$) and calculates the passenger's time to board and/or alight depending on the type of passenger flow. The flow is classified as mainly alighting (70 % or more passengers alighting), mainly boarding (70 % or more passengers boarding), or mixed (all other situations) [8]. Model for mainly alighting passengers (A) is given in Eq. (1), for mainly boarding passengers (B) in eqn (2), and for mixed passenger flow (M) in Eq. (3).

$$t_{d,\,m,\,MLR\text{-}FD(A)} = 8.6 + 0.9\,P_{D,\,A} + 1.5P_{D,\,B} \tag{1}$$

$$t_{d,\,m,\,MLR\text{-}FD(B)} = 7.8 + 0.8\,P_{D,\,A} + 1.3P_{D,\,B} \tag{2}$$

$$t_{d,\,m,\,MLR\text{-}FD(M)} = 7.9 + 1.0\,P_{D,\,A} + 1.7P_{D,\,B} \tag{3}$$

The second model (MLR-D), given in eqn (4), was created by neglecting the type of passenger flow and including only the volumes of alighters and boarders through the busiest tram doors ($P_{D, A}$ and $P_{D, B}$).

$$t_{d,\,m,\,MLR\text{-}D} = 8.2 + 1.0\,P_{D,\,A} + 1.3P_{D,\,B} \tag{4}$$

The third model (MLR-T), given in eqn (5), was created by including only the total number of passengers boarding and alighting per tram ($P_{T, A}$ and $P_{T, B}$).

$$t_{d,\,m,\,MLR\text{-}T} = 8.5 + 0.4P_{T,\,A} + 0.6P_{T,\,B} \tag{5}$$

All models predict a larger influence of boarding passengers on high-floor tram TMK301 dwell time and an operating margin of around 5 s. As it can be seen from the “goodness of model fit” data, given in Table 2, simplifying the input data in the form of neglecting the traffic flow type and busiest doors during model creation did not affect the accuracy of the model.
The LRM models were then applied to the test sample data to cross-validate their results. The comparison of the model’s accuracy and precision was made concerning the modeled results, i.e., predicted dwell time values differences, from the observed dwell time values of the test sample ($t_{d,\,o}$ - $t_{d,\,m}$). Figure 2 presents scatter plots of dwell time values in which the bubble size represents the passenger volumes, and a black line represents the identity line between the observed and calculated dwell time values. The distribution of MLR model errors (differences between observed and by models calculated values) is presented by boxplots in Figure 2, and by five-number descriptive statistics for errors given in Table 3. In general, scatter plots in which the values are more centered around the identity line are considered more accurate, and box plots that show less spread are considered more precise.

Table 2 Regression statistics

Model	Correlation coefficient	Coefficient of determination	Standard error	Observations
MLR-FD (A)	0.58	0.34	2.73	186
MLR-FD (B)	0.71	0.50	2.79	93
MLR-FD (M)	0.65	0.42	3.17	178
MLR-D	0.65	0.43	2.93	457
MLR-T	0.66	0.44	2.90	457

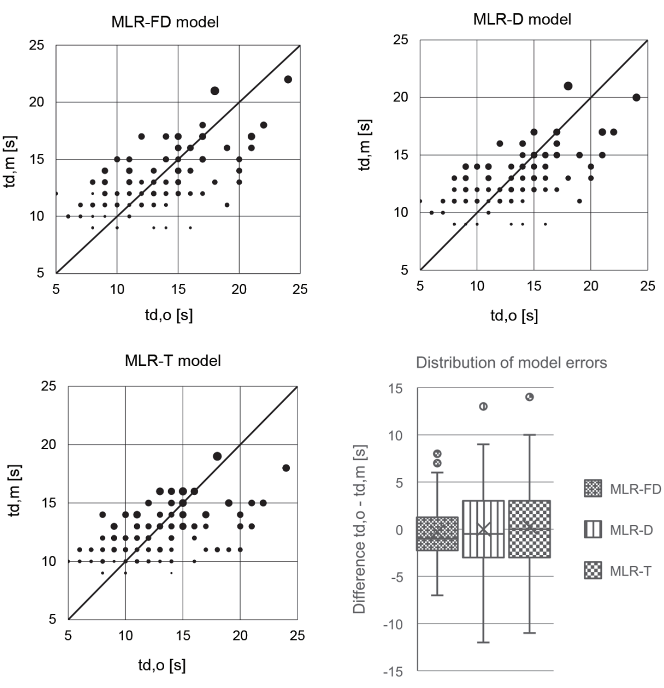

Figure 2 Scatter plots of observed and calculated dwell time values around the identity line (bubble size presents the number of passengers) and a box plot of model errors distribution for the test sample

Table 3 Distribution of MLR model errors

$t_{d,o}$ - $t_{d,m}$ statistics	MLR-FD model	MLR-D model	MLR-T model
Average [s]	-0.3	0.0	0.3
Standard deviation [s]	3.1	4.4	4.5
Minimum [s]	-7	-12	-11
1st Quarter [s]	-2	-3	-3
Median [s]	-1	-1	0
3rd Quarter [s]	1	3	3
Maximum [s]	8	13	14

4 Conclusions

Presented investigation of the tram stop dwell time at the City of Zagreb electric tram network is the first step in reducing the uncertainty of traffic modelling of this specific PT system. Following the analysis of the spatial and traffic characteristics of the entire network, an island tram stop on the southern outskirts of the City of Zagreb centre was identified as an optimal investigation location. The measurements at the selected stop included the field recording of the trams in operation, laboratory processing of collected video data, and creation of a synthesized database on the tram vehicle type, arrival and departure time at the tram stop, and the number of passengers boarding and alighting the tram through a specific door.

To increase the representativeness of the results, statistical analysis and two-step filtering of records collected for the high-floor 4-doors tram vehicle type TMK301 were performed. The filtered tram stop data was divided into training and test sample (in ratio 80:20). The training data sample was used to create three different MLR models for tram dwell time prediction, with the following independent variables: (1) the volume of boarders and alighters and a type of passenger flow transiting through the busiest tram doors, (2) the volume of boarders and alighters transiting through the busiest tram doors, and (3) the total volume of boarders and alighters per tram.

Comparison of the MLR models regression statistics (correlation coefficient, coefficient of determination, and standard error) showed that passenger input data simplification has a minor effect on model's goodness of fit. The model's cross-validation showed that the reduction of complexity of passenger input data by using only the total number of passengers boarding and alighting the tram and neglecting the information on the most intense transit flow through specific tram doors and flow type has a mild effect on model's accuracy and precision. This could be adequately addressed by adding 3 s, as defined by error's interquartile range, to the simpler MLR-D and simplest MLR-T model operating margin.

However, we must emphasize that these results must be taken with caution. Firstly, this investigation assumes that there is a linear relationship between tram stop dwell time and the number of boarders and alighters. Secondly, statistically-based dwell time models are created for a specific TMK301 type tram and are not generally applicable for other types of trams. Thirdly, to develop a more general model, not only vehicle characteristics but island platform conditions should be included in the analysis. For a better understanding of the effects on tram stop dwell time, essential for planning more accurate timetables, reduction of delays, and a higher quality of Zagreb's tram services, further research including these parameters is planned.

References

[1] Li, D., Daamen, W., Goverde, R.M.P.: Estimation of train dwell time at short stops based on track occupation event data: A study at a Dutch railway station, Journal of Advanced Transportation, 50 (2016) 5, pp. 877–896, DOI: 10.1002/atr.1380

[2] Wang, B., Wang, W., Hu, X., Li, X.: A segmented signal progression model for the modern streetcar system, Discrete Dynamics in Nature and Society - Advanced Dynamic Simulations in Transportation, 2015, Article ID 763565, 10 pages, DOI: 10.1155/2015/763565

[3] Altinkaya, M., Zontul, M.: Urban bus arrival time prediction: A review of computational models, IJRTE 2013, 2 (2013) 4, pp. 164–169

[4] Corman, F., Quaglietta, E., Goverde, R.M.P.: Automated real-time railway traffic control: an experimental analysis of reliability, resilience and robustness, Transportation Planning and Technology, 41 (2018) 4, pp. 421–447, DOI: 10.1080/03081060.2018.1453916

[5] Rüger, B.: How platform infrastructure influences passenger behaviour, International Journal for Traffic and Transport Engineering, 8 (2018) 4, pp. 446–454, DOI: 10.7708/ijtte.2018.8(4).04

[6] Lin, T-M., Wilson, N.H.M.: Dwell time relationships for light rail, In Systems Transportation Research Record No. 1361, Light Rail Transit: Planning, Design, and Operating Experience, Sixth National Conference on Light Rail Transit, Calgary, Alberta, Canada, 24–27 May 1992.

[7] Gysin, K.: An investigation of the influences on train dwell time, Master Thesis, Swiss Federal Institute of Technology, ETH, Zurich, Switzerland, 2018.

[8] Highway Capacity Manual (HCM), Transportation Research Board, Washington, DC, United States, 2016.

[9] Fernández, R., Valencia, A., Seriani, S.: On passenger saturation flows in public transport doors, Transportation Research Part A, 2014, 78 (2015), pp. 102–112

[10] Study on the tram traffic development in Zagreb, City of Zagreb, Zagreb electric tram (ZET), University of Zagreb, Faculty of Civil Engineering, Zagreb, Croatia, 2019.

[11] Cvitić, L.: Tram dwell time at stations, Master Thesis, University of Zagreb, Faculty of Civil Engineering, Zagreb, Croatia, 2021.