EVALUATION OF IMPACT OF THE OPERATIONAL AND TECHNICAL FACTORS ON DOWNTIME OF MUNICIPAL BUSES BASED ON A LINEAR REGRESSION MODEL

Joanna Rymarz1,*, Anna Borucka2, Andrzej Niewczas3

1Faculty of Mechanical Engineering, Lublin University of Technology, Lublin, Poland
2Faculty of Logistics and Management Security, Military University of Technology, Warsaw, Poland
3Motor Transport Institute, Warsaw, Poland

*E-mail of corresponding author: j.rymarz@pollub.pl

Resume
The objective of this study was to assess the effect of selected operational and technical factors on downtime of vehicles. The sample consisted of buses from a municipal transport company (Poland). Estimation of parameters of a linear regression model was performed. Month of failure (downtime event) and its type were used as predictors. Failures were divided into three categories: events related to the company’s operations, including vehicle failures (1) and other (organizational) problems (2), as well as failures caused by external factors unrelated to the operations of the transport company (3). The downtime was found to be significantly associated with failure type and month of failure. A linear regression model of downtime with a reduced number of impact factors, taking into account two main failure types and two main periods of their occurrence during the year, was developed.

Article info
Received 18 September 2020
Accepted 16 March 2021
Online 16 September 2021

Keywords:
municipal buses,
downtime,
linear regression,
failure

Available online: https://doi.org/10.26552/com.C.2021.4.A241-A247
ISSN 1335-4205 (print version)
ISSN 2585-7878 (online version)

1 Introduction

In public transport systems, vehicle failures and organizational shortcomings often substantially increase passenger waiting time. In this paper, disablement of a vehicle caused by technical or operational factors has been referred to with the umbrella term “failure”. Vehicle failures are of concern to both drivers and fleet managers. In public transport systems, randomness of bus departure times and travel time has an influence on the quality of transport services [1]. In a situation when compliance with the timetable is the major requirement, the real travel time in the whole transport system is adjusted to a vehicle with the lowest operational speed. In the literature, this phenomenon is known as “bunching”. Bunching forces passengers to arrive early at stations and to budget long travel time [2-5].

The literature describes several corrective strategies to reduce bus bunching. Hickman has proposed a stochastic model of vehicle operations based on recursive equations for expected values of headways and bus loads [6]. His strategy of improving transport services consists in holding operating buses along the service line, in order to regulate the system on an ongoing basis.

Daganzo and other authors have developed a mixed strategy in which passenger boarding and alighting can be limited to improve the regularity of headways [8-10]. Berrebi et. al. have studied the practical effects of implementation of these corrective strategies. They demonstrated that the “bus holding” system reduced bunching, thus also decreasing average passenger waiting time [2]. Adamaki and Turnau have presented a transport system control strategy in which buses were sent at specific times to critical bus stops with high numbers of passengers [11]. However, it is worth mentioning that the “bus holding” method also had negative effects, such as disturbances in traffic flow and an increase in average waiting time [1, 12-18].

An important component of the “bus holding” strategy is prediction of fleet availability. One prediction method involves simulation of readiness based on a regression model developed with use of the retrospective data.

In this paper, a linear regression model is proposed, which links the bus downtime (not-ready time) with the month of the year in which a bus was stationary and type of downtime. A regression analysis of downtime was performed based on data obtained from the municipal transport company in Lublin, Poland. The main objective was to develop a regression model with a reduced number of factors, which could be used to effectively predict bus downtime and ensure continuity of system operation.
status of the vehicles. Unfortunately, the data were not detailed enough to allow to discriminate which failure was caused by which factor. Type B were failures related to events outside the company's control (e.g. collision with another vehicle, freezing of the pneumatic system, vehicle trapped in the snow, a blocked route, an incident inside the vehicle).

### 3 Results

Downtime observation results were divided into "monthly" groups. The number of all the failures recorded in 2018-2019, broken down by month, is given in Table 2. Downtime data for the years 2018-2019, also broken down by month, are shown in Table 3. Monthly downtime duration is presented in Table 2. Figure 1 shows a box plot of downtime per month. As seen in the graph, median values of downtime are different for different months. The highest mean values were recorded in January, September, October and November.

It is worth stressing that the number of failures is different for each month (Table 3). The largest number of failures occurred in March, however, the mean and median downtime values for this month were the lowest (Table 2), which means that the failures were short-term ones.

Significance of differences in downtime between months was assessed using the non-parametric Kruskal-Wallis test.

The Chi-squared statistic was $\chi^2 = 243.98$ and the p-value $< p < 0.0001$, which indicated that, at the significance level $\alpha = 0.05$, the null hypothesis of equality of means was rejected. This demonstrates that there were significant differences between at least two monthly downtime groups.

Another factor that was analysed was the failure type. Three types of most frequent failures (downtime events) were considered. Type A1 were failures related to vehicle damage (e.g. broken/jammed door lock, fluid leakage, broken brakes, no ignition, engine overheating). Type A2 were organizational failures and other technical problems (e.g. a damaged wind shield, mirror, tyre). Some of these failures were related to weather conditions and some to the general status of the vehicles. Unfortunately, the data were not detailed enough to allow to discriminate which failure was caused by which factor. Type B were failures related to events outside the company's control (e.g. collision with another vehicle, freezing of the pneumatic system, vehicle trapped in the snow, a blocked route, an incident inside the vehicle).

### 2 Material and methods

Twenty one buses (8 different makes and models) were studied. The vehicles were between 10 and 16 years old. Observations were conducted in standard public transport conditions over 2 years of operation (2018-2019). The dates and times of bus arrival to and departure from the depot and the vehicle downtime were registered. Source documentation included the company's daily internal reports on the operational status of the fleet. The basic descriptive statistics of the buses are presented in Table 1.

The buses serviced standard routes in municipal traffic. The average monthly mileage was approximately 4637 km. The lowest average monthly mileage of 4313 km was recorded for a bus make 5 (Standard Deviation-SD 1711 km) and the highest mileage of 6041 km for a bus make 2 (SD 1699 km).

Fleet downtime data for the years 2018-2019 were analysed. The impact of two factors, month of failure and failure type, on the dependent variable (downtime) was considered.

Month of failure was analysed repeatedly in each year of observation and was thus an indicator of seasonality related to seasonal changes in weather and vehicle loads (number of passengers) over the year. In winter, many downtime events were caused by door freezing, failures of driver's cabin and passenger compartment heating and power outages. During the summertime, downtime was mainly due to high temperatures, i.e. engine overheating and lack of air conditioning in the vehicle.

The second factor that has been analysed was the failure type. Three types of most frequent failures (downtime events) were considered. Type A1 were failures related to a vehicle damage (e.g. broken/jammed door lock, fluid leakage, broken brakes, no ignition, engine overheating). Type A2 were organizational failures and other technical problems (e.g. a damaged wind shield, mirror, tyre). Some of these failures were related to weather conditions and some to the general status of the vehicles. Unfortunately, the data were not detailed enough to allow to discriminate which failure was caused by which factor. Type B were failures related to events outside the company's control (e.g. collision with another vehicle, freezing of the pneumatic system, vehicle trapped in the snow, a blocked route, an incident inside the vehicle).
The number of type A1 failures was similar (1136 events). The lowest number of failures were the B type events (596 events). Descriptive statistics of downtime for each type of failure is presented in Table 4 and a box-plot of downtime versus failure type is shown in Figure 2.

As Figure 2 shows, downtime duration differed significantly depending on the type of failure. The longest downtimes (though the smallest in number) were caused by the B type failures, which were outside the company's control. The shortest stoppages were related to A1 type events associated with repair of subsystems (mechanisms) or scheduled maintenance of vehicles.
studied category. Variables with extreme average values were selected: type A1 failure and the month of April. In this way, the effect of single-signedness of the remaining parameters with regards to the level of the omitted variable was obtained.

A statistical analysis was conducted to determine the structure of the linear regression model with binary variables, which resulted from the quantitative nature of the dependent variable and the qualitative character of the independent variables. The estimated model parameters are given in Table 5.

Four of the estimated parameters were statistically insignificant. The AIC (Akaike Information Criterion) = 34011 and corrected $R^2 = 48\%$ indicate that the model does not fully explain the observed phenomena. This is also indicated by the residual distribution, which is different from the normal distribution (Lilliefors test statistics $D = 0.088$ and p-value < 0.001). Additionally, an analysis of the autocorrelation function (Figure 3) demonstrated significant dependencies not described by the model. This means that bus downtime is dependent on factors which have not been included in the model.

The regression equation is given by:

$$y = 18.4+36.1 \times A2 + 137.7 \times January - 216 \times February - 13.6 \times March + 3.3 \times May + 1.8 \times July - 3.8 \times June + 23.1 \times August + 31.6 \times September + 30.0 \times October + 32.1 \times November + 2.5 \times December + \varepsilon.$$  

(2)

### Table 4 Descriptive statistics of downtime distribution for different types of failures

| type of failure | median (minutes) | mean (minutes) | standard deviation (minutes) | min. (minutes) | max. (minutes) |
|----------------|------------------|----------------|------------------------------|----------------|----------------|
| A1             | 10               | 23.8           | 26                           | 1              | 119            |
| A2             | 63               | 64.1           | 27.1                         | 10             | 132            |
| B              | 134              | 167            | 124                          | 23             | 399            |

### Figure 2 The box plot of bus downtime versus type of failure
is that the differences are statistically significant. The test statistic is given by:

$$W = \sum_{i=1}^{N} \{ sgn(x_2 - x_1) \times R_i \},$$

(4)

where $sgn$ is the sign function, $R_i = \sum_j R_{ij}$, $R_{ij}$ is the rank of observation, $x_1$, $x_2$ are study groups, $N$ is sample size (number of study groups). Test results are presented in Table 6.

Based on the results, three groups of months were selected for which the downtime distributions were not significantly different. Additionally, the lack of significance of differences in each group was confirmed using the Kruskal-Wallis test. The first group of months (group I) included August, September, October and November. For this group, the chi-squared statistics $\chi^2 = 1.245$ and p-value = 0.742. The second group (group II) included February, March, April, May, June, July

Table 5 Parameters of the linear regression model and evaluation of their significance

| parameter      | estimate $\beta_k$ | std. err or $S(\beta_k)$ | t value | p-value |
|----------------|--------------------|--------------------------|---------|---------|
| $\beta_0$      | 18.4               | 3.379                    | 5.432   | < 0.001 |
| failure A2     | 36.1               | 2.300                    | 15.672  | < 0.001 |
| failure B      | 137.7              | 2.907                    | 47.390  | < 0.001 |
| January        | 29.6               | 4.472                    | 6.621   | < 0.001 |
| February       | -21.6              | 4.357                    | -3.112  | 0.002   |
| March          | -13.6              | 4.357                    | -3.112  | 0.002   |
| May            | 3.3                | 4.357                    | 0.668   | 0.504   |
| July           | 1.8                | 4.357                    | 0.336   | 0.737   |
| June           | -3.8               | 4.357                    | -0.761  | 0.447   |
| August         | 23.1               | 5.341                    | 4.320   | < 0.001 |
| September      | 31.6               | 4.806                    | 6.585   | < 0.001 |
| October        | 30.0               | 4.806                    | 6.184   | < 0.001 |
| November       | 32.1               | 4.755                    | 6.759   | < 0.001 |
| December       | 2.5                | 5.182                    | 0.474   | 0.636   |

where Std. Err or $S(\beta_k)$ explains the accuracy of the parameter estimate ($\beta_k$). Indicates by how many units the assessment value (estimated) differs from the actual value of parameter $\beta_k$.

The quality of fit of the regression model was evaluated using AIC (Figure 3). The value of AIC was found from the following equation:

$$AIC = -2 \ln L + 2k,$$

(3)

where $k$ is the number of model parameters and $L$ is the reliability function.

In accordance with the objective of the study, in the next stage of the calculations, the model was simplified. To limit the number of predictors, it was proposed that months with similar regression coefficients should be aggregated. Statistically similar months were grouped with Pairwise Wilcoxon Rank Sum Test, which is a non-parametric test with multi-testing correction used to compare pairs in groups. The null hypothesis for Pairwise Wilcoxon Rank Sum Test is that there are no differences between distributions, while the alternative hypothesis is that the differences are statistically significant. The test statistic is given by:

$$W = \sum_{i=1}^{N} \{ sgn(x_2 - x_1) \times R_i \},$$

(4)

where $sgn$ is the sign function, $R_i = \sum_j R_{ij}$, $R_{ij}$ is the rank of observation, $x_1$, $x_2$ are study groups, $N$ is sample size (number of study groups). Test results are presented in Table 6.

Based on the results, three groups of months were selected for which the downtime distributions were not significantly different. Additionally, the lack of significance of differences in each group was confirmed using the Kruskal-Wallis test. The first group of months (group I) included August, September, October and November. For this group, the chi-squared statistics $\chi^2 = 1.245$ and p-value = 0.742. The second group (group II) included February, March, April, May, June, July...
seasonality over the year and the effect of organizational and technical factors (type of failure) on bus downtime. The model also permits to predict the availability of a transportation system as part of the strategy of ensuring the continuity of transportation services, e.g. by introducing the “bus holding” control strategy.

From among the three types of downtime events, considered as independent variables, the model includes type B failures (events outside the company’s control) as the dominant type and type A2 (operational and organizational) failures (which have four times less impact than type B events).

Among the selected month groups, the reduced model presents the summer-autumn season, including August, September, October and November, as well as the winter season, which is represented by a single month - January. The effects of the two seasons on downtime duration are comparable and similar to the impact of A2 type failures.

Acknowledgement

The project/research was financed as part of the Lublin University of Technology’s project- Regional Excellence Initiative, funded by the Polish Ministry of Science and Higher Education (contract no. 030/ RID/2018/19).

5 Summary

Based on a study of municipal bus operations, a linear multi-regression model of downtime, as a function of selected groups of months of the year and type of downtime event (failure), was developed.

This model allows to determine the impact of climate and December. For this group $\chi^2 = 10.878$ and p-value = 0.092. The last group consisted of only one month, January, for which no goodness of fit with any other group was observed. The parameters of the estimated model are presented in Table 7.

The final form of the model was the following:

$$y = 12 + 36.8 \times A2 + 135.5 \times B + 36.2 \times grI + 36.3 \times gr III.$$  

(5)

All the model parameters were statistically significant. Corrected $R^2 = 48\%$ and AIC = 34045.64. The values of the parameters describing the quality of the regression model did not differ significantly from the basic formula given by Equation (2). The reduction of the number of factors, achieved by their aggregation in the manner presented in this paper, did not reduce the quality of the initial regression model.
References

[1] SAVSAR, M. Modeling and simulation of maintenance operations at Kuwait public transport company. *Kuwait Journal of Science* [online]. 2013, 40(2), p. 115-129. ISSN 2307-4108, eISSN 2307-4116. Available from: https://journals.kuwait/kjs/index.php/KJS/article/view/120

[2] BERREBI, S. J., WATKINS, K. E., LAVAL, J. A. A real-time bus dispatching policy to minimize passenger wait on a high frequency route. *Transportation Research Part B: Methodological* [online]. 2015, 81(2), p. 377-389. ISSN 0191-2615. Available from: https://doi.org/10.1016/j.trb.2015.05.012

[3] BORUCKA, A., NIEWCZAS, A., HASILOVA, K. Forecasting the readiness of special vehicles using the semi-Markov model. *Eksplotacja i Niezawodnosc - Maintenance and Reliability* [online]. 2019, 21(4), p. 662-669. ISSN 1507-2711. Available from: http://dx.doi.org/10.17531/ein.2019.4.16

[4] CABAN J., DROZDZIEL, P., KRZYWONOS, L., RYBICKA, I., SARKAN, B., VRABEL, J. Statistical analyses of selected maintenance parameters of vehicles of road transport companies. *Advances in Science and Technology Research Journal* [online]. 2019, 13(1), p. 1-13. ISSN 2299-8624. Available from: https://doi.org/10.12913/22998624/92106

[5] CHERNOYAROV, O. V., KUTOYANTS, Y. A., MARCOKOVA, M. On frequency estimation for partially observed system with small noises and observation equations. *Communications - Scientific Letters of the University of Zilina* [online]. 2018, 20(1), p. 67-72. ISSN 1335-4205, eISSN 2585-7878. Available from: http://komunikacie.uniza.sk/index.php/communications/article/view/48

[6] HICKMAN, M. D. An analytic stochastic model for the transit vehicle holding problem. *Transportation Science* [online]. 2001, 35(3), p. 215-237. ISSN 0041-1655, eISSN 1526-5447. Available from: https://doi.org/10.1287/trsc.35.3.215.10150

[7] LALANNE, C., MESBAH, M. Measures of association, comparisons of means and proportions for two samples or more. In: *Biostatistics and Computer-based Analysis of Health Data using Stata*. 1. ed. ISTE Press - Elsevier, 2016. ISBN 9781785481420, eISBN 9780081010846, p. 25-57.

[8] DAGANZO, C. F. A headway-based approach to eliminate bus bunching: Systematic analysis and comparisons. *Transportation Research Part B: Methodological* [online]. 2009, 43, p. 913-921. ISSN 0191-2615. Available from: https://doi.org/10.1016/j.trb.2009.04.002

[9] EDWARD, K., BEATA, K., DARIUSZ, K., DARIUSZ, M. Survival function in the analysis of the factors influencing the reliability of water wells operation. *Water Resources Management* [online]. 2019, 33(14), p. 4909-4921. ISSN 0920-4741, eISSN 1573-1650. Available from: https://doi.org/10.1007/s11269-019-02419-0

[10] MORALES, D., MUNOZ, J. C., GAZMURI, P. A stochastic model for bus injection in a public transport service. *Transportation Research Procedia* [online]. 2019, 38, p. 688-708. ISSN 2352-1465. Available from: https://doi.org/10.1016/j.trpro.2019.05.036

[11] ADAMSKI, A., TURNAU, A. Simulation support tool for real-time dispatching control in public transport. *Transportation Research Part A: Policy and Practice* [online]. 1998, 32(2), p. 73-87. ISSN 0965-8564. Available from: https://doi.org/10.1016/S0965-8564(97)00019-0

[12] NIEWCZAS, A., MORAWSKI, L., DEBICKA, E., BORUCKA, A. Commercial vehicles incapacity risk analysis in the transport company. *Journal of KONBiN* [online]. 2019, 49(4), p. 321-340. ISSN 2083-4608, eISSN 2083-4608. Available from: https://doi.org/10.2478/jok-2019-0089

[13] NIEWCZAS, A., MORAWSKI, L., DEBICKA, E., RYMARZ, J. Predicting costs of city buses’ incapacity risk. *Journal of KONBiN* [online]. 2019, 49(1) p. 425-448. ISSN 2083-4608, eISSN 2083-4608. Available from: https://doi.org/10.2478/jok-2019-0022

[14] NIEWCZAS, A., RYMARZ, J., DEBICKA, E. Stages of operating vehicles with respect to operational efficiency using city buses as an example. *Eksplotacja i Niezawodnosc - Maintenance and Reliability* [online]. 2019, 21(1), p. 21-27. ISSN 1507-2711. Available from: http://dx.doi.org/10.17531/ein.2019.1.3

[15] RYMARZ, J., NIEWCZAS, A., STOKLOSA, J. Reliability evaluation of the city transport buses under actual conditions. *Transport and Telecommunication Journal* [online]. 2015, 16(4), p. 259-266. ISSN 1407-6179, eISSN 1407-6179. Available from: https://doi.org/10.1515/ttj-2015-0023

[16] SANCHEZ, S. A. Optimizing performance-based mechanisms in road management: an agency theory approach. *European Journal of Transport and Infrastructure Research* [online]. 2015, 15(4), p. 465-481. ISSN 1567-7141. Available from: https://doi.org/10.18757/ejtir.2015.15.4.3092

[17] ZAWIERZCHOWSKA, M. Evaluation of the quality of public transport services - selection of variables to be tested by the Delphic method. *Autobusy*. 2018, 19(5), p. 18-21. ISSN 1509-5878, eISSN 2450-7725.

[18] ZHAO, F., ZENG, X. Optimization of transit route network, vehicle headways and timetables for large-scale transit networks. *European Journal of Operational Research* [online]. 2008, 186(2), p. 841-855. ISSN 0377-2217. Available from: https://doi.org/10.1016/j.ejor.2007.02.005