JOURNAL OF MODERN MANUFACTURING SYSTEMS AND TECHNOLOGY (JMMST)
e-ISSN: 2636-9575
VOL. 6, ISSUE 1, 66 – 72
DOI: https://doi.org/10.15282/jmmst.v6i1.7431

ORIGINAL ARTICLE

An Investigation of Travel Time Pattern of T224 RapidKL Bus Using AVL Data

Siti Nur Athirah Ahmad Latfi and Nur Najmiyah Jaafar*

Faculty of Manufacturing and Mechatronics Engineering Technology, Universiti Malaysia Pahang, 26600 Pahang, Malaysia.

ABSTRACT – Transit dependability is a measure of the quality of service provided by public transportation networks. The trip time distribution may represent the nature and pattern of travel time variability, and it is required in the reliability analysis. Predicting transit bus travel times along routes is essential in bus design and operation, particularly in metropolitan regions. Bus riders are more likely to trust a transportation system if journey times can be accurately anticipated within a given margin of error. Many studies have been conducted to better understand travel time distribution, however there are very few studies on heterogeneous traffic in emerging countries such as Malaysia. In this work, the Taguchi T- technique and linear regression scales were used to evaluate the journey time distribution of public bus route T224 utilising AVL data. To increase the accuracy of applications, a prediction mechanism should be developed. Finally, a formulation was constructed to test the influence of side roads on prediction accuracy, and it was discovered that the additional need in terms of location-based data had no discernible effect on prediction accuracy. This clearly proved that the suggested method based on vehicle tracking data is adequate for the contemplated use of bus trip time prediction.

ARTICLE HISTORY
Received: 7th February 2022
Revised: 22nd February 2022
Accepted: 30th March 2022

KEYWORDS
Predictive model
Time series data
Taguchi T-Method
Travel time pattern

INTRODUCTION

The reliability of bus services will have a huge effect on service providers and current and future customers. Passengers are influenced by inadequate infrastructure expectations of service efficiency and bus service relative to other means of choice, while transit systems are affected by lack of riding and sales and increased costs of delivering extra service to cover for weak service operations [1]. Travel time reliability can take several quantitative steps, all of which vary to some degree in the details they contain when determining the reliability of the path. Bus travel time calculation may allow transit operators to understand and enhance the efficiency of their services and draw more public transport users.

Recent developments in automated vehicle location (AVL) systems based on the Global Positioning System (GPS) provided the transit industry and public transport companies with tools to track and coordinate the movement of their vehicles and maintain their fleet reliably and cost-effectively. The benefit of bus automated vehicle location (AVL) data is well known, ranging from passenger-facing applications that forecast bus arrival times to service-facing applications that track network performance and identify performance failures [2]. While these techniques are sufficient for comparatively limited datasets, they are typically not feasible to classify and derive useful findings from far larger datasets derived from AVL schemes. It is also important to collect and summarise the data in a manner that promotes efficient decision-making and helps to concentrate resources.

In the bus industry, AVL systems have been used for many years but there has been no use of the historical data provided by these systems so far. The purpose of the data can be used in a wide range such as timetabling. As the AVL data is a historical archive of how the path currently operates on a day to day, this information can be used to locate locations along the route where buses continually struggle to follow the timetable or periods in the day where length gaps in service often occur [3]. A better schedule will then be drawn up and checked and it is easy to assess the effects of the adjustments to the schedule. The key role of the AVL system is typically to convey real-time information to route controllers about current bus positions to allow them to overcome more effectively with service disturbances. As this data provides a continuous record of the movement of the bus during the day it offers a valuable source of data from which bus speed can be collected.

RELATED WORK

The AVL data for the T224 bus given by RapidKL is the main dataset of this study. This chapter presents the data of the RapidKL T224 bus AVL not only to remind the analyses below but also to put the data in the broad sense of human mobility science. Latest advances in automated vehicle positioning (AVL) technologies based on the Global Positioning System (GPS) have equipped the bus industry and public transport agencies with tools to track and coordinate the operation of their vehicles and maintain their fleet reliably and cost-effectively. There is a component of AVL Data which is the time when the vehicle hits the destination location or the location where the vehicle will be at a given point in time.

*CORRESPONDING AUTHOR | Nur Najmiyah Jaafar | najmiyah@ump.edu.my

RapidKL route T224, which connects LRT Dato Keramat to Desa Pandan, was selected in the current analysis. The study length of 7 km connects LRT Dato Keramat to Desa Pandan to LRT Dato Keramat, which includes 14 bus stops from 300 m to 1km. This route covers various types of urban roads with various characteristics of traffic, geometry, and land use. In the study, the weekdays and weekend data of October 2018 with service operating from 6.00 am to 11.00 pm was used. Details on the scheduled date, schedule ID, route ID, trip ID, latitude, longitude, and GPS time stamp used in the Automatic Vehicle Location Data were changed every 1 to 20 seconds.

To estimate the arrival time of the bus, AVL data must be reduced and altered. The current position, current time, speed, direction and vehicle ID are all part of the AVL data. These raw data were used to extract the input data for predicting bus arrival time.

The Taguchi's T-Method is a branch of the Mahalanobis Taguchi System (MTS) developed by the renowned Japanese Quality Guru Dr. Genichi Taguchi [4]. MTS was introduced in the 1980s, as a pattern information technology that aids the decision-making process by constructing a multivariate measurement scale using data analytic methods [5]. Taguchi's T-method (T-method) is a regression-driven predictive model developed by Genichi Taguchi as part of the Mahalanobis-Taguchi System to forecast the future state or uncertain performance based on current or historical evidence[6]. The T-method, like any regression-based approach, calculated and clarified the relationship between the response variable and the explanatory variable in the context of a statistical model to produce a reliable forecast. T-method is a comparatively recent method, with just a few academic works in the literature focusing on its implementation, contrast, and improvement.

DATA ANALYSIS SIMULATED BY TAGUCHI T'METHOD

Simulated Taguchi Method

The T Method idea is used with the S/N Ratio in the T-Method estimate approach (SNR). The T-concept Method remains focused on three critical characteristics of SNR: uncertainty, adaptability, and linearity [7]. Taking into account the complex world, some value has been applied to the theory established. T-Method used the notion of reference-point proportional equation to build a regression that passes through the origin to make a graph. The unit and signal group are used to design and evaluate the forecasting model. For normalisation, the average value of the unit group is subtracted from each member of the signal group.

Computation of the proportional coefficient (β) and SN ratio (η) is done to determine the items (parameters) that would be used for prediction and estimation. The T-Method computes the proportional coefficient (β) and the SN ratio (η) between the output (M_i) and item value. The proportional coefficient (β) and SN ratio (η) were computed item by item using normalised data (X_{ij}) and normalised performance value (M_i). For each component, the SN ratio and proportional coefficient are calculated. If the SN ratio is negative, the sum would be considered zero. The integrated approximate output value for Signal Data is calculated item by item using the proportional coefficient β and SN ratio η. The larger an item's SN ratio, the greater its contribution to general performance calculation. The following formula can be used to calculate the integrated approximation performance value M_i:

$$\hat{M}_i = \frac{\eta_1 \times \left(\frac{X_{i1}}{\beta_1}\right) + \eta_2 \times \left(\frac{X_{i2}}{\beta_2}\right) + \cdots . + \eta_j \times \left(\frac{X_{ij}}{\beta_j}\right)}{\eta_1 + \eta_2 + \cdots + \eta_j} \tag{1}$$

RESULTS AND DISCUSSION

Weekday and Weekend Data

The analysis described below was performed by combining automatic vehicle location (AVL) data of T224 bus lines in the area of LRT Dato Keramat to Desa Pandan. To begin the study, it is necessary to determine peak and off-peak hours of the day. Figure 1 plots the observed travel times in relation to the travel time of day on working days from Monday to Friday. The weekday data has been sorted by daily travel time which is 30 minutes of the travel time. For example, 6.00 a.m to 6.30 a.m a day for 1 month. It was discovered that the variability in the weekday non-peak period is greater than the variability in the weekday peak period, and the weekday peak period has the least fluctuation in arrival time among all periods.

Figure 2 plots the observed travel times concerning the travel time of the day during the weekend day (Saturday and Sunday). Nonetheless, the weekend statistics reveal that the weight of the significance was not great, implying that there is a lower likelihood of traffic congestion happening on weekends. As can be observed, travel time on weekends is not significantly longer than travel time on weekdays. Weekend vehicle traffic appears to be lower throughout all peak hours as compared to weekdays. In particular, for the investigated bus lines, the variance is higher in the afternoon peak hour than in the morning and evening. Based on the figure, the weekend peak period was in the middle of the graph and it was observed to have high variability compared to the other time data. The graph shows the time travel for non-stationary data varies with time and is not constant.

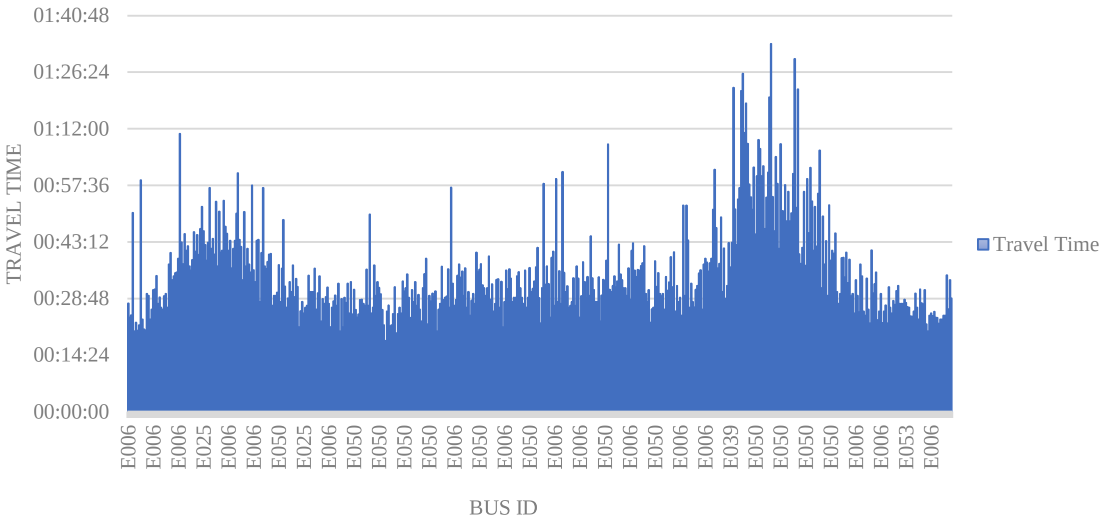

Figure 1. Weekday Data

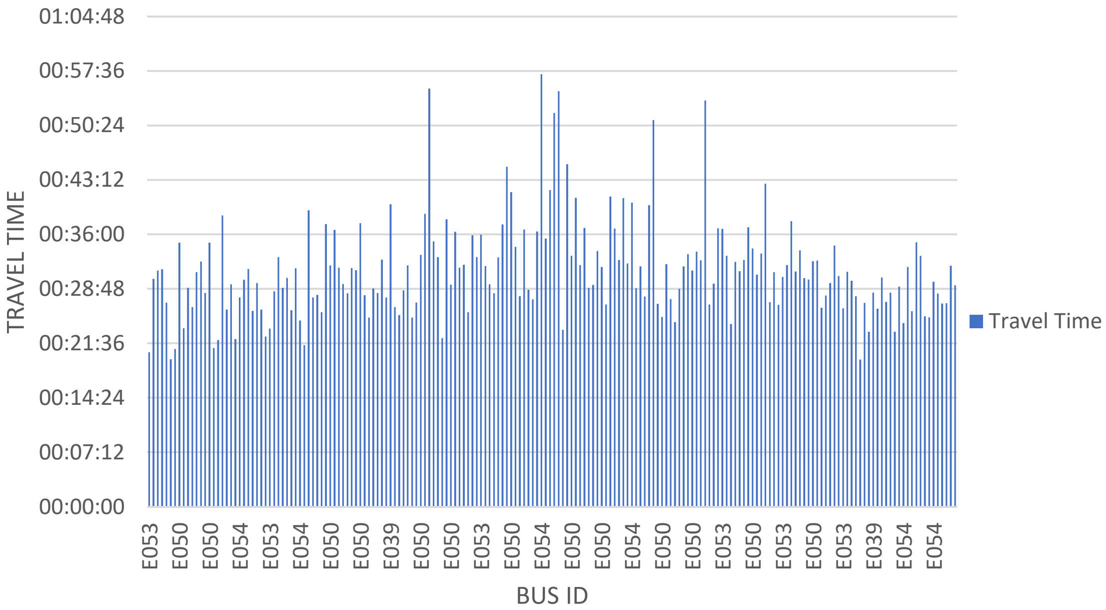

Figure 2. Weekend Data

Morning peak, Evening peak and Non-peak data

The collected data were analysed to comprehend the fluctuation in travel time within a day of the morning peak. The morning peak hour time was chosen for testing and verifying the simulation model since it has the largest passenger demand and the influence of strategies on service dependability would be greatest. Figure 3 illustrates the travel time over bus ID of the day. The data were collected from 6.30 a.m until 10.00 a.m. The data was derived using the average journey time of bus rides to approximate typical travel times. These average travel times could be calculated separately for periods and bus trips. According to the figure below, the maximum travel time relative to other times may be seen in the center of this graph. This is because there is a possibility of traffic jams on the road. After all, that time is peak hour.

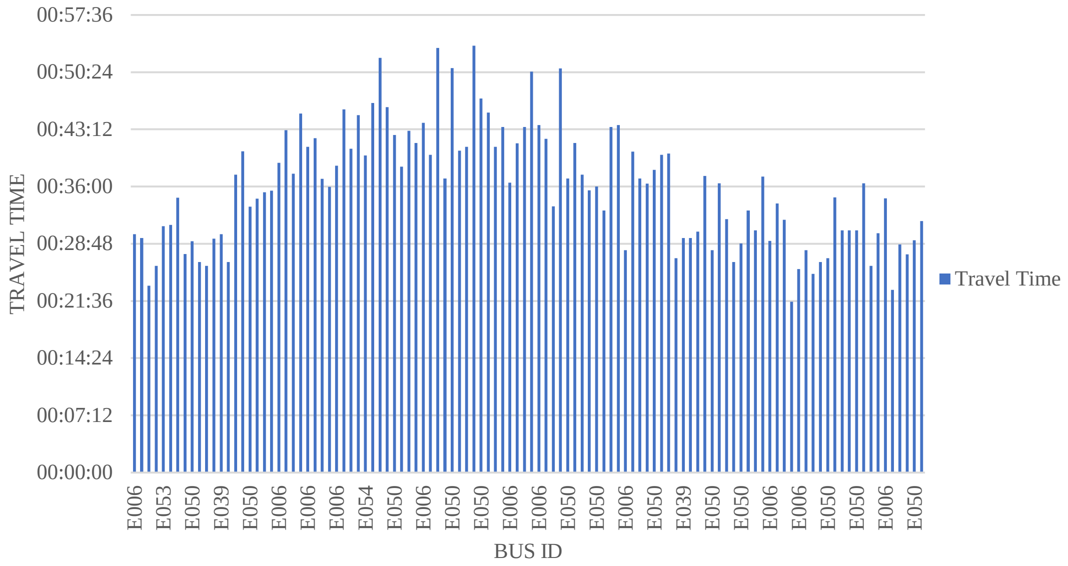

Figure 3. Morning Peak Data

The obtained data were examined in order to understand the variation in travel time within a day of the evening peak. The evening peak hour was chosen for testing and evaluating the simulation model because it has the highest passenger demand and the largest impact of tactics on service reliability. Figure 4 depicts the day's journey time over bus ID. The information was gathered between 5.00 p.m and 8.00 p.m. The data was derived using the average journey time of bus rides to approximate typical travel times. These average journey times might be computed individually for different periods and bus trips. According to the graphic below, the longest journey time relative to other periods may be seen in the middle of the graph. This is because there is a chance of traffic congestion on the road.

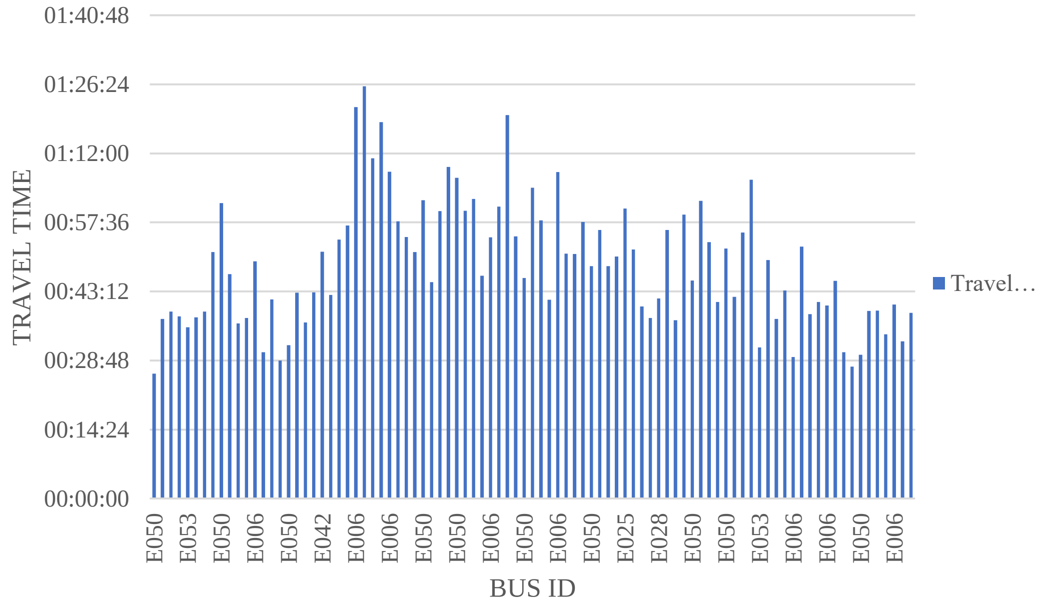

Figure 4. Evening Peak Data

The obtained data were analysed to better understand the variation in travel time during a day of non-peak. The non-peak hour was chosen for testing and assessing the simulation model because it has the highest passenger demand and hence has the greatest influence on service reliability. Figure 5 depicts the day's journey time over bus ID. The information was gathered the time other than morning and evening peaks. To approximate normal travel times, the data was obtained using the average journey time of bus journeys. These average journey times might be calculated separately for various dates and bus rides.

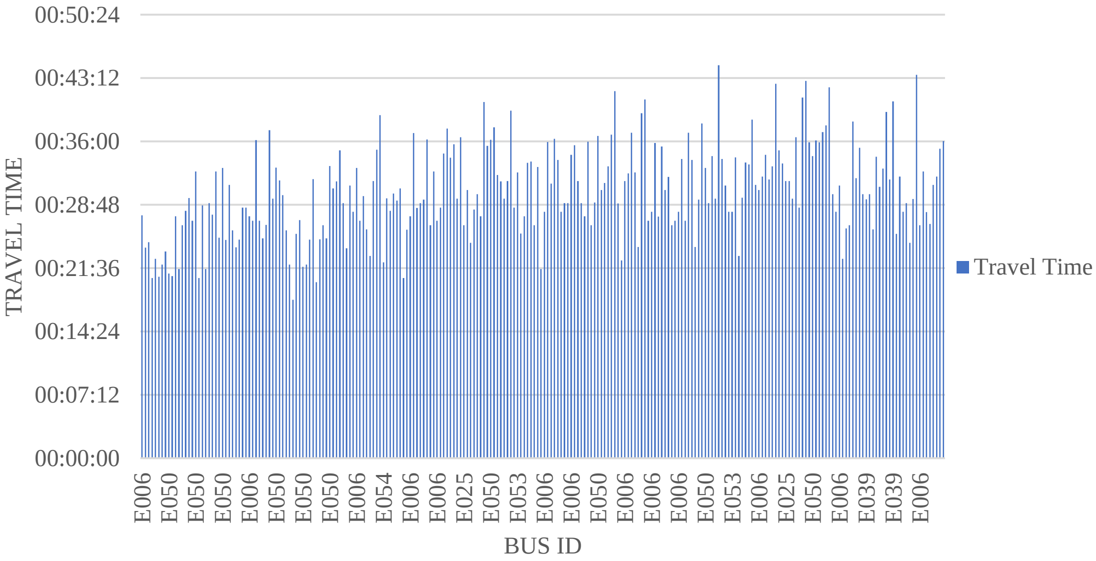

Figure 5. Non-Peak Data

Travel Time Pattern Analysis

Multilinear models of regression are developed in this area. The input variables should not be interrelated in the development of regression models. In this study, the input variables were the average traveling duration and average speed. Recognizing that external influences or other variables might change the time of the bus arrival. In such a scenario the start time, with input or independently determining variables like speed, road conditions, and other potential variables is calculated using a mathematical formula.

After the development of the model, it is necessary to estimate the performance in terms of the accuracy of the data. Figures 5 and 6 show the comparison between data analysis in the morning and evening. The data analysis trip time was compared to check the validity of the models. Figures 5 and 6 show the scatter plots of models for the testing phase. To check the performance of the developed model two measures of effectiveness were used that is MAE (Mean Absolute Error), and R^2 (coefficient of correlation). MAE has defined the sum of absolute differences between our target and predicted variables.

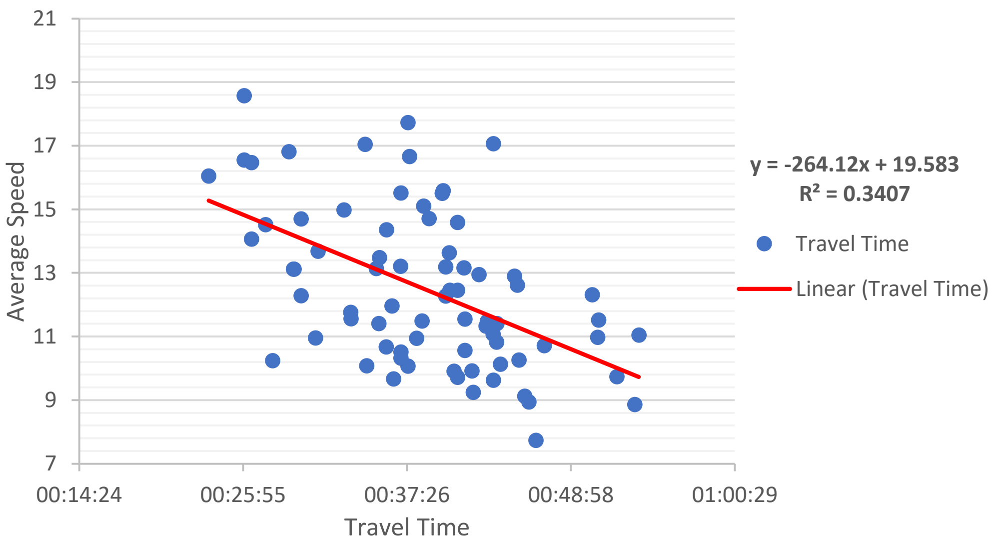

Figure 5. Regression of Morning Peak

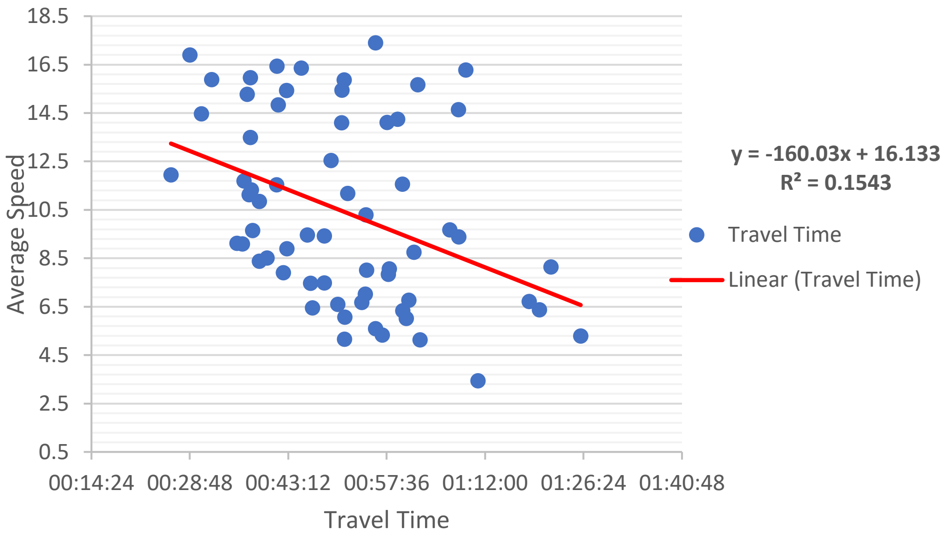

Figure 6. Regression of Evening Peak

Table 1. Result of Mean Absolute Error (MAE) and Coefficient of Correlation (R^2)

Peak Hour	Data Training
Morning	$R^2 =$ 0.3407 MAE = 0.0036
Evening	$R^2 = 0.1543$ MAE: 0.0091

Table 1 displays the MAE and R^2 for both morning and evening data training models. The MAE and R^2 value is obtained from the T-method by using Matlab. The R squared value ranges between 0 and 1, with 0 indicating that the model does not match the given data and 1 indicating that the model fits the data correctly. The R^2 on the evening data is 15% of the variance while the one on the morning data is 34%. As we can see, the r square is low, but data with high variability might have a strong trend. The trend demonstrates that even when data points are further away from the regression line, the predictor variable still gives information about the answer.

The MAE is the total of the absolute differences between our goal and the expected variables. The result of MAE is to measure, without considering direction, the mean size of the errors in a data collection and the value can be range from 0 to ∞. The lower the MSE the higher the accuracy of prediction as there would be an excellent match between the actual and predicted data set. It returns a linear number that equalises the weighted individual differences. The model's performance improves as the value decreases.

CONCLUSION

In this study data from buses have been gathered to do pattern analysis to find the most important travel times for every day using the T-method. The study examined data patterns of public transit bus journey time under diverse traffic circumstances by the weekend, weekday, morning peak, evening peak, and non-peak clustering. The developed T-method and regression model were time discretized and captured the change in trip time in a subsection over time. The MAE (Mean Absolute Error) values for the morning and evening peaks were 0.0036 and 0.0091, respectively. The R^2 value is 34% in the morning and 15% in the evening. The absolute error is defined as the absolute value of the difference between the projected and actual values.

ACKNOWLEDGEMENT

The author would like to thank Puan Najmiyah Jaafar for supervising this work and sharing his expertise in improvements for this research.

REFERENCES

[1] Kumar, B.A., Vanajakshi, L. and Subramanian, S.C., 2017. Bus travel time prediction using a time-space discretization approach. Transportation Research Part C: Emerging Technologies, 79, pp.308-332.
[2] Furth, P.G., Hemily, B., Muller, T. and Strathman, J.G., 2003. Uses of archived AVL-APC data to improve transit performance and management: Review and potential. TCRP Web Document, 23.
[3] Khadhir, A., Anil Kumar, B. and Vanajakshi, L.D., 2021. Analysis of global positioning system based bus travel time data

and its use for advanced public transportation system applications. Journal of Intelligent Transportation Systems, 25(1), pp.58-76.
[4] Taguchi, G., 1995. Quality engineering (Taguchi methods) for the development of electronic circuit technology. IEEE Transactions on Reliability, 44(2), pp.225-229.
[5] Ramlie, F., Jamaludin, K.R., Dolah, R. and Muhamad, W.Z.A.W., 2016. Optimal feature selection of taguchi character recognition in the mahalanobis-taguchi system using bees algorithm. Global Journal of Pure and Applied Mathematics, 12(3), pp.2651-2671.
[6] Negishi, S., Morimoto, Y., Takayama, S. and Ishigame, A., 2017. Daily peak load forecasting by Taguchi's T method. Electrical Engineering in Japan, 201(1), pp.57-65.
[7] Matsui, S. and Nagata, Y., 2019. Proposal for a univariate time series analysis method based on Taguchi’s T-method. Total Quality Science, 5(1), pp.1-10.