ARIMA MODELS FOR BUS TRAVEL TIME PREDICTION

(Date received: 6.3.2009)

Suwardo[1], Madzlan Napiah[2] and Ibrahim Kamaruddin[3]
[1]Civil Engineering Department, Faculty of Engineering,
Gadjah Mada University, Yogyakarta 55281, Indonesia
[2,3]Civil Engineering Department, Universiti Teknologi PETRONAS,
Bandar Seri Iskandar, 31750 Tronoh, Perak
E-mail: [1]suwardo@yahoo.com

ABSTRACT

In this paper, the time series model, Autoregressive Integrated Moving Average (ARIMA) is used to predict bus travel time. ARIMA model is simpler used for predicting bus travel time based on travel time series data (historic data) compared to regression method as the factors affecting bus travel time are not available in detail such as delay at link, bus stop, intersection, etc. Bus travel time prediction is an important aspect to bus operator in providing timetable for bus operation management and user information. The study aims at finding appropriate time series model for predicting bus travel time by evaluating the minimum of mean absolute relative error (MARE) and mean absolute percentage prediction error (MAPPE). In this case, data set was collected from the bus service operated on a divided 4-lane 2-way highway in Ipoh-Lumut corridor, Perak, Malaysia. The estimated parameters, appropriate model, and measures of model performance evaluation are presented. The analysis of both Ipoh to Lumut and Lumut to Ipoh directions is separately performed. The results show that the predicted travel times by using the moving average, MA(2) and MA(1) model, clearly fit with the observed values for both directions, respectively. These appropriate models are indicated by the minimum MARE and MAPPE values among the tentative models. It is concluded that MA(2) and MA(1) models are able to be appropriately applied in this case, and those models can be used for bus travel time prediction which helping in the timetable design or setup.

Keywords: *Autoregressive Integrated Moving Average, Bus Travel Time, Mean Absolute Percentage Prediction Error, Mean Absolute Relative Error, Time Series Model*

1.0 INTRODUCTION

Bus travel time is an important attributes of trip that many people asking for their traveling purpose by using bus service. Generally, people consider travel time as well as other factors such as ticket fare, bus facilities, comfort, safety, *etc*. They prefer to minimize time spent for moving from a place to other place by choosing the best bus service. Therefore, the transport planner needs to elaborate the importance of bus travel time to improve the performance of bus service.

An important aspect of quality of bus service is service reliability. Reliability is used to ensure that the bus service will depart and arrive on time at certain distance accordingly. Bus arrival time is affected by departure time, travel time and delay, therefore, travel time is important factor of bus service reliability. The reliability is improved further for attracting more passengers use bus. Then, bus travel time series data is compiled for predicting bus travel time. Autoregressive Integrated Moving Average (ARIMA) model is simpler kind of time series data analysis than the regression method for prediction while the travel time series data is available.

Time series data was recorded in term of total bus travel time from start to end terminal which including delay (at link, bus stop, intersection, incident, *etc*.). In this case, delay is not separated in detail from the total time. Thus, ARIMA was applied to predict bus travel time based on travel time series data (historic data). That is the reason to apply ARIMA model in predicting bus travel time.

Based on explanation above, ARIMA model is simpler used for predicting bus travel time based on travel time series data (historic data) than regression method as the factors affecting bus travel time are not available in detail such as delay at link, bus stop, intersection and incident location.

2.0 OBJECTIVE AND SCOPE OF STUDY

The objective of this model is to predict the travel time of bus which running from end-to-end bus terminal based on the time series data or historic data of travel time. For further, transportation engineer/planner will take the benefit of this study in the timetable design or setup.

The time series data of bus travel time were collected starting from January to December 2007 for two typical days a month, workday and weekend. The data collection was performed during the running period of a two way bus trip as named on board survey. This paper focuses on deriving and discussing model of bus travel time prediction by using ARIMA model. Model development stages is described in flow chart and followed with model validation (See sub-section 5.2).

3.0 PREVIOUS STUDY ON BUS TRAVEL TIME

The minimum total travel time spent by users is the one of typical users' objectives in using transit service among other measures such as personal accessibility, excess time, congestion, total travel cost, direct travel cost, coverage, traveler satisfaction, accident rate, and criminal activity [1-2].

With the development of Intelligence Transportation System (ITS), the concept of providing users with reliable information about bus arrival time at bus stops has emerged as Advanced Traveler Information Systems (ATIS) and Advanced Public Transportation Systems (APTS). Some of information for users included bus travel time, bus location, speed, passengers on board and dwell time. The information collected by agencies (manually recorded or real time provided) becomes historical information and assists transit agencies with planning, management, and control of the system as well as the improvement of service.

The primary objective of APTS, relating to transit passengers, is to improve the distribution of information about the public transport system such as, travel time, delay, and vehicle position [3]. Hence, the prediction of bus arrival time at bus stop is found to be valuable information for passengers to help reduce their waiting time.

The importance of bus arrival time prediction models as a component of ATIS has been well argued by numerous researchers [4-6]. Bus arrival prediction has become a critical component of the systems [7]. In the prediction of arrival time, the travel time is a part of system as well as delay. In the analysis of historical bus travel time, the coefficient of variation differed between time-points. The coefficient of variation is related to the mean absolute percentage prediction error (MAPPE) achieved using time series (full autoregressive) prediction models. Links with a higher coefficient of variation had a higher MAPPE and links with a lower coefficient of variation had a lower MAPPE [8].

Some models were used to predict travel time such as, regression method and neural network. Regression method requires the availability of dependent variable and whatever factors as being independent variables. Neural network is more complicated approaches and not simple in application. As generally known, ARIMA From above reason, ARIMA is chosen to predict bus travel time based on the bus travel time series data.

4.0 CASE STUDY AND DATA COLLECTION

Ipoh-Lumut corridor is located at Perak State in Peninsular Malaysia which link the Ipoh City (State Center) and Lumut Port (where the Malaysia Navy and Pangkor Island, tourism place located). A long this corridor, there are many attractive and potential land uses which potentially generate and attract the trips of people. This corridor is categorized into rural area with some important central business district (CBD) including such as residential area, school, state building, and universities. The 82.6 km length of divided multiple lane highway is the main road in this corridor connecting three districts, Kinta (Ipoh City), Perak Tengah and Manjung. The regular bus services are operated from 07:00 to 21:00. Figure 1 shows the illustration of bus service operation, such as location, departure time, distance and travel time. The main bus stop locations are indicated in the figure, but the bus actually can stop any where for boarding and alighting of passengers.

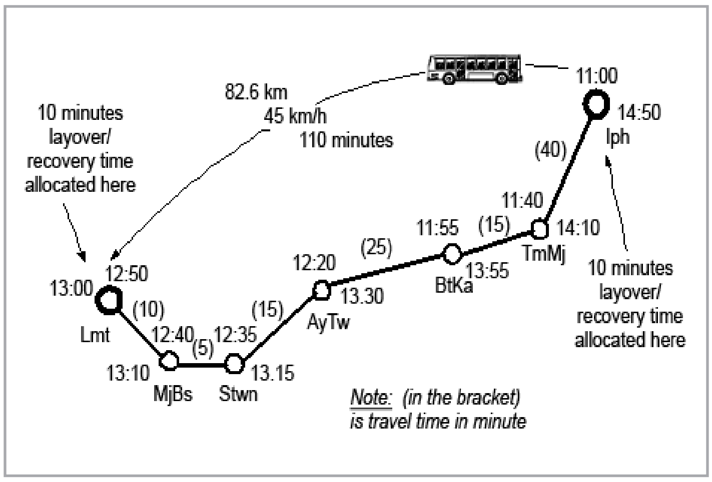


Figure 1: Location and scheduled departure of case study

In this study, data for analysis was collected during full one day period (07:00-21:00), one week period (11:00-15:00) and one year period (11:00-15:00). The data used is primarily collected by on-board survey over the Perak Roadways's 14-hour weekday service, which is playing Ipoh-Lumut corridor. The number of data points used for analysis was 12 months x 2 days per month x 2 trips per day = 48 trips per year. The primary data collected for analysis comprise the following:

a. arrival and departure times of the bus at stop points
b. location of stop points (bus station, bus stop and non bus stop)
c. name of location and environmental situation within the bus route

In addition, secondary data were also used to help surveyor on the primary data collection, such as road network map, timetable and other information on existing bus service.

On-board survey was conducted with more accurate and detailed information gathered about vehicle movement and boarding and alighting of passengers compared to at check point survey. As the survey is conducted by observer getting on a vehicle traveling over the route along a period of time, so that it is called as on-board survey. The observer records the data of location and the time at which the vehicle stop for boarding and alighting of passengers and other data need. For this case the handheld GPS was used for recording the spatial and timely data. For addition, the observer also could count the boarding and alighting passengers over an entire route for a specified time period. At this point, there was no device installed in the bus service system as well as used in more advanced bus service in other city, so that, there was no GPS data collected with advance system in this case.

In this paper, GPS is merely used as instrument or tools in data collection which support in recording the spatial and time based data of bus operation. Here, GPS is not applied specifically in estimating bus arrival in term of providing the real time information.

The type of bus service is regular stage-bus service which operates in mixed traffic. It is true that its travel time is affected by some factors, such as delay at intersection, delay at link and delay at bus stop. The data used for prediction is total bus travel time from the start to the end terminal which is compiled in time series based data. So that this study focus on the total travel time from start to end terminal. The bus travel time in mixed traffic is influenced by factors, such as level of traffic, delay (at link, junction and bus stop), departure time, and number of bus stop. In this case, the focus is not on the detail factors affecting bus travel time. Otherwise, the model is just applied to data series of bus travel time.

5.0 DEVELOPMENT OF ARIMA MODEL

5.1 Time Series Analysis Using ARIMA Model

In this section, it is shortly reviewed the non-seasonal Box-Jenkins Models for a stationary time series data. The Box-Jenkins methodology refers to the set of procedures for identifying, fitting, and checking the ARIMA models with time series data. Prediction or forecasts follow directly from the form of the fitted model [9].

Identification of the type of time series formula within ARIMA models are discussed briefly below.

5.1.1 Autoregressive Model

Autoregressive (AR) is defined as the model of a time series in which the current value of the series is a linear combination of previous values of the series, plus a random error. A p^{th}-order autoregressive model which is noted as AR(p) has the general form,

$$Y_t = \phi_0 + \phi_1 Y_{t-1}, \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \varepsilon_t \quad (1)$$

Where,

Y_t = Response (dependent) variable at time t

Y_t–1, Y_{t-2}, …, Y_{t-p} = Response variable at time lags t–1, t–2, …, t–p, respectively

$\phi_0, \phi_1, \phi_2, \dots, \phi_p$ = Coefficients to be estimated

ε_t = Error term at time $_t$

5.1.2 A.2 Moving Average Model

A Moving Average (MA) is simply a numerical average of the last N data points. The simple moving average is intended for data of constant and no trend nature. A q^{th}-order moving average model which is symbolised as MA(q) has the general form,

$$Y_t = \mu + \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - \dots - \theta_q\varepsilon_{t-q} \quad (2)$$

Where,

Y_t = Response (dependent) variable at time t

μ = Constant mean of the process

θ_1, θ2, …, θq = Coefficients to be estimated

ε_t = Error term at time t

$\varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_{t-q}$ = Errors in previous time periods that are incorporated in the response Y_t

5.1.3 Autoregressive Moving Average Model

As a combinatory model, the Autoregressive Moving Average Model is generally noted as ARMA(p,q) which has the general form,

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \varepsilon_t$$
$$- \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - \dots - \theta_q Y_{t-q} \quad (3)$$

Since the time series data prepared, then it can be produced the graph of sample autocorrelation function (ACF) and the sample partial autocorrelation function (PACF) to determine the form of models, which the processes can be summarised by referring Table 1 as follows:

Table 1: How to determine the model by using ACF and PACF patterns

Model	ACF	PACF
AR(p)	Dies down	Cut off after lag q
MA(q)	Cut off after lag p	Dies down
ARMA (p,q)	Dies down	Dies down

5.1.4 Autoregressive Integrated Moving Average

Autoregressive Integrated Moving Average is abbreviated as ARIMA. ARIMA is a general model widely used in time series analysis. Based on the prior investigation of the behavior of a series, it is specified three numbers that represent the order of autoregressive (p), the degree of differencing (d), and the order of moving average (q). Thus, the general form of model is written as ARIMA(p,d,q). The model can be extended to incorporate seasonality.

5.2 Steps In The ARIMA Model Building

The steps in the ARIMA model-building are briefly explained below. Figure 2 shows the flow chart of model development stages and model validation.

5.2.1 Model Identification

First step is to create ACF and PACF graph. A graph of autocorrelation function (ACF) is used to determine whether the series is stationary or not. The time series is considered stationary if the graph of ACF of the time series values either cuts off fairly quickly or dies down fairly quickly. If the graph of ACF dies down extremely slowly, then the time series values should be considered non-stationary. If the series is not stationary, it can be converted to a stationary series by differencing. The original series is replaced by a series of differences. An ARMA model is then specified for the differenced series. Differencing is done until a plot of the data indicates that the series varies about a fixed level, and the graph of ACF either cuts off fairly quickly or dies down fairly quickly.

Model for non-seasonal series are called Autoregressive integrated moving average model, denoted by ARIMA(p,d,q). Here, p indicates the order of the autoregressive part, d indicates the amount of differencing, and q indicates the order of the moving average part. If the original series is stationary, $d = 0$ and the ARIMA models reduce to the ARMA models.

ARIMA (p,d,q) has the general form,

$$\phi_p(B)(1-B)^d Y_t = \mu + \theta_q(B)\varepsilon_t$$

$$\text{or } \phi_p(B)W_t = \mu + \theta_q(B)\varepsilon_t \tag{4}$$

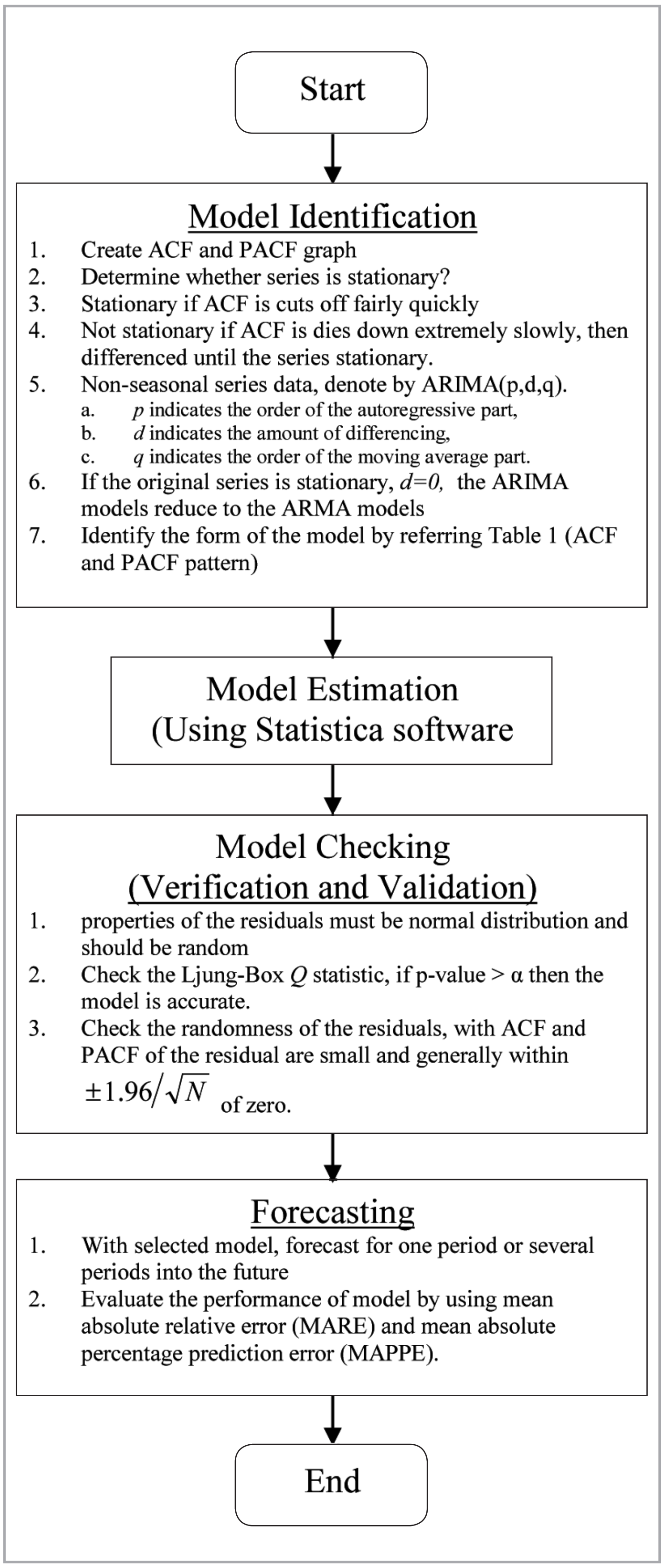


Figure 2: Location and scheduled departure of case study

Where,

ϕ_p = Coefficient to be estimated

B = Lag or backward linear operator defined by $BY_t = Y_t\text{-}1$.

W_t = Stationary series obtained as the d^{th} difference of Y_t., that is $W_t = (1\text{-}B)^d Y_t$

μ = Constant mean of the process

θ_q = Coefficient to be estimated

ε_t = Error term at time t

Second step is to identify the form of the model to be used by using the theory in Table 1, after a stationary series has been obtained.

5.2.2 Model Estimation

Estimate the parameters for a tentative model has been selected.

5.2.3 Model Checking (Verification and Validation)

In this step, model must be checked for adequacy or accuracy by considering the properties of the residuals whether the residuals from an ARIMA model must has the normal distribution and should be random. An overall check of model adequacy is provided by the Ljung-Box Q statistic.

If the p-value associated with the Q statistic is small (p-value $< \alpha$), the model is considered inadequate. The analyst should consider a new or modified model and continue the analysis until a satisfactory model has been determined.

Moreover, it can be checked that the properties of the residual with the graph as follows:

1) Check the normality by considering the normal probability plot or the p-value from the One-Sample Kolmogorov – Smirnov Test.
2) Check the randomness of the residuals by considering the graph of ACF and PACF of the residual. The individual residual autocorrelation should be small and generally within $\pm 1.96/\sqrt{N}$ of zero.

5.2.4 Forecasting with the Model

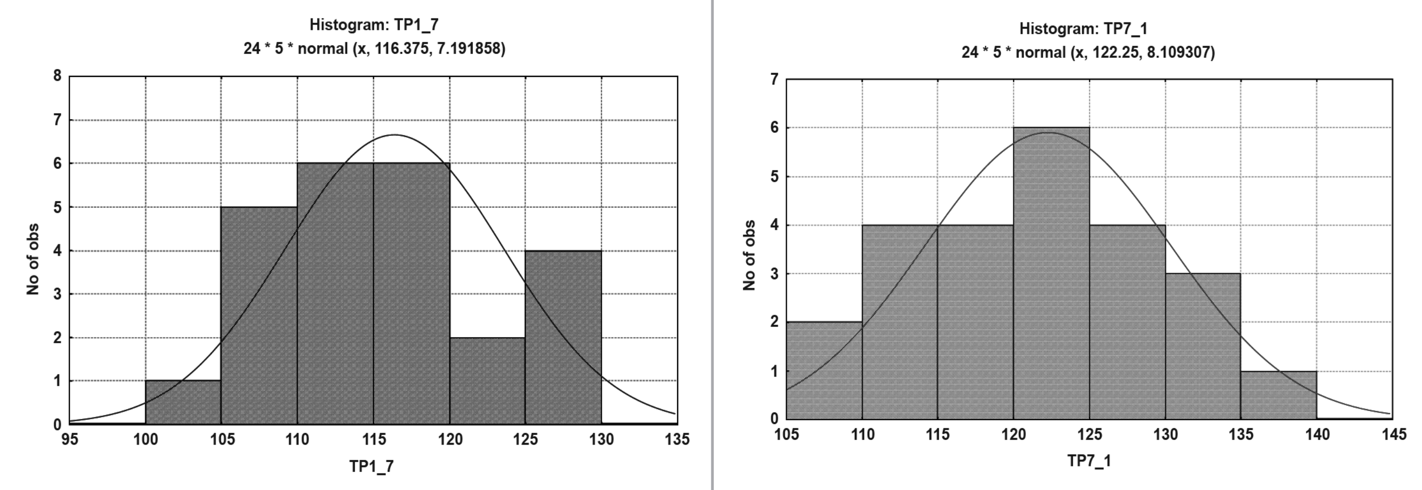
Forecasts for one period or several periods into the future with the parameters for a tentative model have been selected. Evaluation of forecasting was performed by using mean absolute relative error (MARE) and mean absolute percentage prediction error (MAPPE).

6.0 RESULTS AND DISCUSSION

The results obtained and finding models are presented in the following sub-sections.

6.1 Graphical Data Presentation

Data for analysis were series data of bus travel time collected from January to December 2007. Observation was done for two typical days, workday and weekend each month. The



(a) From Ipoh (TP=1) to Lumut (TP=7)

(b) From Lumut (TP=7) to Ipoh (TP=1)

Figure 3: Histogram and normal distribution of bus travel time

unit of travel time is measured in minute, thus, unit of minute is consistently used in the discussion. Descriptive statistics of bus travel time series was summarised in Table 2. From the Figure 3, it was shown that the distribution of travel time met with normal distribution $N(x, 116.38, 7.192)$ for Ipoh to Lumut direction (from time point TP=1 to time point TP=7). Meanwhile, for the opposite direction, Lumut to Ipoh, the normal distribution fitting was $N(x, 122.25, 8.109)$. The difference of average travel time of both directions is 5.9 minutes (5% difference). This is due to the operating speed for Lumut to Ipoh direction is lower (40km/h) with the higher variation of travel time (6.6%) compared to those of Ipoh to Lumut direction. For Ipoh to Lumut direction, the operating speed and variation of travel time are 43 km/h and 6.1%, respectively.

Table 2: Descriptive statistics of bus travel time (ARIMA_TT)

	TP1_7 : ***Travel time from TP=1 to TP=7***	***TP7_1 :*** ***Travel time from TP=7 to TP=1***
Mean	116.375	122.25
Std.Dv.	7.191858	8.109308
Minimum	102	110
Maximum	129	139
First	1	1
Last	24	24
N	24	24

Based on the Box & Whisker plot (See Figure 4), for both directions, the standard deviation and standard error of travel time changed increasing as the distance traveled increased from the starting time point to the downstream time points. The standard deviation plot, for instance, between time points of 1_2, 1_3, 1_4, 1_5, 1_6 and 1_7 tended to increase because there was delay propagation due to various traffic conditions and bus operating speed. The same changes were experienced for other pairs of time points.

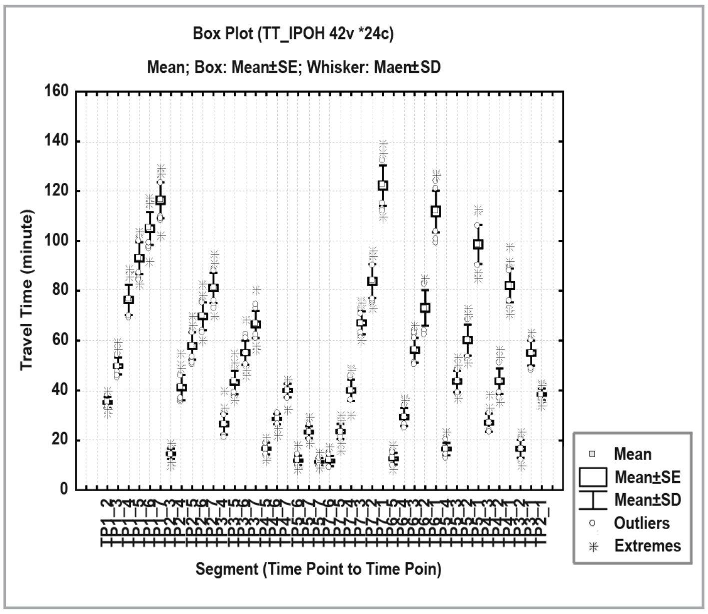


Figure 4: Standard deviation and standard error of bus travel time

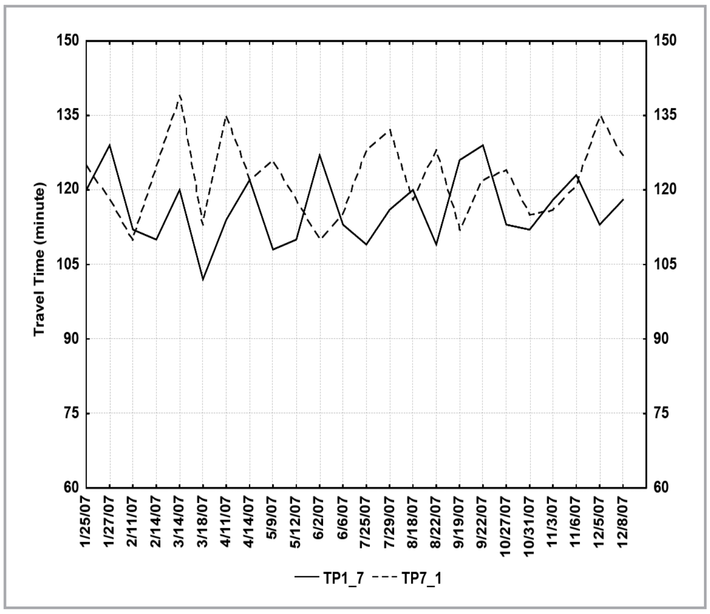


Figure 5: Time series plot of the bus travel time

6.2 The Aart of ARIMA Model Building and Forecasting

By following of the steps in the models-building, the results can be obtained, as shown in Figures 3-5. The time series data are considered stationary at around the mean value. Most of the autocorrelation values are smaller than 1.96 times their standard errors (*i.e.*, the probability of 95% confidence limit) as indicated by the dotted line in the ACF plot in Figure 4 and 5. The time series is stationary because the graph of ACF of the time series values either cuts off fairly quickly or dies down fairly quickly. Thus, it was not necessary to make transformation and differencing of the data, in this case *d*=0. To build the proper model, it was reasonable to applied some tentative ARIMA models for Ipoh to Lumut direction such as, ARIMA(0,0,1), ARIMA(0,0,2), ARIMA(1,0,0) and ARIMA(2,0,0). And, for the Lumut to Ipoh direction, the tentative models considered were as follows: ARIMA(0,0,1), ARIMA(0,0,2), ARIMA(1,0,0), and ARIMA(2,0,0).

Table 3 shows the estimated parameters of selected model of bus travel time prediction. Parameter estimation in ARIMA models is completely achieved by maximising the likelihood (probability) of the data. The method for computing maximum likelihood value for ARIMA is Exact (Melard algorithm) [11]. For Ipoh to Lumut direction (from TP=1 to TP=7), the model was without transformation (the amount of difference, *d*=0), named ARIMA(0,0,2) = MA(2) with MS Residual = 46.722. Furthermore, the model for Lumut to Ipoh direction (from TP=7 to TP=1) is ARIMA(0,0,1) = MA(1) with MS Residual = 68.102.

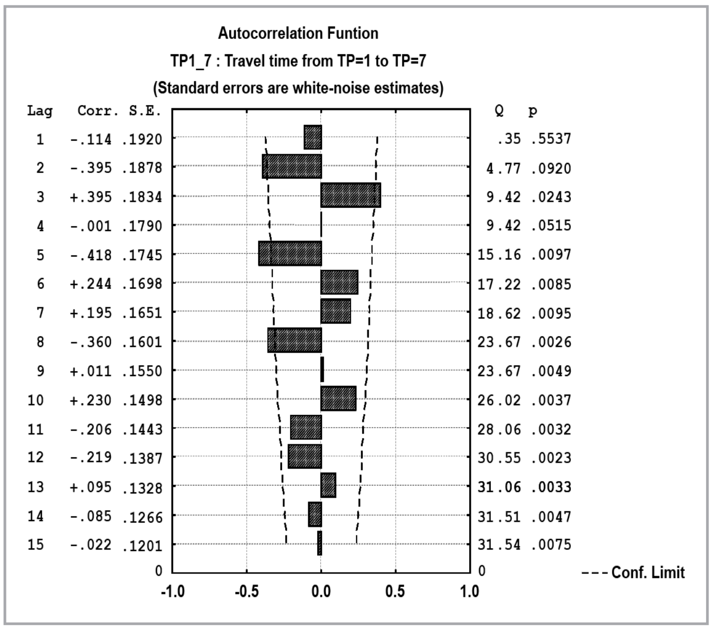


Figure 6: ACF of the bus travel time

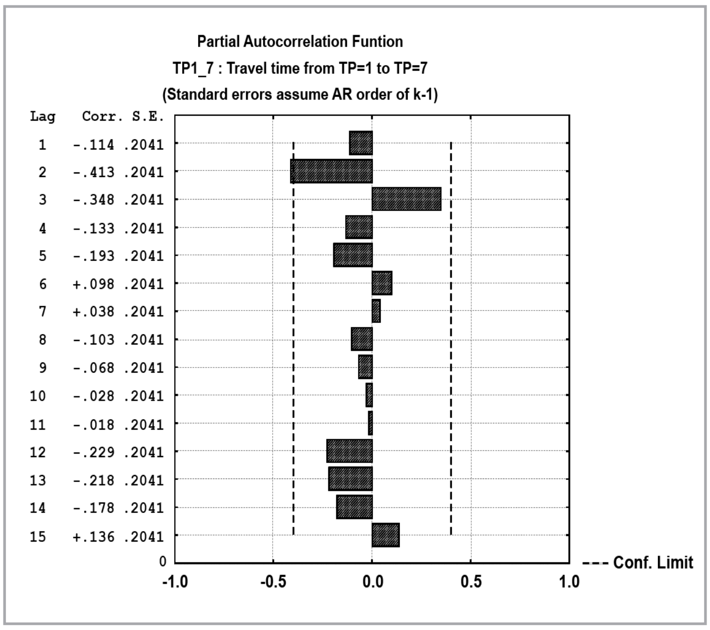


Figure 7: PACF of the bus travel time

For the Ipoh-Lumut corridor, the fitted models are formed as follows:

a) Ipoh to Lumut direction:
$$Y_t = 116.1627 + 0.0447*\varepsilon_{t-1} - 0.4289*\varepsilon_{t-2} \quad (5)$$

b) Lumut to Ipoh direction:
$$Y_t = 122.2072 - 0.1218*\varepsilon_{t-1} \quad (6)$$

More details, all the tentative ARIMA models were shown below. Based on the estimated model parameters of respective bus travel time resulted by using STATISTICA 7 software, it could be obtained the models in the form as the following in Table 4.

Table 3: Estimated mode parameters of bus travel time

Input: TP1_7 : Travel time from TP=1 to TP=7 (ARIMA_TT) Transformations: none Model:(0,0,2) MS Residual= 46.722						
	Param.	***Asympt.Std.Err.***	***Asympt.t(21)***	***p***	***Lower 95% Conf***	***Upper 95% Conf***
Constant	116.1627	0.9341	124.3513	0.0000	114.2200	118.1053
$q(1)$	-0.0447	0.2039	-0.2191	0.8287	-0.4687	0.3794
$q(2)$	0.4289	0.1885	2.2757	0.0335	0.0370	0.8208
Input: TP7_1 : Travel time from TP=7 to TP=1 (ARIMA_TT) Transformations: none Model:(0,0,1) MS Residual= 68.102						
	Param.	***Asympt.Std.Err.***	***Asympt.t(22)***	***p***	***Lower 95% Conf***	***Upper 95% Conf***
Constant	122.2072	1.5267	80.0476	0.0000	119.0411	125.3734
$q(1)$	0.1218	0.2855	0.4266	0.6738	-0.4703	0.7139

Table 4: Model results of bus travel time

Models	***Equations***	***Note***
1. *For Ipoh to Lumut direction:*		
a) ARIMA(0,0,1) = MA(1)	$Y_t = 116.2890 - 0.2602*\varepsilon t\text{-}1$	1st-order moving average
b) ARIMA(0,0,2) = MA(2)	$\mathbf{Y_t = 116.1627 + 0.0447*\varepsilon_{t-1} - 0.4289*\varepsilon_{t-2}}$	**2nd-order moving average**
c) ARIMA(1,0,0) = AR(1)	$Y_t = 116.3530 - 0.1107*Y_{t-1}$	1st-order autoregressive
d) ARIMA(2,0,0) = AR(2)	$Y_t = 116.1699 - 0.1828*Y_{t-1} - 0.4606*Y_{t-2}$	2nd-order autoregressive
2. *For Lumut to Ipoh direction:*		
a) ARIMA(0,0,1) = MA(1)	$\mathbf{Y_t = 122.2072 - 0.1218*\varepsilon_{t-1}}$	**1st-order moving average**
b) ARIMA(0,0,2) = MA(2)	$Y_t = 121.9329 - 0.5090*\varepsilon_{t-1} - 0.4910*\varepsilon_{t-2}$	2nd-order moving average
c) ARIMA(1,0,0) = AR(1)	$Y_t = 122.2269 - 0.0794*Y_{t-1}$	1st-order autoregressive
d) ARIMA(2,0,0) = AR(2)	$Y_t = 122.1416 - 0.0979*Y_{t-1} - 0.1519*Y_{t-2}$	2nd-order autoregressive

From Figure 8-10, proof that the selected ARIMA(0,0,2) or MA(2) is an appropriate model for bus travel time from Ipoh to Lumut direction, and is better than other tentative ARIMA models. Similarly, by using the other graph for Lumut to Ipoh direction it can be obtained the ARIMA(0,0,1) or MA(1) which is proper model for bus travel time in the case.

The models were checked for accuracy by considering the properties of the residuals whether the residuals from an ARIMA model has the normal distribution and should be random. For overall, as shown in Figure 9, the models are considered adequate as the p-value of associated with the *Ljung-Box Q* statistic is large (p-value > α).

The equation of the final result can be used to approximately generate the historical patterns of bus travel time in a time series and forecast the future value of the time series of bus travel time.

From Figure 8, the histogram shows that it looks like the residuals are normally distributed. Apparently, it seems how well the normal distribution fits the actual distribution of residuals. Moreover, it can be checked the properties (randomness) of the residuals with the graph of ACF and PACF of the residual (See Figure 9-10). The individual residual autocorrelation was small and generally falling within limit $\pm 1.96/\sqrt{N}$ of zero (there is no residual serial correlation). In other words, the residuals are independent of each other as second criteria of ARIMA models.

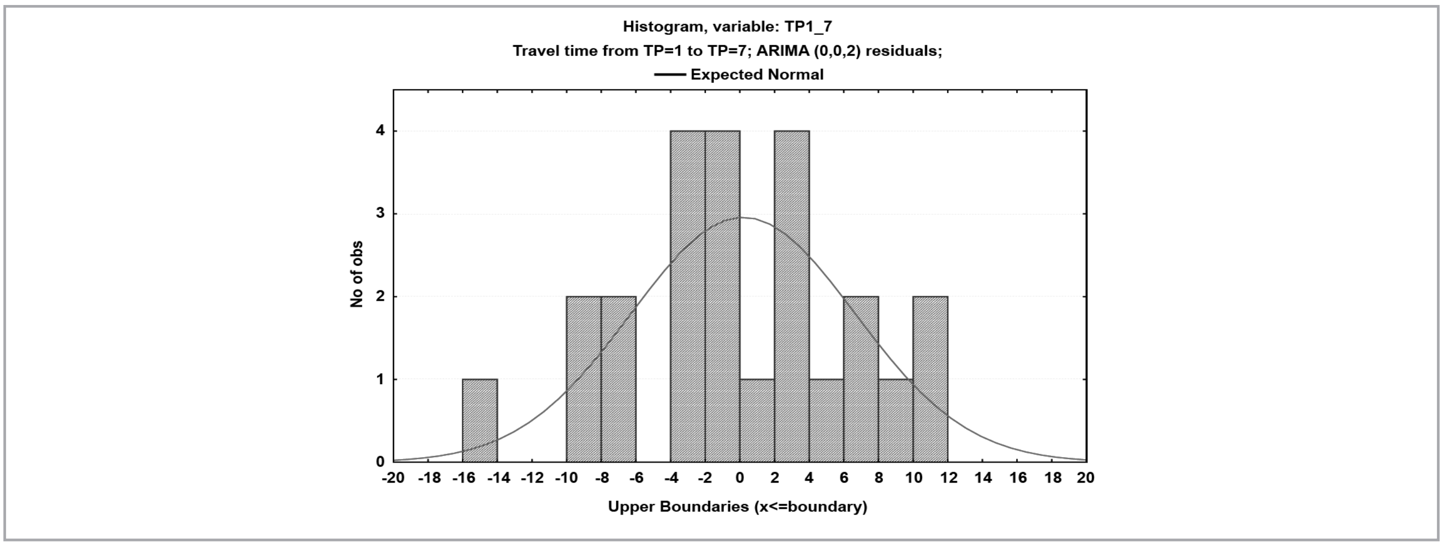


Figure 8: Histogram of the residuals of the bus travel time

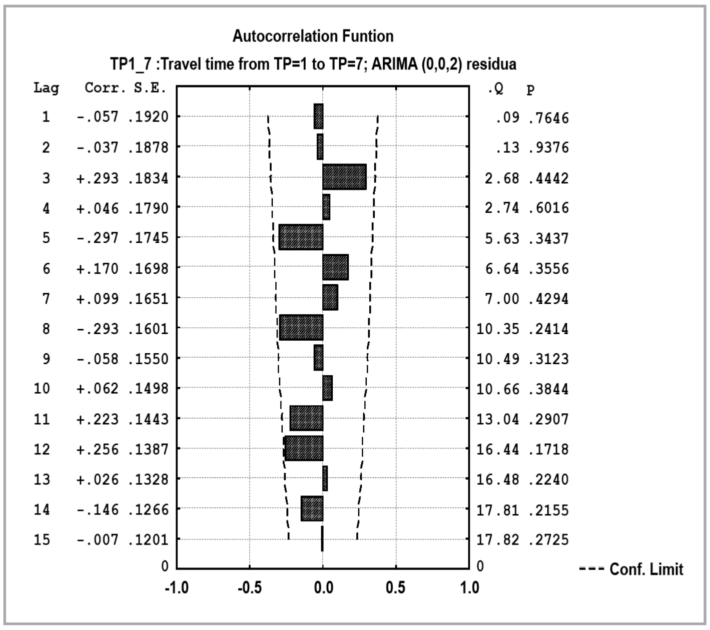


Figure 9: ACF of the residuals of the bus travel time

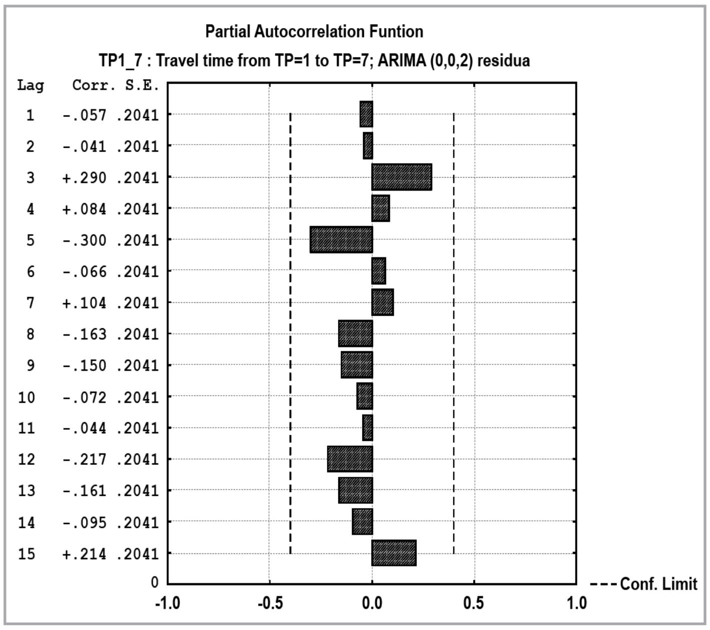


Figure 10: PACF of the residuals of the bus travel time

7.0 APPLICATION OF BUS TRAVEL TIME PREDICTION

Data set used for prediction (validation) consists of data set for model estimation and being added with any observed data else.

The performance of model is measured by the degree of accuracy. Accuracy of the model is indicated by statistical closeness such as mean absolute relative error (MARE) and mean absolute percentage predicting error (MAPPE). Both are indicator of model performance. The model which has minimum value of MARE and MAPPE is the accurate model (the best) among the several tentative models in predicting bus travel time. In other word, the minimum residual (error) indicate high accuracy model.

In this application, the prediction of bus travel time with the model obtained is properly done by considering the value of MARE (mean absolute relative error) and MAPPE (mean absolute percentage prediction error). Performance of models was measured by calculating MARE and MAPPE for each tentative models as being formulated:

$$\text{MARE} = \frac{\sum_{i-1}^{n} \left| ActualValue_i - ForecastValue_i \right.}{n} \tag{7}$$

$$\text{MAPPE} = \frac{\sum_{i-1}^{n} \left| \dfrac{ActualValue_i - ForecastValue_i}{ActualValue_i} \right.}{n} \tag{8}$$

where,
n : number of cases or data points
MARE in respective unit of minute
MAPPE is measured in %

The proper model is determined by comparing the MARE and MAPPE values among the tentative models as indicated by the minimum value of MARE and MAPPE. Table 5 shows the MARE and MAPPE values. ARIMA(0,0,2) = MA(2) and ARIMA(0,0,1) = MA(1) are selected because of the smallest MARE and MAPPE values.

The MA(2) and MA(1) equations can be used to approximately generate the historical patterns of bus travel time in a time series and forecast the future value of the time series of bus travel time. For instant, the residual of travel time of 10% was tolerable, therefore, the route distance of 82.6 km which was traveled in a round trip within 240 minutes (4 hours), then the delay would likely being 24 minutes. The assumption of delay of 10% travel time seemed still reasonable for the regular stage bus which traditionally operated in mixed traffic. Meanwhile, for bus rapid transit (BRT) system generally concern with delay of 5 minutes which was tolerable.

In this case, both the model results could describe well the historical pattern of bus travel time with the minimum MAPPE values less than 10% (See Table 5). On the other hand, the delay as indicated by MARE values, both 4.44 and 6.77 minutes, were quite tolerable because those were not significant delay compared to the bus travel time of 117 and 123 minutes, respectively.

Predicted and actual (observed) bus travel times were plotted with 95% confidence limit in Figure 11-12. It was clearly shown that the bus travel time series vary stationary in mean value of 116.38 minutes for Ipoh to Lumut direction. Similarly, it varies in mean value of 122.25 minutes for Lumut to Ipoh direction. For both directions, there is no trend or linear relation between date and month of the year and bus travel time series. Most of data and predicted travel time for both Ipoh to Lumut and Lumut to Ipoh direction fell within 95% confidence limit, meaning that the models are reasonable and acceptable.

Table 5: Residual analysis and performance of models

Models	*MARE (minute)*	*MAPPE (%)*
1. Ipoh to Lumut Direction:		
a) ARIMA(0,0,1) = MA(1)	5.74	4.67
b) ARIMA(0,0,2) = MA(2)	4.44	3.88
c) ARIMA(1,0,0) = AR(1)	12.83	10.30
d) ARIMA(2,0,0) = AR(2)	74.34	64.24
2. Lumut to Ipoh Direction:		
a) ARIMA(0,0,1) = MA(1)	6.77	5.64
b) ARIMA(0,0,2) = MA(2)	7.86	6.42
c) ARIMA(1,0,0) = AR(1)	10.10	8.06
d) ARIMA(2,0,0) = AR(2)	30.56	24.66

As shown in Figure 11-12, also, the forecasting results indicated that there was no specific trend line showing the changes of bus travel time in the following period. In the following period, the average travel time for both directions would be 117 and 123 minutes, respectively.

The prediction was only based on the historical travel time data, other factors, like as road and traffic conditions were not considered. As well known, the upgrading of the existing highway to be divided 4-lane 2-way highway in this corridor was done from middle of 2006 and finished totally in September 2007. The traffic volume might change from time to time before and after upgrading work. Those factors were not accounted to this study as the upgrading of road was done per segment while data was collected by on board survey. In addition, the mixed-traffic condition would be varies per segment along the corridor regarding the constructing work.

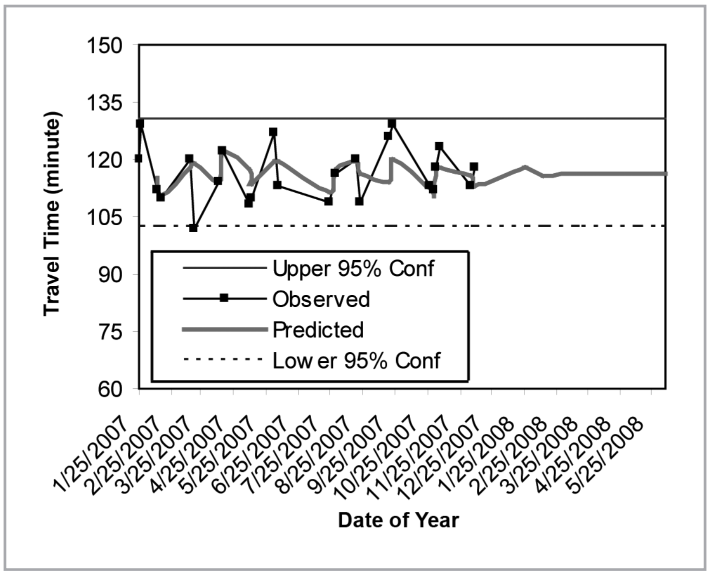


Figure 11: Travel time prediction (Ipoh to Lumut direction)

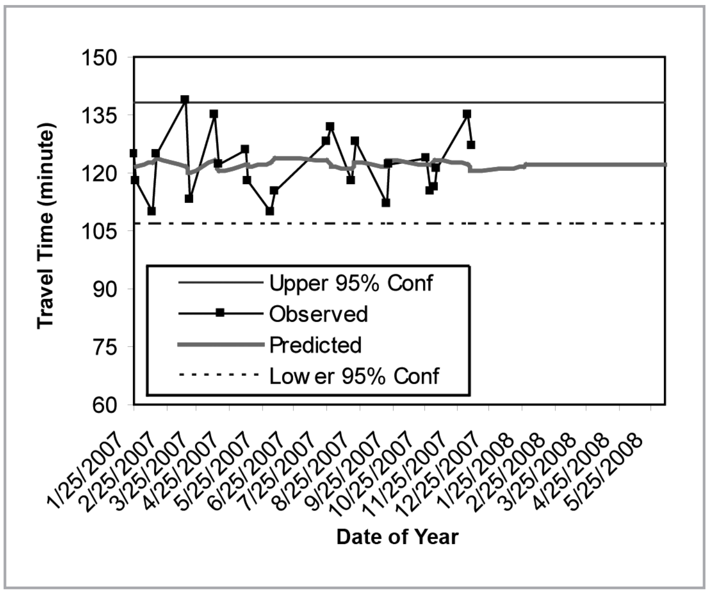


Figure 12: Travel time prediction (Ipoh to Lumut direction)

8.0 CONCLUSION

The results show that the predicted travel times by using the moving average model, MA(2) and MA(1) are close to the observed values. Those are indicated by the MARE and MAPPE values. The moving average models obtained had minimum MARE and MAPPE values compared to other tentative ARIMA models which being assessed. Those mean that MA(2) and MA(1) models are appropriate to be applied for the bus travel time prediction for Ipoh to Lumut and Lumut to Ipoh direction, respectively. And, those models can be used for bus travel time prediction in the case and they were statistical acceptable to be used in timetable design.

9.0 ACKNOWLEDGMENT

The author would thank to the Perak Roadways Sdn. Bhd. for the help to facilitate on conducting the on board survey during data collection. A good appreciation is awarded to the anonymous reviewer for their valuable suggestion on the previous manuscript. And, many thanks are delivered to Universiti Teknologi PETRONAS for the facilities provided and financial support in this research. ■

REFERENCES

[1] Peat, Marwick, Mitchell, and Co., *Transportation SystemsEvaluation Indicators*, U.S. DOT, Washington, DC., 1973.

[2] Khisty, C.J, and Lall, B.K., *Transportation Engineering: An Introduction*, 3rd edition, pp. 464-466, Prentice Hall, New Jersey, 2003.

[3] Dailey, D.J., Wall, Z.R., Maclean, S.D., and Cathey, F.W., "An Algorithm and Implementation to Predict the Arrival of Transit Vehicles", *Intelligent Transportation Systems Conference Proceeding*, Michigan, IEEE, pp. 161-166, 2000.

[4] Dailey, D.J., and Cathey, F.W., "Transit Vehicles as Traffic Probe Sensors", *Transportation Research Board*, 81st Annual Meeting, Washington, D.C., 2002.

[5] Abdelfattah, A.M., and Khan, A.M., "Models for Predicting Bus Delay", *Transportation Research Record 1623*, Transportation Research Board, 1998.

[6] Park, D., Rilett, L.R., and Han, G., "Spectral Basis Neural Networks for Real-Time Travel Time Forecasting", *Journal of Transportation Engineering*, pp. 515-523, 1995.

[7] Chien, I.J., and Kuchipudi, C.M., "Dynamic Travel Time Prediction with Real-Time and Historical Data, *Transportation Research Board*, Washington, D.C., 2002.

[8] Rajat Rajbhandari, *Bus Arrival Time Prediction Using Stochastic Time Series and Markov Chains*, Dissertation, pp. 5-56, January 2005.

[9] Box, G.E.P., and Jenkins, G.M., *Time Series Analysis: Forecasting and Control*, Holden-Day, San Francisco, California, 1970.

[10] Day, R.A., *How to Write and Publish a Scientific Paper*, 3rd edition, Oryx Press, Canada, 1988.

[11] StatSoft, Inc., *STATISTICA (data analysis software system)*, version 7, www.statsoft.com, 2004. Retrieved on 5 January 2009.

PROFILES

SUWARDO

Currently, he is pursuing PhD in Transportation Engineering and Traffic, Civil Engineering Department, Universiti Teknologi PETRONAS, Bandar Seri Iskandar, 31750 Tronoh, Perak Darul Ridzuan, Malaysia.
Tel: +6053688191, Fax: +6053654090.
E-mail: suwardo@yahoo.com
or suwardo@utp.edu.my

His permanent address is Civil Engineering Department, Faculty of Engineering, Gadjah Mada University. Jl. Yacaranda 1, Sekip, Yogyakarta 55281, Indonesia. Tel: +622747112126, Fax: +62274545193

He received his certificate of education and B. Eng in Civil Engineering (Teknik Transportation) from Gadjah Mada University in 1996. He obtained his M.Sc. (Magister Teknik, Rekayasa Transportasi) from Bandung Institute of Technology in 2001. His field of study and the research interest are in urban transportation engineering, public transport, transport project evaluation, traffic engineering, and road safety engineering. He is involved in professional association such as member of Inter-University Transportation Study Forum (FSTPT, Forum Studi Transportasi Perguruan Tinggi), member of Indonesian Road Development Association (HPJI, Himpunan Pengembangan Jalan Indonesia), member of Masyarakat Transportasi Indonesia (MTI), member of Eastern Asian Society for Transportation Studies (EASTS).

MADZLAN B NAPIAH

His current address is Civil Engineering Department, Universiti Teknologi PETRONAS, Bandar Seri Iskandar, 31750 Tronoh, Perak Darul Ridzuan, Malaysia. His academic experience is as follow:

- Doctor of Philosophy, University of Leeds
- Masters of Science (Engineering), University of Leeds
- Bachelor of Science, Michigan State University

Areas of specialization:
Highway and Transport Planning Engineering
Tel: +605-3687281
E-mail: madzlan_napiah@petronas.com.my

IBRAHIM B KAMARUDDIN

His current address is Civil Engineering Department, Universiti Teknologi PETRONAS, Bandar Seri Iskandar, 31750 Tronoh, Perak Darul Ridzuan, Malaysia.
His academic experience is:

- Doctor of Philosophy, University of Leeds
- Masters of Science in Civil Engineering, Louisiana State University
- Bachelor of Science, University of Sussex

Areas of specialization: Highway and Transportation
Tel: +605-3688585
E-mail: ibrakam@petronas.com.my