Abstract: Space-mean speed is an important parameter in the capacity analysis of rural and suburban highway. The ability to estimate space-mean speed from time-mean speed or spot speed will assist the study to analyse more data, besides saving time and cost. This paper presents the linear regression models developed for estimating space-mean speed for rural and suburban highways in Malaysia. However, due to the different characteristics and driving behavior between two-lane highways and multilane highways, using only one general equation to estimate space-mean speed for both two-lane and multilane highways is unrealistic and inaccurate. Therefore, it is essential to derive separate relationships for two-lane highways and multilane highways in Malaysia. Hence, in this study, specific models for each road facility which are the two-lane highways and multilane highways were derived. Statistical analyses were also conducted and discussed in this paper. The models derived are statistically accepted.

Key Words: Space-mean speed, time-mean speed, rural and suburban highways, linear regression

1. INTRODUCTION

Speed is an important transportation consideration because road users relate speed to economics, safety, time, comfort and convenience. Speed is a fundamental measurement of the traffic performance on the road system. Speed is defined as the rate of movement of a vehicle in distance per unit of time. Common units are miles per hour (mph), feet per second (fps), kilometres per hour (kph) and meters per second (mps). Spot speed is the instantaneous measure of speed at a specific location on a roadway. Journey speed is the average speed on a journey including stops and running speed is the average speed while in motion. If there is no stop delay, the average running speed is equal to the journey speed or the average travel speed.

The mean value of the speeds of individual vehicles can be determined either as space-mean or time-mean. Time-mean speed is the arithmetic mean of individual speed and space-mean speed is the harmonic mean. Time-mean speed is the sum of the measured spot speeds divided by the number of measurement and space-mean speed is another type of average speed where, the length of a segment is divided by the mean travel time of several vehicles or trips over the segment. The U.S HCM2000 (2000) used average travel speed which is the space-mean speed of vehicles in the traffic stream to analyze the capacities and level of service of uninterrupted flow facilities.
The difference in speed computations is attributed to the fact that the space-mean speed reflects the average speed over a spatial section of roadway, while the time-mean speed reflects the average speed of the traffic stream passing a specific stationary point. Specifically, Daganzo (1997) demonstrates that the space-mean speed is a density weighted average speed, while the time-mean speed is a flow weighted average speed. Given that a stationary observer will observe faster vehicles more often than slower vehicles while an aerial photograph would show more slow moving vehicles than faster vehicles over a fixed roadway length, it should come as no surprise that the time-mean speed is greater than or equal to the space-mean speed.

The same theory derived by Wardrop (1952), which shows the general relation between time mean speed and space mean speed is as follows:

\[
\bar{V}_t = \bar{V}_s + \frac{\sigma_s^2}{\bar{V}_s}
\]

where the standard deviation of the space mean speed is:

\[
\sigma_s = \sqrt{\sum_{i=1}^{n} f_i (v_i - \overline{v}_s)^2}
\]

It showed that the time mean speed is never less than the space mean speed on all occasions \( \bar{V}_t \geq \bar{V}_s \), and if there were any variations in speed at all, \( \sigma_s > 0 \) and \( \bar{V}_t > \bar{V}_s \).

Garber and Hoel (2002) describe a more direct relationship between time mean speed and space mean speed as follows:

\[
\bar{v}_t = 0.966\bar{v}_s + 3.541
\]

However, the parameters in the model are specific to the local roadway and traffic stream characteristics. Wang (2000) has conducted a study at the I-880 freeways in U.S. Based on the results obtained, the optimum model constant was 2.389, as opposed to 3.541, and the model slope was 0.986, as opposed to 0.966. Consequently, the model proposed by Garber and Hoel (2002) would require calibration to local roadway and traffic conditions and could not be generalized.

Apart from Garber and Hoel (2002), Drake et al. (1967) also did a similar study. Typical relationship between time-mean speed and space-mean speed developed by Drake et al. (1967) as presented in the U.S. HCM 2000 (Transportation Research Board, 2000) is as follows:

\[
\bar{v}_s = 1.026\bar{v}_t - 3.042
\]

However, due to the different characteristics and driving behavior between two-lane highways and multilane highways, using only one general equation to estimate space-mean speed for both two-lane and multilane highways is unrealistic and inaccurate. Therefore, it is essential to derive separate relationships for two-lane highways and multilane highways in Malaysia. Hence, in this study, specific models for each road facility which are the two-lane highways and multilane highways were derived.
2. METHODOLOGY

Spot speed data was recorded by using radar guns while space-mean speed data was estimated using car chasing method. Speeds of vehicle were measured at midpoint of each segment under fair weather conditions. Only vehicles that were separated by headways of more than 5 seconds were sampled. The speed measurements were made under very light flow conditions to make sure the sampled vehicles had headways far longer than 5 seconds. The sites for this study were selected based on the following criteria:

- The segment lengths are limited to 3.5 km. If the segments are too long, the effects on speed are different. Also, the segment cannot be less than 3.2 km as U.S HCM2000 (2000) defined the rural and suburban highway spacing between signalized should be more than 3.2 km. Karl L. Bang (1996) observed speed data from specially designated long-based rural and suburban highway sites ranging from 3 km to 7 km. Tseng (2005) collected data for multilane rural and suburban highways at segments that have spacing between signalized intersections in between 0.4 km to 5 km. The reason Tseng (2005) collected data less than 3.2 km was to calibrate the minimum length of segment where the speeds of vehicles were stable. He found that the speed increased rapidly when the spacing between signalized intersections increased from 0.4 km to about 2.5 km, after that (about 3.2 km) the speed record reached a steady value.

- Every segment should be at least 1 km from signalized intersections or major intersections. This is to eliminate the effects of stopping or slow vehicles due to crossing behaviour at signalised intersection or major intersection, the segment should be at least 1 km from signalised intersections or major intersections.

3. DATA COLLECTION

Time-mean speed and space-mean speed data were collected at 84 segments of rural and suburban two-lane highways and 66 segments of multilane highways throughout Malaysia. Table 1 show the summary of data collected for rural and suburban two-lane highways and multilane highways.

<table>
<thead>
<tr>
<th>Facilities</th>
<th>Two-lane highways</th>
<th>Multilane highways</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of sites</td>
<td>84</td>
<td>66</td>
</tr>
<tr>
<td>Locations (state)</td>
<td>Perak, Kedah, Penang, Perlis, Selangor, Melaka, Negeri Sembilan, Johor, Kelantan, Terengganu, Pahang</td>
<td>Perak, Kedah, Penang, Melaka, Negeri Sembilan, Johor, Kelantan, Terengganu, Pahang</td>
</tr>
</tbody>
</table>

4. RESULTS AND DISCUSSIONS

Initially, t-test was conducted to compare the means of space-mean speed observed at two-lane highways and multilane highways. The null hypothesis for t-test is that means for the two groups of cases are the same:

\[ H_0: \mu_1 = \mu_2 \]  

(5)

where \( \mu_i \) is mean of group \( i \). And the alternate hypothesis is that the means for the two groups
of cases are significantly different:

$$H_1: \mu_1 \neq \mu_2$$  \hspace{1cm} (6)

However, in order to select the appropriate $t$-test (equal or unequal variance), Levene test for equality of variances is used to test on whether the two samples have statistically equivalent variances. The null hypothesis is that the two samples come from populations with the same variances. In this test, 95% confidence interval was used. The results obtained from Levene’s Test are as shown in Table 2. Referring to Table 2, based on the observed significant level of 0.535 which was greater than 0.05 (hence not significant at 0.05 level), null hypothesis of equal variances was not rejected. Therefore, equal variances were assumed. Subsequently, based on the observed significant level of less than 0.001 in Table 2, null hypothesis was rejected and can be concluded that the means of space-mean speed obtained at two-lane highways and multilane highways were significantly different at 95% confidence interval.

<table>
<thead>
<tr>
<th>Levene's Test for Equality of Variances</th>
<th>t-test for Equality of Means</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>Sig.</td>
</tr>
<tr>
<td>-------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Space-mean speed</td>
<td>Equal variances assumed</td>
</tr>
<tr>
<td>Space-mean speed</td>
<td>Equal variances not assumed</td>
</tr>
</tbody>
</table>

Hence, the results obtained show that the means of space-mean speed are significantly different at two-lane highways and multilane highways. This indicates that it is important to derive separate equations for space-mean speed at two-lane highways and multilane highways.

Subsequently, linear regression analyses were conducted to derive the space-mean speed models. However, before linear regression analyses can be conducted, outliers were first identified. Data screening was applied to the data to find values outside the reasonable range for a variable and to determine whether they are real outliers or errors. Boxplots, stem-and-leaf or histogram can be used to identify outliers. Boxplots can clearly indicate the range in which the central 50% of the observation fall but can mask gaps or separations in the distributions in the distribution and also hide the existence of multiple outliers. In stem-and-leaf diagram, gaps are exposed and the leading digits of each outlier are reported (SPSS Inc., 1999). Histogram with normal curves overlaid are used to assess the distribution assumptions of individual variables, it is the simplest diagnostic test for normality. Normally distributed variables are very important particularly in regression analysis because if it were used to predict one variable from another, it will yield poor result due to highly skewed variable.

For two-lane highways, from a total of 84 data points, 3 data points were identified as outlier based on regression standardized residual. Data that have standardized residual value more than 2.5 are considered as outlier. Figure 1 shows histogram of regression standardized residual before and after removing the outlier for two-lane highway data.

1477
Hence, only data from 81 sites were used to develop the model. Linear regression was used to derive and determine the accuracy of the equation. Model summary obtained in this study for two-lane highways is shown in Table 3.

Table 3 Model summary (two-lane highways)

<table>
<thead>
<tr>
<th>Correlation coefficient, $R$</th>
<th>Coefficient of determination, $R^2$</th>
<th>Adjusted, $R^2$</th>
<th>Std. Error of the Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.939</td>
<td>0.881</td>
<td>0.879</td>
<td>3.24799</td>
</tr>
</tbody>
</table>

The result of the ANOVA test is as shown in Table 4. At 95% confidence level, the value of the observed significance level for the $F$-statistic which was less than 0.0005 indicated that the test of the coefficient of 0 was rejected.

Table 4 Results of ANOVA test (two-lane highways)

<table>
<thead>
<tr>
<th>Sum of Squares</th>
<th>d.f.</th>
<th>Mean Square</th>
<th>$F$</th>
<th>$P$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>6158.371</td>
<td>1</td>
<td>6158.371</td>
<td>583.764</td>
</tr>
<tr>
<td>Residual</td>
<td>833.404</td>
<td>79</td>
<td>10.549</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>6991.775</td>
<td>80</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The estimation results obtained in this study are as shown in Table 5. Based on the results obtained, the $t$-statistic value was well above 2, which indicates that the test of the coefficient of 0 was rejected.

Table 5 Estimation results obtained for this study (two-lane highways)

<table>
<thead>
<tr>
<th>Variables</th>
<th>$\beta$</th>
<th>Std. Error</th>
<th>$T$</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-1.704</td>
<td>2.871</td>
<td>-0.593</td>
<td>0.555</td>
</tr>
<tr>
<td>$\bar{V}_r$</td>
<td>1.016</td>
<td>0.042</td>
<td>24.161</td>
<td>&lt; 0.001</td>
</tr>
</tbody>
</table>

The equation derived to estimate space mean speed for two-lane highways based on regression analysis is as shown below:

$$\text{Equation}$$
\[
\bar{v}_s = 1.016\bar{v}_t - 1.704 \quad (R^2 = 0.881) \tag{7}
\]

Figure 2 shows the relationship between time-mean speed and space-mean data collected at two-lane highways in Malaysia.

Subsequently, after identifying the outliers, models were derived using linear regression for each facility. Residuals analyses were conducted in order to check the validity of linear regression assumptions. If the assumptions required for a regression analysis are met, then the residuals should have the following characteristics:

- The error terms should be approximately normally distributed with mean 0
- The error terms should have constant variance
- The error terms should be independent of each other

The normal probability plot can be used to test whether the error terms are normally distributed or otherwise. It plots a variable’s cumulative proportions (the proportion of the distribution that is less than the specified value) against the cumulative proportions of any of a number of test distributions. If the error terms are normally distributed, the points cluster around a 45° straight line.

Subsequently, normal probability plot was generated for the space-mean speed model for two-lane highways to test whether the error terms are normally distributed or otherwise. As shown in Figure 3, data points are distributed closely around the 45° straight line, hence indicating that the normality assumption is satisfied.
As for the subject of homogeneity of variance, plotting residuals against fitted values will help in checking whether non-constant variance exists. If the points in the residual plot show no pattern, then the variance is constant. The residuals plotted in Figure 4 appear to be randomly scattered around the horizontal axis and thus variance is constant. Therefore, this relationship is accepted.

For multilane highways, from a total of 66 data points, 2 data points were identified as outlier. Figure 5 shows histogram of regression standardized residual before and after the outliers were removed for data collected at multilane highways.
Figure 5 Histogram of regression standardized residual (a) before and (b) after the outliers were removed for data collected at multilane highways.

Hence, only data from 64 sites were used to develop the model. Linear regression was used to derive and determine the accuracy of the equation. Model summary in this study for multilane highways is shown in Table 6.

### Table 6 Model summary (multilane highways)

<table>
<thead>
<tr>
<th>Correlation coefficient, $R$</th>
<th>Coefficient of determination, $R^2$</th>
<th>Adjusted, $R^2$</th>
<th>Std. Error of the Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.918</td>
<td>0.842</td>
<td>0.839</td>
<td>3.48537</td>
</tr>
</tbody>
</table>

The result of the ANOVA test is as shown in Table 7. The value of the observed significance level for the $F$-statistic which was less than 0.0005 indicated that the test of the coefficient of 0 was rejected at 95% confidence level.

### Table 7 Results of ANOVA test (multilane highways)

<table>
<thead>
<tr>
<th>Sum of Squares</th>
<th>d.f.</th>
<th>Mean Square</th>
<th>$F$</th>
<th>$P$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>4012.854</td>
<td>1</td>
<td>4012.854</td>
<td>330.335</td>
</tr>
<tr>
<td>Residual</td>
<td>753.166</td>
<td>62</td>
<td>12.148</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>4766.019</td>
<td>63</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 8 shows the estimation results obtained for this study. Based on the results obtained, the $t$-statistic value was well above 2, which indicates that the test of the coefficient of 0 was rejected.

### Table 8 Estimation results (multilane highways)

<table>
<thead>
<tr>
<th>Variables</th>
<th>Coefficients</th>
<th>$T$</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta$</td>
<td>-2.528</td>
<td>-0.572</td>
<td>0.569</td>
</tr>
<tr>
<td>$\bar{v}_t$</td>
<td>1.021</td>
<td>18.175</td>
<td>&lt; 0.001</td>
</tr>
</tbody>
</table>

The regression equation derived to estimate space-mean speed for multilane highways is as in equation (8).

$$\bar{v}_s = 1.021\bar{v}_t - 2.528 \quad (R^2 = 0.842) \quad (8)$$
Figure 6 shows the relationship between time-mean speed and space-mean data collected at multilane highways in Malaysia.

![Figure 6 Relationship between time mean speed and space mean speed for multilane highways](image)

Figure 7 shows the normal probability plot. From the plot, the data points are distributed closely around the 45º straight line, indicating that the normality assumption is satisfied.

![Figure 7 Normal probability plot (multilane highways)](image)

Residual plot is as shown in Figure 8. The residuals plotted in Figure 8 are randomly scattered around the horizontal axis and thus variance is constant. Hence, this relationship is accepted.
Therefore, based on the statistical analyses, both of the linear regression models developed for two-lane highways and multilane highways were accepted and can be used to estimate space-mean speed at two-lane highways and multilane highways in Malaysia.

5. CONCLUSIONS

The models to estimate space-mean speed from time-mean speed for rural and suburban two-lane highways and multilane highways were developed and discussed in this paper. These models are statistically accepted but are yet to be validated. Therefore, further analysis will be conducted to validate these models.

ACKNOWLEDGEMENTS

The authors wish to express their sincere gratitude to Highway Planning Unit, Ministry of Works, Malaysia for funding this study under the project entitled “Malaysian Highway Capacity Study – Stage 3 (Inter-Urban)”.

REFERENCES

SPSS Inc. (1999) SPSS based 17.0 Applications Guide, Chicago, USA.
and suburban highway, Journal of the Eastern Asia Society for Transportation Studies,
Vol. 6, 1484–1495.
from Dual and Single Loop Detectors, Transportation Research Record, Journal of the
Transportation Research Board, TRB, National Research Council, Washington D.C.,
120-126.
Wardrop, J. G. (1952) Some Theoretical Aspects of Road Traffic Research, Proceedings of
the Institute of Civil Engineers, Vol. 1-2, 325-378.