Statistical Correlation between Road Surface Roughness and Traffic Accidents

Conference Paper · January 2018

DOI: 10.1049/cp.2018.1563

5 authors, including:

Soon Jiann Tan
Universiti Teknologi Brunei

Yok Hoe Yap
Universiti Teknologi Brunei

Statistical Correlation between Road Surface Roughness and Traffic Accidents

N. Hikmah[1], S. J. Tan[1], E. S. M .M. Zahran[1], Y. H. Yap[1], Hasnanizan Taib[2]

[1]Centre for Transport Research, Universiti Teknologi Brunei, Brunei Darussalam
[2]School of Applied Sciences and Mathematics, Universiti Teknologi Brunei, Brunei Darussalam
hikkiemah@gmail.com

Keywords: Road roughness, International roughness index, Crash risk, Statistical regression.

Abstract

Road surface condition is widely accepted as one of the most important factors that affect the safety of a land transport system. Although various researchers examined different types of surface condition, only a few focused on how road roughness, particularly quantified by International Roughness Index (IRI), can contribute to the occurrences of Road Traffic Accidents (RTA). Previous studies have found no consistent correlation and trend between both variables due to different regression techniques adopted, and the inclusion and exclusion of different contributory factors. This paper presents an analysis of the influence of IRI and various types of road properties on the observed number of accidents (μ-value) using a statistical regression analysis. The study road, where the accident and road properties were examined, was a high-traffic volume and high-speed highway section in Brunei Darussalam. The outcome of the analysis has indicated that the IRI values have a weak correlation to the observed number of accidents given that other explanatory variables are included into the regression. Limitations of the statistical regression acquired are discussed and recommendations to refine the selected variables and analysis methods are provided.

1 Introduction

The level of road safety is reflected by the frequency and severity of Road Traffic Accidents (RTA), and they may subsequently be applied to evaluate the risk which traffic users are exposed to. Statistically, in Brunei Darussalam, there was a slight increase of 0.9% between 2015 and 2016 with a total of 3,345 and 3,375 accident cases reported in 2015 and 2016 respectively [1]. The fatality rate has shown a declining trend in the long-term and is less significant when compared countries such as Malaysia and USA [2]. However, due to its small population and heavy reliance on private transport [3], RTA remains an important concern. Various initiatives are continually being made to improve the understanding of the causes of RTA, and thus to achieve a safer road system. Generally, road pavement deteriorates during its service life due to the continuous combined loading from road traffic and other unfavourable environmental conditions. The deterioration of road pavement would lead to an increase in pavement roughness that affects ride quality, and eventually may result in the occurrences of RTA [4]. Hence, the performance of road pavement, including road roughness, has a significant impact on road safety. If the influence of road surface roughness on RTA is to be confirmed and quantified, one approach is to analyse statistically the relationship between these two set of variables.

Several previous studies have shown that there is a relationship between these variables. Along the road network of USA, multiple research studies have suggested that the extreme values of IRI (in mm/km) have resulted in a significantly higher crash rate [5]. A similar result was reported from a case study in Victoria, where a clear increase in road crashes in segments with roughness value beyond 150 counts/km was found [6]. With the inclusion of speed as a contributory factor, as investigated by [7] in Southern Queensland, a positive relationship was concluded stating that roughness affected the speed, which consequently impacted on road safety. These outcomes, however, contradicted another study in the State of Victoria which found either uncorrelated or showed small enough correlations to disregard possible interactions among the variables [6]. In summary, it is still difficult to determine accurately how road roughness influence road safety.

This research paper analyses statistically how road roughness, expressed as International Roughness Index (IRI), and other engineering properties of the study road have influenced the observed number of accidents (μ-value). This was achieved by analysing historical RTA along the study road, which is a high traffic volume and high-speed road section, using the ESRI ArcGIS software. The objectives are to: (1) assess the key variables for the analysis; (2) develop an empirical regression model to assess the dependent changes of the μ-value on the changes of the selected variables; and (3) draw preliminary conclusions on the correlation between road roughness and the μ-value.

2 Methodology

2.1. Digitisation of data

All RTA that were reported in Brunei Darussalam would typically contain information such as crash type, time, and weather during the accident, along with other related details. However, road surface condition, particularly road roughness, would not be included. One particular highway, Lebuhraya Muara-Tutong, was previously found to have the highest RTA cases reported per year and has been selected as a case study for this paper. This road is a major highway that connects the Brunei-Muara District in the West to the Tutong District in the East. It has a high traffic volume with a posted speed limit of 100 km/h. By using a road surface profiler, the IRI data (expressed in m/km) for every 100-m intervals in 2012 were previously collected and will be used for analysis.

This research study also utilised RTA data for Lebuhraya Muara-Tutong from 2011 to 2015 that were collected from an earlier research that focused on the identification and ranking of hotspots along several major roads in Brunei Darussalam road network [8]. The IRI values were manually appended to the attribute table of the previously analysed hotspot zones. The raw IRI data were filtered and refined further to match the available number of RTA for GIS analysis which are spatially distributed on the base maps of the study road. Both existing and new data were then overlapped so that the full range of IRI and various types of RTA data were projected along Lebuhraya Muara-Tutong.

2.2. Regression methodology and development

The relationship between various types of road properties, IRI values and μ-value (i.e. the observed number of accidents) were analysed and interpreted using empirical regression models that were developed in ArcGIS. The Ordinary Least Square (OLS) method was selected to model and examine the statistical relationships between all the variables along Lebuhraya Muara-Tutong. Table 1 lists and classifies the selected data into dependent and explanatory variables in the analysis.

Table 1: Classification of variables included in the analysis.

Dependent (μ-value)	Explanatory
RTA during night or during day; RTA during wet or dry weather; Self-accidents or multi-vehicle accidents	IRI; No. of junctions; No. of conflict points; Road width

Prior to the regression analysis, the collinearity and hence redundancy of the candidate variables was determined through OLS tool. This procedure reduces bias caused by multicollinearity among the variables during the analysis. Subsequently, the tool generated an "OLS statistical report" that provides various statistical parameters of the regression model – these parameters were used to evaluate the developed regression model. The resultant model can then be used to determine the statistical correlation between μ-value and road roughness.

3 Preliminary Results and Discussion

Prior to the OLS analysis, exploratory data analysis in the form of bivariate scatterplots was performed to identify the potential relationships among the variables. With the scatterplot matrices, variables that might have similar links to the dataset would be determined. When the scatterplot matrix tool was run, several matrices did not show a clear trend, i.e. the points were clustered around a diagonal line on individual scatterplots between multiple variables (as seen in Figure 1). This indicated that the independent variables were not correlated to each other. Explanatory variables that were found to be correlated to each other would violate a key assumption of regression modelling, i.e. explanatory variables are independent of each other. Violation of this assumption could result in miscalculation and create redundancy of the model's statistical significance [9]. From the initial analysis, RTA during night or day, RTA during dry or wet weather, self-accidents or multi-vehicle accidents, IRI values, road width, number of junctions and conflict points projected unclear trends and were chosen for the OLS regression analysis.

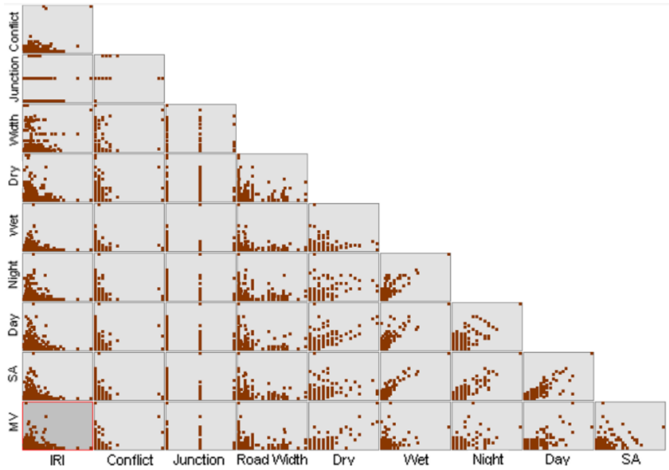

Figure 1: The scatterplot matrices of explanatory variables.

Since there were six types of μ-value selected as the dependent variables, they were assessed individually of their impact on the selected explanatory variables. After running the OLS tool, six OLS reports were provided. Figure 2 illustrates an example of the report.

Summary of OLS Results

Variable	Coefficient [a]	StdError	t-Statistic	Probability [b]	Robust_SE	Robust_Pr [b]	VIF [c]
Intercept	1.340123	0.405743	3.302887	0.001031*	0.533585	0.012281*	--------
IRI_L2	-0.227257	0.068348	-3.325004	0.000955*	0.075892	0.002876*	1.104429
NCONFLICTS	0.341968	0.053484	6.393824	0.000000*	0.149059	0.022123*	1.653545
NJUNCTIONS	-0.268453	0.158013	-1.698933	0.089879	0.254833	0.292568	1.495595
ROADWIDTH	-0.025711	0.050167	-0.512503	0.608506	0.055075	0.640810	1.031456

Input Features:	RTA	Dependent Variable:	SUM_MV
Number of Observations:	583	Akaike's Information Criterion (AICc) [d]:	2090.533480
Multiple R-Squared [d]:	0.077975	Adjusted R-Squared [d]:	0.071594
Joint F-Statistic [e]:	12.220202	Prob(>F), (4,578) degrees of freedom:	0.000000*
Joint Wald Statistic [e]:	24.291656	Prob(>chi-squared), (4) degrees of freedom:	0.000070*
Koenker (BP) Statistic [f]:	7.196246	Prob(>chi-squared), (4) degrees of freedom:	0.125874
Jarque-Bera Statistic [g]:	32230.530245	Prob(>chi-squared), (2) degrees of freedom:	0.000000*

Figure 2: An example of an OLS statistical report.

To avoid redundancy among the variables, it is important to ensure that the value of "Variance Inflation Factor" (VIF) does not exceed 7.5. This was confirmed in the OLS reports where VIF values for IRI, number of conflicts, number of junctions and road width were found to be 1.10, 1.65, 1.50 and 1.03 respectively.

The relatively small coefficient of each of the explanatory variables in the OLS reports showed that the variables did not show a strong influence on the μ-values. Both the number of junctions and conflicts points along the road have positive relationships with RTA frequencies, while IRI had the most negative coefficients of between -0.40 and -0.19 for each additional unit increase.

The consistency of the explanatory variables in the model in relation to the dependent variable in both geographical and data space was determined through the Koenker Statistic, i.e. the p-value. The p-values acquired from the regression analysis have indicated that the variability of the variables was statistically insignificant and equal across the range of values of the other variable that predicted it. The insignificance indicated that the variables did not behave randomly everywhere in the study. In other words, the variation in the relationship between the predicted value and each explanatory variable changed constantly with changes in the variable's magnitude. When the variables were analysed individually in each regression equation, the t-statistic of IRI values yielded the least magnitude of t-statistics for all reports. This indicated that the magnitude of IRI in relation to the other variables was more likely to be not significantly different relative to the variation in the overall dataset, i.e. IRI value is not highly correlated with the dependent variable. Table 2 shows the results of the t-statistic extracted from the OLS regression reports.

Generally, a t-statistic would assess only one variable per run but the F-statistic compares the fits of multiple coefficients simultaneously. As shown in Table 2, the joint F-statistics indicated the dataset have provided sufficient evidence to conclude that the model is significant, but not enough to conclude that any individual variable is significant. This suggested that the model significance was contributed by the other road properties instead of a clear contribution from the IRI values.

Table 2: t-statistic and F- statistic results.

Regression on (μ- value)	t-statistic (lowest value)	F- statistic
RTA during Day	IRI= -3.688	7.08581
RTA at Night	IRI= -2.449	2.22224
Self-Accident	IRI= -2.618	2.46152
RTA during wet weather	IRI= -1.537	1.11772
RTA during dry weather	IRI= -4.300	9.10441
Multi-vehicle accidents	IRI= -3.325	12.22020

A t-statistical probability of less than 5% for the coefficient of a particular explanatory variable suggests that the coefficient of the variable is unlikely to be zero at a 95% confidence level. However, this was not the case for the coefficients of IRI and frequency of conflict points, suggesting that these two variables do not have a statistically significant effect on the μ-values. This explained that it was likely that the IRI and conflict points do not contribute to the performance of the model.
OLS regression model requires the residuals to be normally distributed, and this was assessed through the Jarque-Bera statistic. All the regression models however showed that these statistics were significant, and therefore, the residuals were not normally distributed. This indicated that the model was strongly biased and the dataset is skewed.

Aside from the skewness of the dataset, the Spatial Autocorrelation tool was used to observe the correlations of the explanatory variables in the spatial dimensions. All the output reports showed that the regression residuals were clustered with a Z-score listed in Table 3. This indicated that there is a less than 5% likelihood that the clustered pattern could be the result of random chance. Although this could be accounted for by the fact that RTA's are constrained to the road network, there is also the possibility that not all key explanatory variables were present and that there was a misspecification within the model [10]. An example of spatial autocorrelation report for a regression equation where the observed number of accidents during the day was a dependent variable is shown in Figure 3.

Table 3: Z-score of respective μ- value.

Regression on (μ- value)	z-score
RTA during Day	5.81
RTA at Night	4.05
Self-Accident	6.86
RTA during wet weather	8.83
RTA during dry weather	2.98
Multi-vehicle accidents	2.50

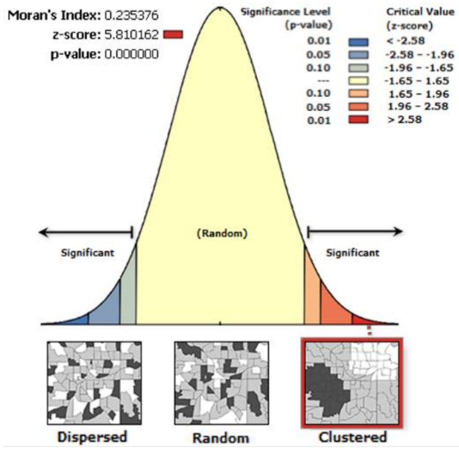

Figure 3: The Spatial Autocorrelation report from ArcGIS.

As shown by the assessments of bias, all models were shown to afflict by misspecification. This was consistent with low adjusted and multiple R-squared values of the models, which indicate how well the explanatory variables were able to explain the variation in the RTA μ-values. The adjusted R-squared obtained varies from 0.8% to 7.8% only, and thus at most, only 7.8% of the variation in the output variable was explained by the subset of input variables. This may also explain the weak relationship between IRI and RTA, as the model significance was contributed by the other included road properties. Another likely reason was that there were other factors and variables which likely had an impact on the RTA frequencies, but could not be included in this study.

The main objective of this paper was to determine how road roughness would influence the observed number of accidents; however, based on the results, IRI value exhibited statistically insignificant contribution throughout most of the tests when compared to the other independent variables shortlisted for this study. The results thus indicated no clear relationship between IRI and the μ-value, and hence IRI value has a weak relationship (at best) to the observed number of accidents. This may be due to the limitation of the current approach – most RTA were likely to have been caused by a combination of contributory factors (e.g. traffic volume, speed and so on) and these variables have not been analysed in the current model.

4 Conclusion and Recommendations for Future Work

Road surface condition is widely accepted as one of the most important factors that affect the safety of a land transport system; however, only a few researchers examined how road roughness, quantified by International Roughness Index (IRI), can contribute to the occurrences of Road Traffic Accident (RTA). The Ordinary Least Square (OLS) tool in ArcGIS was used to develop a regression model on a GIS database of RTA locations and IRI measurements, to assess the relationship between the individual explanatory variables and the overall model. Based on the summary of OLS reports, several interpretations were made. Amongst the various types of RTA, the sum of RTA during night or day, RTA during dry or wet weather, self-accidents or multi-vehicle accidents, IRI values, road width, number of junctions and conflict points were shown to be qualified as variables for the analysis. The main objective of this paper was to determine the influence of IRI on the μ-value (i.e. RTA data), and the statistical results have shown a weak correlation between the two variables. As the relationship of other variables to the μ-value may have overshadowed the correlation of IRI, this approach was unable to conclude the link between them. Past data did not indicate a strong enough relationship between IRI and the number of road accidents to enable the future μ-value (or any dependent variable) to be better predicted with the inclusion of IRI in the regression model.

Several recommendations can be made for further research to improve the outcome of the study. Other factors, such as speed and traffic volume, should be explored for inclusion into the analysis. In the process of data preparation, it is suggested to remove any possible influential outliers. Where justified by exploratory data analysis, normalising or linearising the data should be considered to possibly represent the data better for OLS analysis and thereafter produce a better outcome. The sole influence of IRI alone on the μ-value should be analysed before including other variables to allow comparison between different sets of explanatory variables.

The Geographical Weighted Regression (GWR) tool in ArcGIS, can also be used to investigate the spatial relationship between IRI and other variables across the study area. Other statistical methods such as factor analysis could also be considered, particularly given the multiple contributory factors typically involved in RTAs. The developed model should also be validated using further data collected from the field. Nevertheless, given the limitations inherited within the model development here, it is clear that further analysis will be required to determine the influence of IRI on the observed number of accidents.

Acknowledgements

The authors are grateful to the Public Works Department and the Royal Brunei Police Force for providing the necessary road accident and road condition data. The views presented in this paper are solely those of the authors and not of the supporting authorities and organisations.

References

[1] http://www.rtbnews.rtb.gov.bn, accessed June 2018.

[2] Tan, S. J., Ladi, S., Zahran, E. S. M. M., Yap, Y. H., Rahman, E. K. and Abdullah, N.: ‘A review of road traffic accident statistics and trends in Brunei Darussalam’, Brunei Int. Conf. of Eng. and Tech. (BICET), November 2016, pp. 1–4.

[3] Koh, W. C.: ‘Review to formulate a roadmap and draft national masterplan for a sustainable land transportation system for Brunei Darussalam’, Centre for Strat. and Pol. Stud., 2014, (5), pp.1-246.

[4] Adlinge, S.: ’Pavement Deterioration and its Cause’, Journal of Mechanical & Civil Engineering, 2009, 6, pp. 09-15.

[5] Elghriany, A., Ping, Y., Liu, P. and Yu, Q.: ’Investigation of the effect of pavement roughness on crash rates for rigid pavement’, Jour. of Trans. Safety and Security, 2015, 8, (2), pp. 164–176.

[6] Cairney, P. and Bennet, P.: ’An exploratory study of surface characteristics and crash occurrence on selected roads in Australia’, Research Report ARR, 2013, (382), pp. 1–89.

[7] King, B.: ’The Effect of Road Roughness on Traffic Speed and Road Safety’, 2014, pp.1-146.

[8] Zahran, E. S. M. M., Tan, S. J., Yap, Y. H., Rahman, E. K. and Abdullah, N.: ‘A GIS-based approach to identification and prioritisation of road traffic accident hotspots’, Brunei International Conference of Engineering and Technology (BICET), November 2016, pp. 1–489.

[9] Scull, P.: ‘Spatial Statistical Analysis’, Colgate University, 2000.

[10] http://desktop.arcgis.com/en/arcmap/10.3/tools/spatial-statistics-toolbox/ interpreting-ols-results.htm, accessed March 2018.