

Assessment of the effects of expressway geometric design features on the frequency of accident crash rates using high-resolution laser scanning data and GIS

Maher Ibrahim Sameen & Biswajeet Pradhan

To cite this article: Maher Ibrahim Sameen & Biswajeet Pradhan (2017) Assessment of the effects of expressway geometric design features on the frequency of accident crash rates using high-resolution laser scanning data and GIS, Geomatics, Natural Hazards and Risk, 8:2, 733-747, DOI: [10.1080/19475705.2016.1265012](https://doi.org/10.1080/19475705.2016.1265012)

To link to this article: <https://doi.org/10.1080/19475705.2016.1265012>



© 2016 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 14 Dec 2016.



Submit your article to this journal [↗](#)



Article views: 2565



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 13 View citing articles [↗](#)

Assessment of the effects of expressway geometric design features on the frequency of accident crash rates using high-resolution laser scanning data and GIS

Maher Ibrahim Sameen ^a and Biswajeet Pradhan ^{a,b}

^aDepartment of Civil Engineering, Faculty of Engineering, Geospatial Information Science Research Center (GISRC), University Putra Malaysia, Selangor, Malaysia; ^bDepartment of Energy and Mineral Resources Engineering, Sejong University, 209 Neungdong-ro Gwangjin-gu, Seoul, Republic of Korea

ABSTRACT

Accurate information on accidents and on the relevant factors that affect them is critical for establishing the relationship between accident frequency and explanatory factors. In this study, we present a simplified method to extract road geometric features accurately from very high-resolution laser scanning data to analyze accident frequency on the North-South Expressway in Malaysia. Using expressway geometric features (i.e. horizontal and vertical alignments) extracted from laser scanning data and accident histories, this research first developed an APM based on geometric regression and a geographic information system (GIS). Then, an elasticity analysis was conducted to investigate the relationship between accident occurrence and road geometric design features. Results of the case study showed that the length of the road segments (mean = 0.014, elasticity = 0.122), the number of vertical curves in a road section (mean = 4.797, elasticity = 0.999), and the presence of a horizontal curve in a road segment (mean = 2.746, elasticity = 0.877), the average distance to the nearest access point (mean = -0.001, elasticity = -0.035), and AADT (mean = 3.01, elasticity = 0.881) determined accident occurrence, all at a significance level of 5%. This study shows that laser scanning systems can provide an easy and efficient method to collect transportation data, particularly those for accident analysis.

ARTICLE HISTORY

Received 24 June 2016
Accepted 3 November 2016

KEYWORDS

Highway accidents; Bayesian approach; LiDAR; GIS; Malaysians expressway

1. Introduction

Recent studies have predicted that road accidents will become the fifth leading cause of death worldwide by 2030 (Pei et al. 2011). In Malaysia, current statistics show that the deaths per 100,000 people are close to 24 for all road users (Global Status Report on Road Safety 2015). Therefore, many researchers have realized the seriousness of the matter and have conducted an extensive body of research to improve road safety. The most common approach in studying explanatory factors associated with various traffic entities is analyzing the frequency of accidents on roadway segments over a specified period (Huang & Abdel-Aty 2010). The frequency of road accidents on definite roadway segments is typically analyzed using accident prediction models (APMs) (Wang et al. 2011). Such models are significant in emergency planning through site ranking, identifying prone areas and key factors that affect the severity of an accident, and in

improving road geometric design by incorporating safety consideration into road designs and standards. Road accidents are not negative, discrete, and random (Huang & Abdel-Aty 2010). Transportation data are commonly modelled using two approaches: statistical and computational intelligence (Karlaftis & Vlahogianni 2011). Statistical methods are highly recommended for developing APMs. Statistics have solid and widely accepted mathematical foundations and can provide insights into the mechanisms that create data. Statistical methods are best performed when researchers have knowledge or prior information regarding the functional relationship of the variables in a problem. In addition, statistical methods should be conducted when interpreting results and causalities is important (Karlaftis & Vlahogianni 2011).

A wide variety of statistical and machine learning approaches for modelling accident data is available in literature. One of the most popular methods used to estimate accident frequency is the count data model (Abdel-Aty & Radwan 2000; Anastasopoulos & Mannering 2009; Ayati & Abbasi 2011; Gomes 2013). The popularity of this method is mostly attributed to its suitability for count data modelling (Wang et al. 2011). Poisson distribution requires the variance of the count data to be equal to the mean. However, this case is not always true in accident data. Moreover, directly using Poisson models will underestimate the standard error of parameters, and thus, cause a biased selection of parameters. Therefore, negative binomial (NB) models have been developed to accommodate the overdispersion problem in accident data by including an error term in the Poisson model; adding this term enables the variance to differ from the mean (Hosseinpour et al. 2014). However, the presence of excess zero limits the use of these models because they cannot predict the existence of excess zero. To address this problem, zero-inflated models have been proposed (Hosseinpour et al. 2014). These models allow zeros to be generated through the following processes: generating structural zeros estimated from a binary (logit) distribution and generating sampling zeros derived from the NB distribution. Nonlinear relationships sometimes exist between dependent and explanatory variables, which cannot be identified by NB and zero-inflated methods. The use of support vector machine (SVM) algorithm has been proposed to address the nonlinear relationship between variables (Deublein et al. 2014; Dong et al. 2015). Some studies indicated that SVM is better than NB models in terms of goodness of fit (Li et al. 2008). The strength of SVM is ascribed to its basis on structural risk minimization, which provides a trade-off between hypothetical space complexity and the quality of fitting the training data (Li et al. 2008). However, the major challenge associated with the SVM model is the optimal input feature subset, particularly in complex and highly multivariate prediction models. This difficulty mostly arises from the feature subset selection influencing the appropriate kernel parameters, and vice versa (Dong et al. 2015).

In addition to the aforementioned models, generalized linear models (Cafiso et al. 2010), which exhibit the advantages of overcoming the limitations of conventional linear regression in traffic accident modelling, are widely used to model accidents. These models facilitate the assumption of an NB error structure, which is pertinent to accident frequency variation (Cafiso et al. 2010). However, these models are fixed parameter models, and thus, they can be limited for roads on which traffic is not mixed with numerous internal variabilities. For highways on which traffic is mixed with numerous internal variabilities ranging from differences in vehicle type to variations in driver behaviour, these models can result in varying effects of explanatory variables on accidents across locations.

Although selecting appropriate statistical models is critical to ensure accurate prediction model development, the quality and accuracy of traffic and highway geometric data are also critical. Matching data from different sources and then combining them require considerable effort and mostly produce low-quality data. By contrast, laser-scanning systems can provide high-quality, detailed, and accurate highway geometric data by capturing dense point clouds of selected highway sections. A complete highway geometric environment can be modelled using the point clouds collected by laser scanning systems (Poullis & You 2010). In addition, accident data are difficult to collect and only become available gradually, such as yearly. Possible variations can also exist for APMs as a result of changes in several influential factors, such as widening of road lanes, construction of new bridges, or installation of traffic lights.

Thus, the present study aims to use very high-resolution laser scanning data and a geometric regression model to fit the geometric design features of a subset extracted from the longest expressway in Malaysia (i.e. North-South Expressways [NSE]) into the number of accidents that occurred from 2009 to 2015. This study is different from those listed in literature because it uses a new data source (i.e. laser scanning systems) for geometric road modelling and accurate road geometric design features extraction. Furthermore, a geographic information system (GIS), which is capable of spatial analysis, is used for accurate road modelling through the proposed simplified methodology. First, this study presents a simplified methodology for extracting road geometric parameters from airborne laser scanning data. Then, using the extracted parameters and geometric regression, an APM is developed with the aid of GIS to evaluate the effects of road geometric design features on accident frequencies. The main contribution of this study is a new method for extracting accurate road geometric design features from laser scanning data and its application for road accident analysis. In this research, ArcGIS 10.3 is used for LiDAR data and orthophotos processing. AutoCAD Civil 3D 2015 is used to extract road geometric design features from LiDAR-based digital surface model, and regression analysis is conducted in NCSS 11 statistical analysis software (<http://www.ncss.com/>, accessed on 15 May 2016).

2. Geometric regression

Geometric regression (Hilbe 2014) is a generalization of Poisson regression which loosens the restrictive assumption that the variance is equal to the mean made by the Poisson model. In geometric regression, the Poisson distribution is generalized by including a gamma noise variable with a mean of 1 and a scale parameter of ν . This results in an NB distribution as follows (Hilbe 2014):

$$\Pr(Y = y_i | \mu_i, \alpha) = \frac{r(y_i + \alpha^{-1})}{\Gamma(y_i + 1)\Gamma(\alpha^{-1})} \left(\frac{\alpha^{-1}}{\alpha^{-1} + \mu_i} \right)^{\alpha^{-1}} \left(\frac{\mu_i}{\alpha^{-1} + \mu_i} \right)^{y_i} \quad (1)$$

where $\Gamma(\cdot)$ is the Gamma function, $\mu_i = t_i\mu$, $\alpha = 1/\nu$, μ is the mean frequency rate of y per unit of exposure. Exposure is often a period of time and the symbol t_i is usually used to represent the exposure for a particular observation. When no exposure given, it is assumed to be one. When the dispersion parameter α is set to one, the result is called the geometric distribution.

In geometric regression, the mean of y is determined by the exposure time t and a set of k regressor variables (the x 's). The expression relating these quantities is

$$\mu_i = \exp(\ln(t_i) + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki}) \quad (2)$$

Often, $x_1 \equiv 1$, in which case β_1 is called the intercept. The regression coefficients $\beta_1, \beta_2, \dots, \beta_k$ are unknown parameters that are estimated from a set of data. Their estimates are symbolized as b_1, b_2, \dots, b_k . The regression coefficients are estimated using the method of maximum likelihood. The method is explained in details in Cameron (2013, p. 81).

3. Materials and study area

The study area is a subset corridor from the longest expressway in Malaysia (i.e. NSE), with a total length of 44 km running from the stretch of Pedas Linggi to Ayer Keroh (239.7–195.4 km) (Figure 1). This study aims to determine the relationship between road geometric variables and the number of accidents occurring on an urban expressway. The prediction model proposed in this study aims to understand the relationship between road geometric variables and the number of accidents. Therefore, to establish the APM, accident data were collected from the police department in digital form (Excel sheets) over a six-year period (2009–2015). The accident data were collected from a 44 km

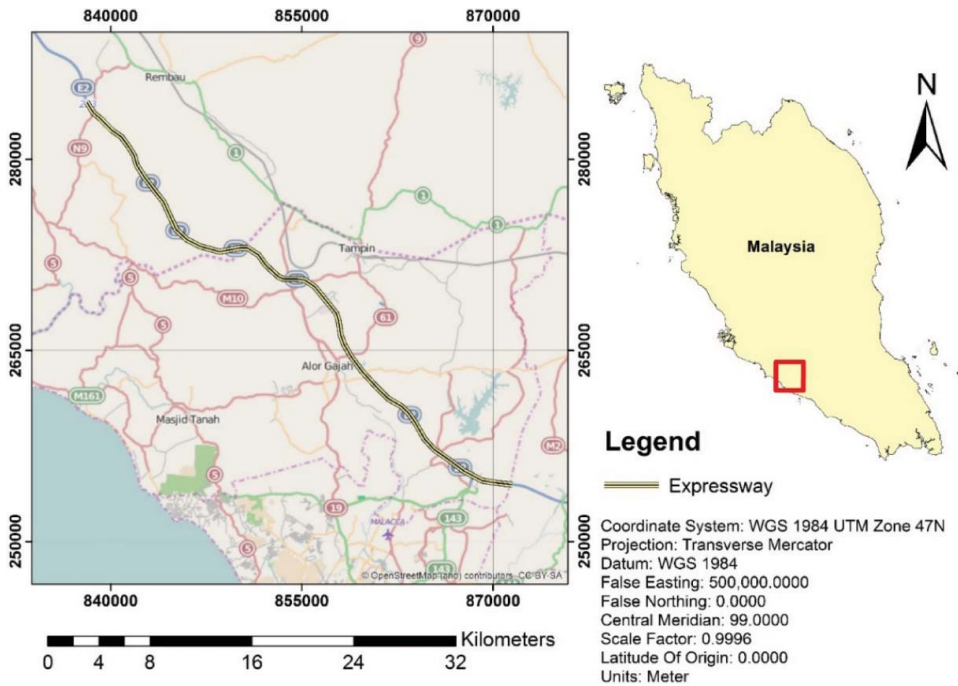


Figure 1. Location of the studied expressway.

section of the three-lane expressway (433 segments). The average number of accidents is five per road segment during the study period (2009–2015).

The LiDAR data used in this study were collected on 8 March 2015 using Riegl LM Q5600 and Hassleblad 39 Mp camera. The device had a spatial resolution of 13 cm, a laser scanning angle of 60°, and a camera angle of approximately -45°. In addition, the point spacing density of the LiDAR data was 3–4 pts/m². Figure 2 presents an example of the point clouds (a) and orthophotos (b) collected by the airborne LiDAR system for a small part of the study area.

4. Methodology

4.1. Extraction of road boundary from LiDAR point clouds and orthophotos

The overall methodology used to extract road boundary from LiDAR point clouds and the aerial photo is shown in Figure 3. In general, three main steps are used: preprocessing the original data,

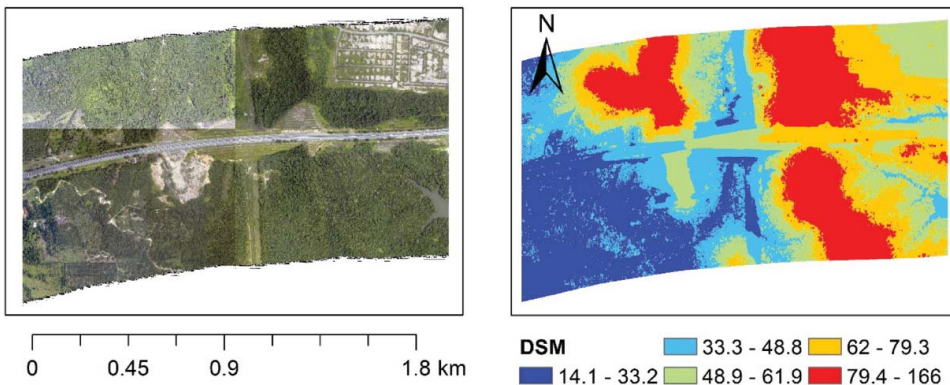


Figure 2. Example of LiDAR data, (a) a digital surface model constructed from LiDAR point clouds, (b) orthophoto.

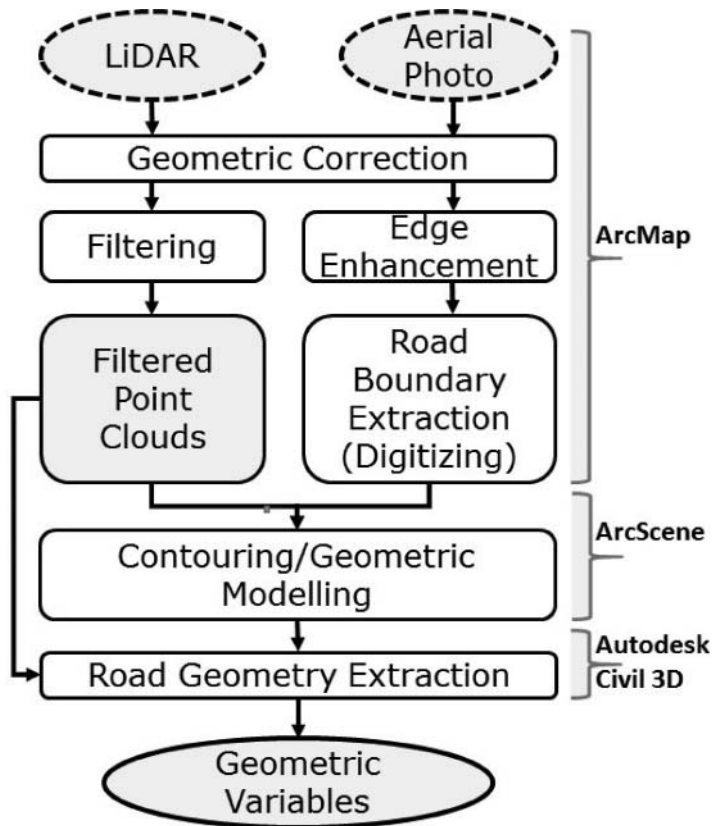


Figure 3. Overall workflow of extraction of road geometry from LiDAR point clouds.

processing the enhanced data, and road boundary extraction from the enhanced data. LiDAR systems capture point clouds and aerial photos using two separate sensors, and thus, geometric correction is an essential step. In the geometric correction step, the aerial photo is corrected based on a LiDAR intensity image to ensure accurate geometric matching between the LiDAR and aerial photo data. The steps of geometric correction method include identification of tie points in the LiDAR intensity image and the orthophoto. The points were uniformly distributed in the data-set. After that, the least square method (Kardoulas et al. 1996) was applied to estimate the coefficients which are important for geometric transformation process. After the least-squares solution, the polynomial equations were used to solve for X and Y coordinates of the GCPs, and to determine the residuals and the root mean square (RMS) errors between the source X , Y coordinates and the retransformed X , Y coordinates. The orthophoto was corrected using 34 GCPs identified from the LiDAR intensity image at clearly identifiable points (i.e. road intersections, corners, power lines). The geometric correction was done in ArcGIS 10.3 software.

Once the LiDAR and aerial photo are geometrically corrected, applying filtering techniques is important to eliminate or reduce noises in the data. LiDAR point clouds are then filtered using the threshold technique to eliminate the outliers from the data (Brink 1996). In particular, the threshold technique is adopted to evaluate the difference between ground elevation and surface elevation by checking whether the difference is less than a certain value; otherwise, the point is considered an outlier. The aerial photo is used to simplify road boundary extraction in two dimensions (X , Y). In this study, the aerial photo was first filtered using Canny edge detection to enhance the edges in the image. The enhanced aerial photo was then manually digitized to extract the road boundary as vector features (Esri vector data storage format).

4.2. Extraction of road geometric design parameters

In this research, six road geometric design parameters were identified from literature that has a potential relationship with road accident count and can be extracted from LiDAR data. These parameters are the radius of a horizontal curve, length of a vertical curve, *K* value of vertical curves, number of vertical curves in a segment, presence of a horizontal curve in a segment, and distance to the nearest access point. This section explains the steps of extracting these design parameters from LiDAR data and orthophotos.

At first, the LiDAR point clouds and the extracted road boundary were prepared and organized in a personal geodatabase. Note that the vector data in the personal geodatabase can be converted into AutoCAD Civil 3D by using conversion tools of ArcGIS. The data was converted from shapefile (.shp) to CAD (.DWG) file formats during data processing stages. After the data was prepared in correct formats, the LiDAR point clouds were then clipped based on the road boundary extent. Afterwards, the new set of point clouds was used to generate three-dimensional (3D) surface in AutoCAD Civil 3D software. The 3D surface generated from the previous step allowed modelling of the complete road geometry. For each road segment (see Section 4.2), design parameters were calculated from the generated 3D surface. Figure 4 illustrates the process of curve fitting for one road segment (Figure 4(a)) and also shows a basic diagram of vertical curves (Figure 4(c)). The horizontal curvature of a road segment was estimated by fitting a curve using the least square method. In this step, the centreline of the road was generated by offsetting the road boundary. Based on the centreline, a curve was constructed and its parameters were calculated (Figure 4(b)). Among the curve parameters, radius was used to describe the relationship between horizontal curve geometry and accident count. On the other hand, vertical curves are important transition elements in geometric design for highways (Figure 4(c)). Most vertical curves in road design are symmetrical parabolic curves for a good reason. The parabolic curve is the natural vertical curve followed by any projectile. The quadratic equation of vertical curves is expressed as:

$$y = \frac{g_1}{100}x + \frac{g_1 - g_2}{200L}x^2 \tag{3}$$

where *g*₁ is the grade of tangent in, *g*₂ is the grade of tangent out, and *L* is the length of the curve.

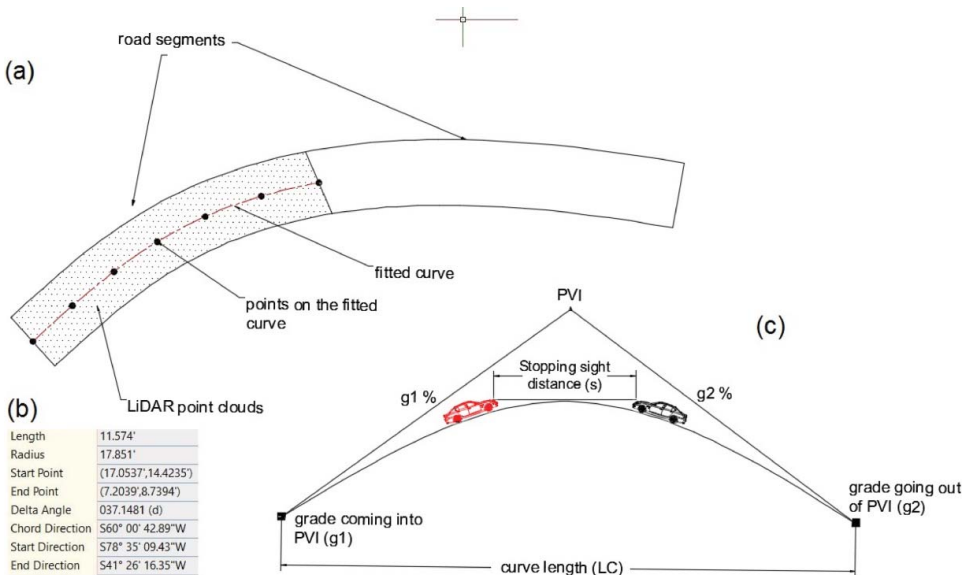


Figure 4. Estimation of road geometry features, (a) the process of curve fitting, (b) an example of attribute information about the road segments and their design parameters, and (c) a basic diagram of vertical curves.

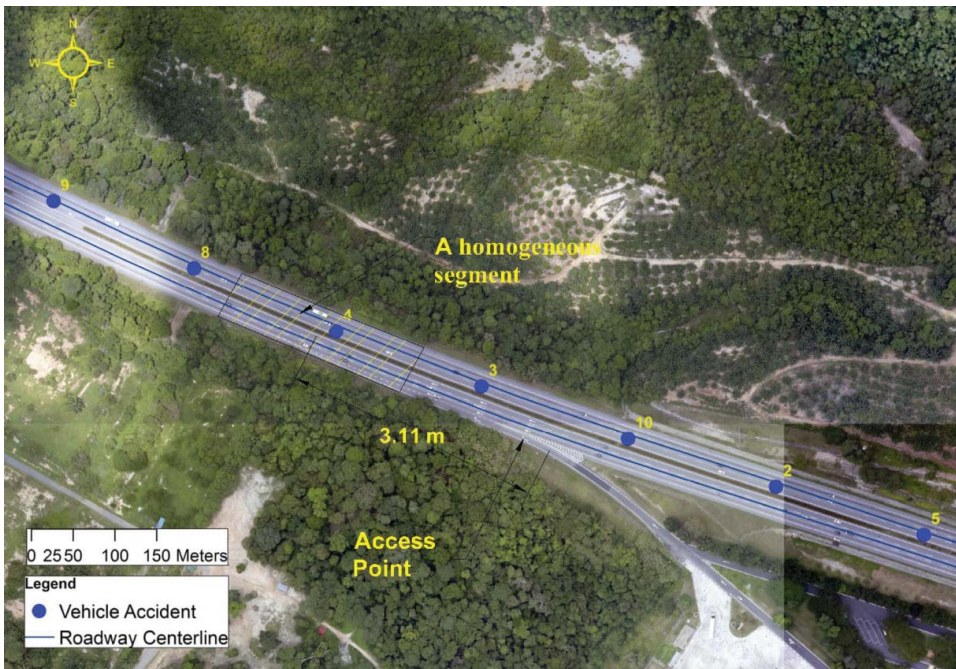


Figure 5. An example of the factor 'distance to the nearest access point' used in the accident count modelling.

In addition, K value is another factor used to predict the accident frequency. It is a design parameter related to vertical curves. This value represents the horizontal distance along which a 1% change in grade occurs on the vertical curve. It is a function of grade change and length of the curve and it can be expressed as:

$$K = \frac{L}{|g1 - g2|} \quad (4)$$

Calculation of the K value requires two parameters, length of curve and grade changes. The vertical curves were estimated by fitting a parabolic curve using the least square method with manually selected points. After that, the grade changes were estimated from curve tangents and the curve length. The estimation of grade changes and the length of the curve allowed to calculate the K value which was used for accident investigation.

Furthermore, Figure 5 shows an example of the distance to the nearest access point. The distance was calculated from the centre of a road segment to the nearest access point using the 'Euclidean Distance' tool in ArcGIS. Previous studies showed that vehicles entering from the access point were involved in serious traffic conflicts (Manan & Várhelyi 2015). This could cause increasing number of accidents. The number of accidents of road segments is represented by blue dots in Figure 5. It can be observed that the number of accidents occurred in the segment at access point is relatively higher than other neighbouring segments. This may not be always true and detailed analysis should be made for general conclusion. Some observations are presented related to this factor in the results and discussion section of the current manuscript.

4.3. Road segmentation

The two possible means to segment a road into small sections, which can be used in accident prediction modelling, are fixed length and homogenous segmentations. Homogenous segmentation is

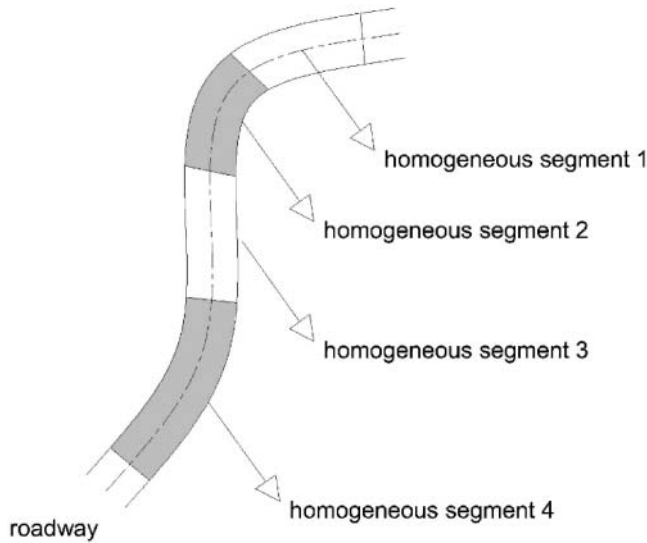


Figure 6. Example for segregation of a road section into four homogenous segments based on values of the observed explanatory variables.

mostly recommended in recent literature (Cafiso et al. 2010; Deublein et al. 2013; Fernandes & Neves 2013). A homogeneous segment is a road section on which the values of all explanatory variables to be included in the model can be assumed to be constant, and thus, risk is uniform (Deublein et al. 2013). Although homogeneous road segmentation is suggested, some considerations should be made when segmenting a road into small sections based on the included explanatory variables. For example, a study focusing on a single characteristic of a road segment (e.g. circular or straight alignments, width of lanes, etc.) is extremely limiting because the influences of other possible variables of the road environment are disregarded. To overcome the limitations inherent in considering only uniform road segments, high-quality road geometric data can be used (Fernandes & Neves 2013). In addition, accidents may not be reported on some roadway segments during the period over which the data are collected. In this case, the data are considered left-censored at zero. This censoring may occur for a number of reasons ranging from the possibility that no accident occurred on the roadway segment during the study period to the possibility that accidents that do not involve injury may not be reported (Anastasopoulos et al. 2012). If fixed length sections are used, then longer sections are suggested to obtain reliable APMs (Ackaah & Salifu 2011). In particular, accident rates should be computed from 0.8 km or longer sections.

On the basis of the preceding discussion, the homogeneous road segmentation method is used in this study because accurate road geometric parameters can be extracted from LiDAR point clouds. Figure 6 illustrates the process of sub-dividing a road section into four homogeneous segments based on the values of six explanatory variables (i.e. radius of a horizontal curve, length of a vertical curve, K value of vertical curves, number of vertical curves, presence of a horizontal curve, and distance to the nearest access point). The change in the value of any explanatory variable results in the start of a new homogeneous segment.

4.4. Modelling accident frequency using road geometry

The summary of the descriptive statistics of the explanatory variables derived from the 3D road geometric model is presented in Table 1. These explanatory variables were fitted into the number of accident occurrences in each homogeneous section using the geometric regression model.

Table 1. Summary of characteristics of homogeneous section variables considered for model development.

Explanatory variable	Mean	Min	Max	Standard deviation
Length of road segment	1388.86	577.69	2342.28	404.17
Radius of road horizontal curvature	34,804.73	1432.32	159,233.37	44,754.26
Length of vertical curve	612.75	0.00	2867.75	874.82
K value	249.94	0.00	1636.57	432.37
Number of vertical curves	0.56	0.00	4.00	0.89
Presence of a horizontal curve	0.47	0.00	1.00	0.51
Distance to nearest access point	1606.75	74.68	4217.52	1238.67
Accident occurrence	88.97	32.00	179.00	33.90
AADT (2010-2015)	3376	3231	3940	2896.92

Note: Number of observations (road segments) = 34.

The maximum-likelihood estimates of the regression coefficients are then estimated by the geometric regression model. These coefficients were then converted into elasticity values which could provide causalities better than the estimated coefficients by the geometric regression. Elasticity is the ratio of the per cent change in one variable to the per cent change in another variable. The coefficient in a regression is a partial elasticity since all other variables in the equation are held constant.

5. Results and discussion

Accurate road geometric parameters were extracted from very high-resolution laser scanning data using the proposed simplified approach. Afterwards, the geometric regression was estimated using the maximum likelihood approach as described in Section 2. Table 2 shows the intercept β_0 and the β coefficients associated with each explanatory variable used. The precision of the regression coefficient is shown as standard error. The fourth column in Table 2 shows the Wald statistic which represents the Chi-square test value. In this test, the null hypothesis $\beta_i = 0$ was tested against the two-sided alternative $\beta_i \neq 0$. *P*-value indicates the significance level of the test. In this study, the value less than 0.05 indicates that the explanatory variable is said to be statistically significant. On the other hand, the lower and upper confidence limits provide a large-sample confidence interval for the values of the coefficients.

Table 2 also shows the indicators estimated for checking model performance (i.e. Log likelihood, Deviance, and AIC). The Log likelihood is the value of log-likelihood function for the model that

Table 2. The geometric regression model and the estimated coefficients.

Explanatory variable	Regression coefficient $b(i)$	Standard error $Sb(i)$	Z value H0: $\beta = 0$	Two-sided <i>P</i> -value	Lower 95.0% confidence limit	Upper 95.0% confidence limit
Intercept	-6.534	0.668	-9.780	0	-7.843	-5.224
Length of road segment	0.014	0.001	27.050	0	0.013	0.015
Radius of horizontal curve	0	0	1.950	0.051	0	0
Length of vertical curve	-0.001	0.001	-2.020	0.044	-0.002	0
K value	-0.004	0.001	-3.750	0	-0.006	-0.002
Number of vertical curves	4.797	0.284	16.860	0	4.239	5.354
Presence of a horizontal curve	2.746	0.534	5.140	0	1.699	3.792
Distance to nearest access point	-0.001	0	-3.650	0	-0.001	0
AADT	3.01	0	6.312	0	2.65	4.74
Log Likelihood				-516.430		
Deviance				664.115		
AIC(1)				1048.861		

Table 3. The analysis of variance and the estimated elasticities.

Explanatory Variable	DF	Deviance	χ^2	P-value	Elasticity
Intercept only	1	564.934			-
Length of road segment	1	598.082	66.03	0*	0.122
Radius of horizontal curve	1	663.966	0.15	0.699	0.001
Length of vertical curve	1	665.623	1.51	0.219	-0.011
K value	1	662.081	2.03	0.153	-0.052
Number of vertical curves	1	671.224	7.11	0.007*	0.999
Presence of a horizontal curve	1	659.571	4.54	0.033*	0.877
Distance to nearest access point	1	672.584	8.47	0.003*	-0.035
AADT	1	633.678	6.91	0.009*	0.881
(Full model)	9	664.115			-

*Statistically significant at $\alpha = 0.05$.

fits the data perfectly. The deviance is the measure of the discrepancy between the fitted values and the data. AIC (1) is the Akaike Information Criterion and is one of the most commonly used fit statistics to compare different types of models.

The analysis of variance and the estimated elasticities are shown in Table 3. The information in the table provides the effect of each explanatory variable on the frequency of traffic accidents analyzed by the geometric regression model. The variables that found to be statistically significant at 5% confidence level are: length of road segment (P -value = 0), number of vertical curves (P -value = 0.007), presence of a horizontal curve (P -value = 0.033), distance to nearest access point (P -value = 0.003), and AADT (P -value = 0.009). On the other hand, the remaining variables [i.e. radius of horizontal curve (P -value=0.699), length of vertical curve (P -value = 0.219), K value (P -value = 0.153)] are found to be statistically insignificant at 5% confidence level.

The proposed model fits the common road geometric design features into the number of accidents that occurred on the homogeneous sections and offers insight into the effect of geometric features. The overall accuracy of the developed model for predicting the number of accidents with only the geometric variables was $R^2 = 0.637$. The scatter plot of the actual accident numbers versus the predicted number of accidents is shown in Figure 7. Given that road characteristics are not the major factors that contribute to accident occurrence, the overall accuracy of the proposed model is acceptable in this study. The output of the proposed model allows designers to determine the thresholds of geometric variables such as maximum grade, minimum horizontal curve radius, and the number of vertical curves in each section. In addition, the proposed model can be used as a guideline for setting highway design standards and developing highway geometric design policies or manuals.

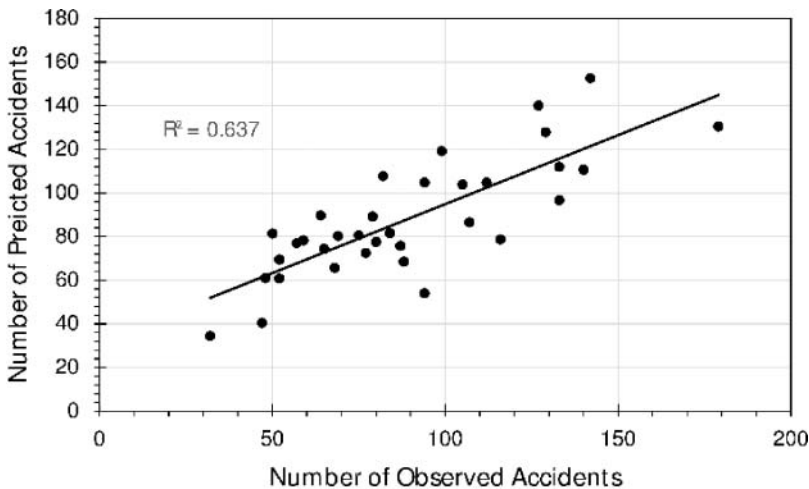


Figure 7. Scatter plot of observed accidents versus those predicted by the Bayesian logistic model.

As shown in Table 3, the length of a road segment (regression coefficient = 0.014, elasticity = 0.122), the number of vertical curves (regression coefficient = 4.797, elasticity = 0.999), the presence of a horizontal curve in a section (regression coefficient = 2.746, elasticity = 0.877), and AADT (regression coefficient = 3.01, elasticity = 0.881) determine accident occurrence, all at a 5% significance level. By contrast, other variables (i.e. radius of a horizontal curve, length of a vertical curve, and K value of a vertical curve) are statistically insignificant. The interpretations of these results, which are based on elasticity values (Couto & Ferreira 2011), are presented in the subsequent sections.

The roadway-geometric variables included in the model are the length of road segment, number of vertical curves in a road section, presence of a horizontal curve in a road section, average distance to the nearest access point from the centre of the road section, radius of a road horizontal curvature, length of a vertical curve in a road section, K value, and also is included one traffic related variable AADT. The analysis showed that the length of a road segment tends to increase the number of accidents. On the basis of the value of the estimated coefficient (mean = 0.014) associated with road segment length and the elasticity value (0.122) with respect to accidents, the number of accidents is expected to increase with the length of road segment. This assumption is consistent with the logic that shorter road segments are less likely to experience accidents than longer road segments because of decreased exposure. Moreover, this finding is consistent with that of other studies (Anastasopoulos et al. 2009; Wang et al. 2009, 2011; El-Basyouny and Sayed 2010; Couto and Ferreira 2011). In addition, determining road segment length from other geometric parameters (i.e. homogeneous segments) results in varying segment lengths across observations. It is also found that the number of vertical curves in each homogeneous section tends to increase accident frequency. This parameter is statistically significant and positively associated with accident frequency. This finding is consistent with those of other studies (Anastasopoulos & Mannering 2009). In addition, by comparing the elasticity values of the variables included in the proposed model, the number of vertical curves exhibits the highest value. Therefore, this variable is the most risky factor that contributes to accidents. One possible and logical explanation for this finding is that a large number of vertical curves reduce the sight distances on the sections where the vertical curves are located. Thus, drivers are more likely to experience accidents. Improving the vertical alignment of a section is one way to reduce accidents. However, improving vertical alignment is extremely costly. Accordingly, other improvements, such as optimizing the vertical alignment/curve with respect to accident frequencies and vehicle dynamics, are appreciated.

In addition, this study found that the presence of a horizontal curve in a road section increases accident frequency. Evidently, drivers are more likely to commit errors and lose control of a vehicle in the presence of a horizontal curve. One possible solution to reduce the effects of this variable is to use horizontal curves with larger radius and to design distributed super elevations on the horizontal curves. Another possible solution that can be practiced to reduce the effects of this variable is to use a good pavement material that helps improve the control of car by the drivers. On the other hand, an increase in the distance to the nearest access point decreases accident frequency. Greibe (2003) found that access points had no significant effect on a two-lane roadway but had a significant effect on multi-lane roadways. This is the illustrative of the fact that drivers close to an access point are usually less attentive at merging and diverging areas and tend to change their speed. Moreover, the current research indicated that increasing the traffic volume (AADT) tends to decrease the frequency of traffic accidents. Crash frequency increases in congestion because of the increased interactions among vehicles.

The remaining geometric design features are found to be statistically insignificant at 5% confidence level. The estimated regression coefficient of the radius of a road horizontal curvature is 0, suggesting that there is no effect for this parameter on the frequency of traffic accidents. According to the literature review, the main reason for the insignificant relationship between the radius of a horizontal curvature and accidents is attributed to the combined effects of road curvature and the limited variation in the degree of curvature (Milton & Mannering 1998; Wang et al. 2009). Wang

et al. (2009) used road segments with a standard deviation of less than 6 km. In the case of the current study, the standard deviation of road segment length used is larger than 44 km (Table 1) because of the fact that the approximately straight segments have an extremely high degree of curvature value. Although high variation was found in the radius of the horizontal curvature, the result of the model estimates showed that this variable remained insignificant at a 5% confidence level. Length of a vertical curve in a road section is intended to assess the effects of the length of vertical curves in road segments on accident frequency. It is mostly used during the process of designing vertical alignments, and thus, understanding its effects on the number of accidents is important to improve the design of vertical alignments for urban expressways. On the basis of the elasticity value associated with this variable, our finding indicates that an increase in the length of vertical curves in road segments slightly increases the number of accidents. However, this variable is found to be statistically insignificant at a 5% confidence level. The reason for this finding is that the positive relationship of this variable with accidents is probably attributed to the effect of road segment length, given that the length of a vertical curve and that of a road segment are highly correlated. In road segments wherein no vertical curve is found, the length of a vertical curve is set to zero. This variable can also be affected by other geometries (i.e. presence of a vertical curve and number of vertical curves). Furthermore, the K value, which is defined as the horizontal distance in feet (meters), is required to make a 1% change in gradient. This variable is important and is commonly used to determine the minimum length of vertical curves in the road design process. The K value is a function of the algebraic difference between the upgrade and downgrade of a vertical curve (A) and the length of a vertical curve ($K = L/A$). The length of vertical curves is assumed to have no significant effect on accidents, as analyzed from the preceding variable. The K value is inversely correlated with the gradient of the vertical curves. Therefore, this finding is consistent with those of other researchers (Wang et al. 2009). Moreover, it supports the interpretation of the effects of the length of a vertical curve as explained earlier. Therefore, vertical gradients are positively associated with the number of accidents. Accordingly, an increase in K value tends to decrease the number of accidents, assuming that the length of a vertical curve has no significant effect on accidents.

In addition, the regression analysis was repeated for fixed-length road segments to show the difference of this method with the homogeneous segments. Table 4 shows the results of coefficients estimates by the geometric regression model. Note that the variable, length of road segment was discarded from the analysis because it is constant throughout all the segments. Based on the analysis, the following variable were statistically significant at 5% confidence level: number of vertical curves, presence of a horizontal curve in a road segment, and traffic volume (AADT) and they are shown in Table 4. The remaining variables were found to be statistically insignificant at 5% confidence level. Comparing the results of fixed-length and homogenous segment methods, the finding is almost same. This was confirmed by the Mann-Whitney test. The test value (U) was found to be 19.5 and by comparing this value with the expected value (24.5), the estimated P -value (two-tailed) was found to be 0.565 because the estimated P -value (0.565) is greater than the set alpha (0.05). This suggests that there is no significant difference between the estimated regression based on the two methods (i.e. homogeneous segments and fixed-length method).

Table 4. Coefficients and elasticity estimates for the explanatory variables based on fixed-length segments.

Explanatory variable	Coefficient	Mean	Elasticity
Intercept	Beta0	2.35*	–
Radius of horizontal curve	Beta1	–3.05E-06	–0.00146
Length of vertical curve	Beta2	9.66E-04	0.01122
K value	Beta3	–0.015	–0.0529
Number of vertical curves	Beta4	11.09*	0.987
Presence of horizontal curve	Beta5	7.87*	0.974
Distance to nearest access point	Beta6	0.0014	0.0351
AADT	Beta7	12.89*	0.999

*Statistically insignificant at the 5% level.

6. Conclusion

In this study, very high-resolution laser scanning data were investigated for road geometric modeling and its application for accident frequency analysis. A geometric regression model, combined with various geometric design features extracted from laser scanning data, was used to develop associations between road geometry and traffic accidents that occurred on NSE in Malaysia. The development of an APM was followed by an elasticity analysis, which was conducted to understand the effects of the geometric design features included in the model on the number of accidents.

In literature, several studies have indicated that accurate road geometric parameters and accident data are critical for accurate association development between explanatory factors and accidents. Data collection on high-speed expressways is a difficult and challenging task because of the high volume of traffic. This study demonstrates that laser scanning systems can provide an easy and efficient method for collecting transportation data, particularly for accident analysis. Data collected by laser scanning systems can be used to model road environments in 3D with high accuracy. The road geometric model developed for NSE in Malaysia can be used to derive various geometric variables, particularly those related to vertical and horizontal alignments. Accurate road geometric variables can improve our understanding on the causes of traffic accidents occurring on high-speed highways and expressways in modern megacities. Improvements in the APMs can be conducted using two approaches. The first approach involves using highly accurate accident data and explanatory factors, whereas the second approach involves using robust statistical models that consider missing data and the combined effects of various explanatory variables. Therefore, using laser scanning systems is suggested for collecting transportation data for accident analysis.

The three main factors that contribute to the number of accidents are road environments, vehicle factors, and human factors. Although road environment is not a major factor, investigating the effects of road environment on accidents is important to improve road designs and develop road design policies and manuals. Using only road geometric parameters, the model based on a Bayesian logistic approach proposed in this study can predict the number of accidents on NSE in Malaysia with reasonable accuracy ($R^2 = 63.7\%$). The aim of this study is not to propose a model that can predict the number of accidents but to propose a model that can interpret the contributing factors. Interpreting the contributing factors is relatively more important than only predicting accidents. This method can improve planning and decision making to reduce accident frequency on such expressways.

The results of the elasticity analysis conducted in this study indicate that four main factors contribute significantly to the number of accidents. These factors are the length of a road segment (elasticity = 0.122), the number of vertical curves (elasticity = 0.999), the presence of a horizontal curve in a road section (elasticity = 0.877), the average distance to the nearest access point (elasticity = -0.035) and AADT (elasticity = 0.881) all at a 5% significance level. By contrast, other variables (i.e. radius of a horizontal curve, length of a vertical curve, and K value of a vertical curve) are statistically insignificant at a 5% significance level. On the other hand, the analysis of regression based on fixed-length segments shows relatively similar results analyzed by the Mann-Whitney test. Although most of the findings of this research are consistent with those of previous studies, using only geometric parameters may affect model estimates, and consequently, final interpretations. Therefore, including other traffic-, human-, and vehicle-related factors into the model is suggested for a more accurate development of APMs.


Acknowledgments


The authors acknowledge and appreciate the provision of LiDAR, traffic and accident related data by the PLUS Berhad. Also, the authors gratefully acknowledge the financial support from the UPM-PLUS industry project grant, vote number: 6300193.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Maher Ibrahim Sameen  <http://orcid.org/0000-0003-0490-7796>

Biswajeet Pradhan  <http://orcid.org/0000-0001-9863-2054>

References

- Abdel-Aty MA, Radwan AE. 2000. Modeling traffic accident occurrence and involvement. *Accident Anal Prev.* 32:633–642.
- Ackaah W, Salifu M. 2011. Crash prediction model for two-lane rural highways in the Ashanti region of Ghana. *IATSS Res.* 35:34–40.
- Anastasopoulos PC, Mannering FL. 2009. A note on modeling vehicle accident frequencies with random-parameters count models. *Accident Anal Prev.* 41:153–159.
- Anastasopoulos PC, Mannering FL, Shankar VN, Haddock JE. 2012. A study of factors affecting highway accident rates using the random-parameters tobit model. *Accident Anal Prev.* 45:628–633.
- Ayati E, Abbasi E. 2011. Investigation on the role of traffic volume in accidents on urban highways. *J Safety Res.* 42:209–214.
- Brink AD. (1996). Using spatial information as an aid to maximum entropy image threshold selection. *Pattern Recogn Lett.* 17:29–36.
- Cafiso S, Di Graziano A, Di Silvestro G, La Cava G, Persaud B. 2010. Development of comprehensive accident models for two-lane rural highways using exposure, geometry, consistency and context variables. *Accident Anal Prev.* 42:1072–1079.
- Cameron AC, Trivedi PK. 2013. *Regression analysis of count data* (Vol. 53). New York: Cambridge University Press.
- Couto A, Ferreira S. 2011. A note on modeling road accident frequency: A flexible elasticity model. *Accident Anal Prev.* 43:2104–2111.
- Deublein M, Schubert M, Adey BT. 2014. Prediction of road accidents: comparison of two Bayesian methods. *Struct InfrastructEng.* 10:1394–1416.
- Deublein M, Schubert M, Adey BT, Köhler J, Faber MH. 2013. Prediction of road accidents: A Bayesian hierarchical approach. *Accident Anal Prev.* 51:274–291.
- Dong N, Ma M, Huang H, Zheng L. 2015. Support vector machine in crash prediction at the level of traffic analysis zones: Accessing the spatial proximity effects. *Accident Anal Prev.* 82:192–198.
- El-Basoumy K, Sayed T. 2010. Application of generalized link functions in developing accident prediction models. *Safety Sci.* 48:410–416.
- Fernandes A, Neves J. 2013. An approach to accidents modeling based on compounds road environments. *Accident Anal Prev.* 53:39–45.
- Global status report on road safety [Internet]. 2015. [cited 2015 Oct 23]. Available from: http://www.who.int/violence_injury_prevention/road_safety_status/2015/en/
- Gomes SV. 2013. The influence of the infrastructure characteristics in urban road accidents occurrence. *Accident Anal Prev.* 60:289–297.
- Greibe P. 2003. Accident prediction models for urban roads. *Accident Anal Prev.* 35:273–285.
- Hilbe JM. (2014). *Modeling count data*. New York: Cambridge University Press.
- Hosseinpour M, Yahaya AS, Sadullah AF. 2014. Exploring the effects of roadway characteristics on the frequency and severity of head-on crashes: Case studies from Malaysian Federal Roads. *Accident Anal Prev.* 62:209–222.
- Huang H, Abdel-Aty M. 2010. Multilevel data and Bayesian analysis in traffic safety. *Accident Anal Prev.* 42:1556–1565.
- Kardoulas NG, Bird AC, Lawan AI. 1996. Geometric correction of SPOT and Landsat imagery: a comparison of map- and GPS-derived control points. *Photogramm Eng Rem S.* 62:1173–1177.
- Karlaftis M, Vlahogianni E. 2011. Statistical methods versus neural networks in transportation research: Differences, similarities and some insights. *Transport Res C Emerg Technol.* 19:387–399.
- Li X, Lord D, Zhang Y, Xie Y. 2008. Predicting motor vehicle crashes using support vector machine models. *Accident Anal Prev.* 40:1611–1618.
- Manan MMA, Várhelyi A. 2015. Motorcyclists' road safety related behavior at access points on primary roads in Malaysia – A case study. *Safety Sci.* 77:80–94.
- Milton J, Mannering F. 1998. The relationship among highway geometrics, traffic-related elements and motor-vehicle accident frequencies. *Transportation* 25:395–413.

- Pei X, Wong S, Sze N-N. 2011. A joint-probability approach to crash prediction models. *Accident Anal Prev.* 43:1160–1166.
- Poullis C, You S. 2010. Delineation and geometric modeling of road networks. *ISPRS J Photogramm Remote Sensing.* 65:165–181.
- Wang C, Quddus MA, Ison S. 2009. The effects of area-wide road speed and 7 curvature on traffic casualties in England. *J Transp Geogr.* 17:385–395.
- Wang C, Quddus MA, Ison SG. 2011. Predicting accident frequency at their severity levels and its application in site ranking using a two-stage mixed multivariate model. *Accident Anal Prev.* 43:1979–1990.