Scientia Iranica B (2015) 22(4), 1594–1603

Sharif University of Technology
Scientia Iranica
Transactions B: Mechanical Engineering
www.scientiairanica.com

Spatial analysis methods for identifying hazardous locations on expressways in Korea

S.h. Park[a], K. Jang[b], D.-K. Kim[c], S.-Y. Kho[c] and S. Kang[d,*]

a. *Keimyung University, 1095 Dalgubeol-daero, Daegu 704-701, Republic of Korea.*
b. *Korea Advanced Institute of Science and Technology, N-22 #206, Daehak-ro, Yuseong-gu, Daejeon 305-701, Republic of Korea.*
c. *Seoul National University,1Gwanak-ro, Gwanak-gu, Seoul 151-744, Republic of Korea.*
d. *Korea University,145 Anam-ro, Seongbuk-gu, Seoul 136-713, Republic of Korea.*

Received 22 May 2014; received in revised form 26 February 2015; accepted 11 April 2015

KEYWORDS
Hazardous locations;
Spatial interaction;
Geographic information system;
Geographically weighted regression;
Kernel density estimation.

Abstract. Identifying hazardous locations on highways is a fundamental step in safety improvement programs and projects since it provides decision makers with a basis for allocating budgets and other resources in a cost-effective manner. Extensive research has been conducted to identify such locations. However, most studies have ignored the spatial characteristics of crash occurrences and the relative significance of injury severity. In this study, we develop a procedure for identifying hazardous locations on expressways based on Geographically Weighted Regression (GWR) and Equivalent Property Damage Only (EPDO). GWR is a spatial regression method that can reflect spatial dependency and heterogeneous relationships between crash occurrences and other explanatory variables. EPDO is a comprehensive measure of crash occurrences weighted by the level of injury severity. We apply this procedure to a case study in Gyeongbu Expressway in Korea. The findings from our case study show that the procedure can identify hazardous locations on roadways while reflecting crash frequency and injury severity simultaneously with the comprehensive measure.

1. Introduction

Identifying hazardous locations (hotspots) and understanding the process of crash occurrences in those locations are important for the appropriate allocation of resources for safety improvements [1,2]. Failure to identify true hazardous locations such as false positives (i.e. identifying sites for safety improvements that should not have been selected) or false negatives (i.e. not identifying sites that should have been selected) reduces the effectiveness of safety improvement projects by wasting resources.

Hazardous locations on highways have been defined or ranked by various crash outcome measures, including crash rate (crashes per vehicle miles traveled), frequency, severity, density, or a combination of these measures within divided highway segments. Many recent studies have suggested determining these locations based on the amount of excessive risk, which can be estimated by the difference between observed and expected counts [3-5]. However, the previous models are limited in their ability to 1) measure crash outcomes that can reflect both severity and occurrence and 2) reflect the effect of spatial interaction among traffic crashes that occur in nearby segments because the regression models are specified based on the assumption of independence between samples.

*. *Corresponding author. Tel.: +82-2-3290-4862; Fax: +82-2-928-7656*
E-mail addresses: shpark@kmu.ac.kr (S.H. Park); kitae.jang@kaist.ac.kr (K. Jang); dongkyukim@snu.ac.kr (D.-K. Kim); sykho@snu.ac.kr (S.-Y. Kho); s_kang@korea.ac.kr (S. Kang)

In this study, we use the Equivalent Property Damage Only (EPDO) method, which considers the total number of crashes weighted by the severity of each crash [6]. Geographically Weighted Regression (GWR) is then applied to modeling EPDO as an outcome variable. GWR is used to model spatially varying relationships in the geographic space [7]. Based on these two methods, this study proposes a procedure for identifying hazardous locations on expressways in Korea. The result is then compared to hazardous locations identified by Kernel Density Estimation (KDE), which is an effective method for analyzing the first order properties of a point event distribution [8,9].

In practice, hotspots are identified by the difference between the actual and the expected number of crashes, which can later be used to estimate benefits of safety improvement projects. Since the KDE cannot provide this quantitative selection criterion, it may not be used independently for selecting safety improvement projects. However, it is an efficient method for visualizing geographically concentrated high-accident segments, and it can identify the list of hotspots that are quite comparable to that of the GWR model.

The remainder of this paper is organized as follows. In the second section, we discuss the importance of spatial interaction in analyzing traffic crashes. The third section explains the measure of the EPDO and the estimation procedure for GWR. The proposed approach is applied to an actual highway, Gyeongbu Expressway in Korea, in the fourth section. The final section provides concluding remarks and suggestions for future research.

2. Spatial interaction in traffic crashes

Traffic accidents are spatial events occurring over a network in geographic space. They are caused by various factors such as human behavior, the mechanical failure of vehicles, roadway geometry, and environmental conditions. Spatial interaction among these factors, which are often missing in collision reports, can also be a significant latent factor in accidents. Therefore *Tobler's first law of geography* [10], which states that objects in geographic space are not distributed randomly but interact with each other, also holds true for crash occurrences, resulting in spatial dependency and spatial heterogeneity.

Spatial dependency refers to a certain degree of redundancy in the additional information that is provided by nearby locations within a geographic space [11]. Spatial dependency leads to the spatial autocorrelation problem in conventional statistics; however, this can be viewed as evidence for important spatial processes [12]. In contrast, spatial heterogeneity is characterized by a spatial or regional dissimilarity between the locations of objects within a geographic space. The results of any analysis over a limited area can be different from the results that would be obtained for other areas. These concepts tend to affect almost any kind of spatial analysis conducted on geographic data, including crashes on highways [13].

Previous studies of crash prediction models have analyzed crash causal factors from a traditional statistical standpoint, without sufficient consideration of spatial interaction. The Safety Performance Function (SPF), for example, estimates the expected number of traffic accidents per unit time interval using independent variables, such as traffic flow rates and geometric design features [14,15]. The SPF assumes that 1) the rate of traffic collisions along a highway is spatially uncorrelated, 2) the rate at which collisions occur within the segment remains constant, and 3) the factors causing high collision rates reside within the segment [16,17]. The first assumption can be invalidated by the spatial autocorrelation among accidents, as examined in the fourth section of this paper. The second and third assumptions are also invalid because crash risks vary by location even within the same highway segment, and the factors causing a high number of collisions may spread out over contiguous segments.

Other studies such as those of Miaou and Lum [18] and Shankar et al. [19] developed models based on homogeneous road sections divided by explanatory variables, such as the geometric features of the site, traffic volume, and other environmental features. However, these models could not fully examine the spatial relationships between traffic accidents because of the limits of traditional statistical analysis methods.

3. Methodology

3.1. Equivalent Property Damage Only (EPDO)

Since the impacts of crashes can vary according to their accompanying injuries, it is important for a crash outcome measure to reflect injury severity. We propose the EPDO method, which assigns severity weights to individual crashes. Higher levels of injuries are given higher weights based on the ratio of their average crash costs to Property Damage Only (PDO) crashes [20].

In this study, we use a severity index using the EPDO method, which is calculated as follows:

$$\mathrm{SI}(n) = W_F(n) \cdot C_F(n) + W_I(n) \cdot C_I(n) + C(n), \quad (1)$$

where:

$\mathrm{SI}(n)$: Severity index at location n;

$W_F(n)$: Weight for a fatal crash at location n;

$W_I(n)$: Weight for an injury crash at location n;

$C_F(n)$: The number of fatal crashes at location n;

$C_I(n)$: The number of injury crashes at location n;

$C(n)$: The number of PDO crashes at location n.

By assigning greater weight to severe injury crashes, bias due to the underreporting of minor injuries and PDO crashes can be partially corrected [6]. Furthermore, since EPDO is not only a comprehensive measure but also a continuous variable, it can obviate issues in other conventional count models for crash occurrences and allows more advanced and appropriate statistical analyses for safety evaluation [21].

3.2. Geographically weighted regression method

In the Ordinary Least Square (OLS) regression, parameters are estimated globally. Once estimated, the same parameter values are applied over all highway segments, although the influence of some independent variables (e.g. geometric attributes or traffic volumes) on the dependent variable (e.g. crash frequencies or severity) may vary across space. By not reflecting spatial variability, the estimated model may include some biases and result in low explanatory power of a model [22]. In GWR, local parameters are estimated for each location or highway segment with different weights for observations relative to those positions. In other words, in estimating parameters for a model at a certain location, observations that are made at nearby locations should have greater weights in the estimation than observations that are made farther away [23].

The general form of the GWR model is modified from the OLS and can be written as:

$$y_i = \beta_{i0} + \sum_k \beta_{ik} x_{ik} + \epsilon_i, \tag{2}$$

where:

y_i : Dependent variable in highway segment $i \in L$, where L is a set of all highway segments in the route;

x_{ik} : Independent variable of the kth parameter in highway segment i;

β_{ik} : kth parameter in highway segment i;

ϵ_i : Error term in highway segment i.

The estimation procedure of the GWR model is composed of four parts: i) Select h (bandwidth), ii) Calculate weight matrix, W_i, iii) Calculate coefficient matrix, $\hat{\beta}_i$, and iv) Calculate model fitness, Akaike information criterion with a sample size correction (AICc). This procedure is repeated until the AICc is minimized. At minimum AICc, the bandwidth, h, is determined.

$\hat{\beta}_i$ is a vector of local estimated parameters for a given highway segment i, where $\hat{\beta}_i^T = \{\hat{\beta}_{i1}, \hat{\beta}_{i2}, ..., \hat{\beta}_{ik}, ...\}$. This is calculated with weights given on the location relative to the other observations in the dataset. The estimator can be represented as follows:

$$\hat{\beta}_i = \left(X^T W_i X\right)^{-1} X^T W_i y, \tag{3}$$

where $\mathbf{X}$ is the matrix of the independent variables, x_{ik}, y is the vector of observed dependent variable values, and W_i is a $p \times p$ matrix of weights relative to the position of i in the study area, where p is the total number of segments. Off-diagonal elements of W_i are zero, and its diagonal element in j-th row is represented by a Gaussian form as:

$$W_i(j) = \frac{1}{\sqrt{2\pi}} e^{-0.5(d_{ij}/h)^2}, \tag{4}$$

where $W_i(j)$ is the geographical weight of the data point i related to the regression point of segment $j \in L$, where L is the set of total segment; d_{ij} is the distance between data point i, and the regression point of segment j, and h is the bandwidth. In this study, the data point is the corresponding crash location, and a middle point of each roadway segment is used as the regression point. At minimum AICc, the bandwidth, h, is 70,034 m.

Hazardous locations are selected based on the residuals-highway segments that have residuals greater than the predetermined threshold from the estimated value of GWR [24,25]. The threshold value can usually be represented as a certain statistically significant level. For example, the California Department of Transportation [26] and Sung [27] used 0.5% and 5% significance levels, respectively, and Kononov and Allery [28] set the threshold to be 1.5 times the standard deviation. The threshold value may depend on site-specific conditions or on a budget size for the safety improvement program.

4. Application to Gyeongbu expressway in Korea

4.1. Data description

Gyeongbu Expressway is a 416 km expressway connecting Seoul, the capital of Korea, and Busan, the second largest city in Korea. Data for a total of 842 crashes, which occurred on the northbound lanes of the expressway from 2006 through 2008, are used in this study (see [29,30] for details of the site description and the database of geocoded collision records). We divided the expressway into 524 segments, with an average length of 794 m for the estimation of both OLS and GWR, according to their horizontal curve

Table 1. Description of variables.

Category	Variables	Description
Dependent variable	EPDO	Sum of scores for each crash multiplied by a weight based on crash severity EPDO = (1 × crashes with property damage only) + (3 × number of injury crash) + (12 × number of fatal crash)
Explanatory variable	NumLane	The number of lanes in a segment
	Bridge	The number of bridges in a segment
	Tunnel	The number of tunnels in a segment
	Camera	The number of speed cameras in a segment
	OffRamp	The number of off-ramps in a segment
	OnRamp	The number of on-ramps in a segment
	Restarea	The number of rest areas in a segment
	TG	The number of toll booths in a segment
	EXPO	Exposure of a segment to traffic $\text{EXPO} = \frac{D\times 365}{10^6} \times \left(\sum_{g=2006}^{2008}\left(\text{AADT}_g \times \frac{\text{Freq}_g}{\sum_{h=2006}^{2008}\text{Freq}_h}\right)\right)$ where D is the length of a road segment, AADT_g is annual average daily traffic for year g, and Freq_g is crash frequency in year g.
	HR	Radius of horizontal curve (m)

attributes, as suggested by Kwak et al. [31] and Park et al. [32] who, using comparable data sets, showed horizontal curvature to be an influential segmentation criteria. The dependent and explanatory variables are described in Table 1.

In general, EPDO weights are estimated based on the societal costs of fatal, injury, and PDO crashes. Therefore, they may vary with time and location. In this study, we used 12 for a fatal crash and 3 for an injury crash, which are Equivalent Property Damage Only (EPDO) values officially employed by the Korea Road Traffic Authority, an organization that is affiliated with the National Police Agency (NPA).

4.2. Verification of spatial relationship

Spatial autocorrelation can be defined as the existence of a positively (or negatively) systematic pattern in the spatial distribution of the variable. Contiguous spaces have a positive spatial autocorrelation if they are more alike, and vice versa. This is important because most traditional statistics are based on the assumption that the values of observations in each sample are independent of one another.

Moran's autocorrelation coefficient (often denoted as I) [33] is an indicator of global spatial autocorrelation, and it is used to measure the spatial autocorrelation of crash occurrences [34]. It compares the value of the variable at any one location with the value at all other locations. Moran's I can be calculated as follows:

$$I = \frac{p\sum_i\sum_j w_{ij}(R_i-\bar{R})(R_j-\bar{R})}{(\sum_i\sum_j w_{ij})\sum_i(R_i-\bar{R})^2}, \tag{5}$$

where p is the total number of segments, and i and j are highway segment indices. R_i is the difference between the observed and the estimated variable value (e.g. fatality rate, severity, or frequency) at location i, $\bar{R}$ is the mean of R_i, and w_{ij} is a distance-based weight represented by the reciprocal of the distance between segments i and j.

Given a set of accident points and associated attributes of the accidents, Moran's I evaluates whether the pattern expressed is clustered, dispersed, or random. The Z score of Moran's I is used to examine its statistical significance. To test whether we can reject the null hypothesis (i.e., there is no spatial clustering), the Z score is calculated as:

$$Z(I) = \frac{I-E(I)}{S_{E(I)}}, \tag{6}$$

where $E(I)$ is the expected value of Moran's I and $S_{E(I)}$ is an estimate of the theoretical standard deviation. To determine if the Z score is statistically significant, it is compared to the range of values for a particular confidence level [12].

The results of Moran's I and Koenker's studentized Breusch-Pagan (K-BP) statistics for OLS statistics are summarized in Table 2. Moran's I results show

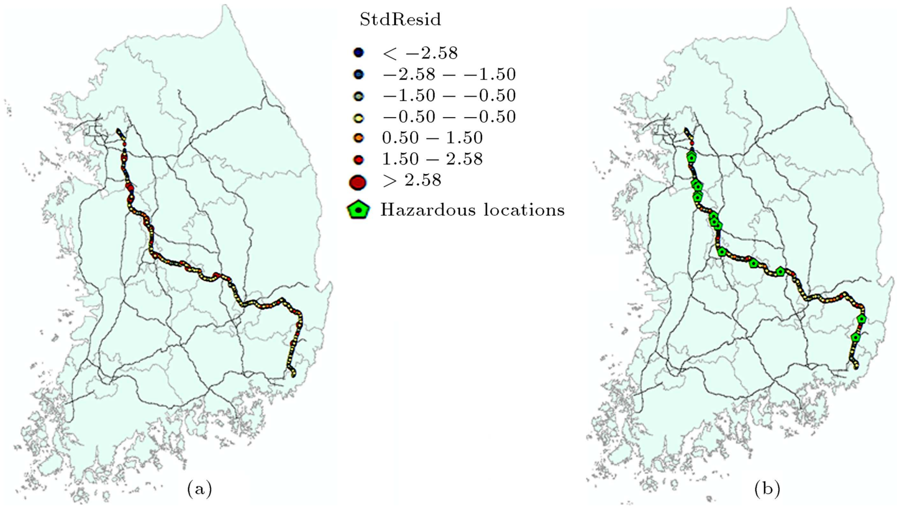

Figure 1. Identified hazardous locations by GWR Method: a) Symbolized by the value of standardized residual; and b) hazardous locations of which the standardized residuals are bigger than 2.58.

Table 2. A summary of statistics results of Moran's I test and OLS.

Test	Statistics name	Statistics value
Moran's I	Moran's Index	0.043
	Z-Score	2.012
	p-value	0.043
K-BP for OLS	K-BP	52.358
	K-BP_Prob	0.000000

that all Z-scores are greater than 1.96 and p-values are smaller than 0.05, which implies the existence of spatial autocorrelation with a 95% level of statistical significance.

The K-BP statistic shows whether the explanatory variables in the model have a consistent relationship to the dependent variable over the geographic space (i.e. spatial heterogeneity) [35]. For the OLS result, a K-BP probability of less than 0.05 indicates that the dependent variable has statistically significant heteroscedasticity and non-stationarity over the target area. Resulting from Moran's I test and K-BP statistic, the spatial interaction in crash data raises an important problem to be considered in accident analyses.

4.3. GWR estimation

The OLS and GWR results are compared in Table 3. Akaike Information Criterion (AIC) is a relative measure of performance that is used to compare statistical models [36]; the smaller AICc indicates a better goodness of fit. While EPDO shows similar AICc values for both models, both R-square and adjusted R-square values show that the GWR model is superior to the OLS model for explaining the dependent variable. This is as expected because GWR can better describe the different relationships between the dependent and explanatory variables at different locations.

Table 3. Comparison of OLR and GWR results.

Test	Statistics name	Statistics value
OLS	AICc	3,390.38
	R^2	0.477
	Adjusted R^2	0.467
GWR	AICc	3,390.46
	R^2	0.520
	Adjusted R^2	0.484

4.4. Identifying hazardous locations by GWR

The next step is to select hazardous locations using the GWR results. From the model with EPDO, 12 highway segments with residuals larger than 2.58 standardized residuals, representing high residual sections with a 1% significance level, are identified, as shown in Figure 1.

In Table 4, selected segments are ranked by the higher unit accident costs that imply the higher social costs from the accidents in the unit segment; thus, there could be more room for saving the social cost by a given investment for the safety improvement project.

4.5. Comparison of GWR and KDE results

In general, crash maps do not precisely reflect the crash concentrations of locations having more than one crash because the symbols for each of the crashes at one location lie on top of each other and thus are not shown distinctly [37]. Kernel density, on the other hand, calculates the magnitude of a crash frequency or severity per unit area from every crash point on the highway using a kernel function to fit a smoothly tapered surface to each point. The surface value is the highest at the location of the crash point, diminishing as it moves away from the point and reaching zero at the radius distance (or bandwidth) from the point [38].

In the KDE method, the highway route is first

Table 4. Identified hazardous locations by GWR

Priority	From (km)	To (km)	Length (m)	EPDO	Acc._cost* (USD)	Unit accident cost (USD/m)
1	307.380	307.891	511	36	1,341,469	2,625
2	310.978	311.380	402	24	842,726	2,096
3	343.015	343.720	705	28	916,462	1,300
4	268.780	269.720	940	26	919,779	978
5	55.270	55.760	490	25	461,852	943
6	389.877	390.720	843	31	589,565	699
7	344.960	346.900	1,940	38	1,333,483	687
8	188.710	190.110	1,400	28	912,701	652
9	226.320	227.820	1,500	25	898,804	599
10	386.920	388.510	1,590	29	921,068	579
11	298.716	300.375	1,659	28	940,166	567
12	353.940	356.340	2,400	31	933,960	389

* Accident cost is a social cost for accidents represented by the sum of the property damage costs, the injury costs, and the fatality costs (Road Traffic Authority of Korea, 2011).

divided into small cells, and the kernel function value at each cell is calculated. The value increases as it gets closer to accident locations, and the location having a large cell value can be regarded as an accident-prone location.

In this study, we propose a severity-weighted kernel function, which is a product of a severity index and a distance function. The severity index in Eq. (1) defines the kernel function at cell a, represented by a Gaussian form:

$$K_a = (2\pi)^{-1/2} \sum_{n_a} \exp\left(-\frac{(d(a,n_a))^2}{2b^2}\right) \mathrm{SI}(n_a), \quad (7)$$

where K_a is a kernel function value at cell a, n_a is an index of crash location within a bandwidth from cell a, $d(a, n_a)$ is the distance from crash location n_a to cell a, and b is a bandwidth for KDE.

Two key inputs of the KDE method, cell size and bandwidth, are determined for the clear representation of hotspots. Using large cell size and bandwidth results in data aggregation in distant segments, so that blurred images of hotspots are provided, resulting in a false positive (i.e. unnecessary identification of low risk segments). Meanwhile, when a small cell size and bandwidth are used, statistical fluctuations in the data cannot be removed, so that an increased number of hotspots are identified. Therefore, in this study, we set the cell size and bandwidth at 100 m and 500 m to match the length of the safety improvement project in Korean expressways. For the comparable number to the selected locations by the GWR method, 15 segments with highest kernel function values are selected as shown in Table 5.

Figure 2 and Table 6 show locations selected by both the GWR and KDE methods. In Table 6, the

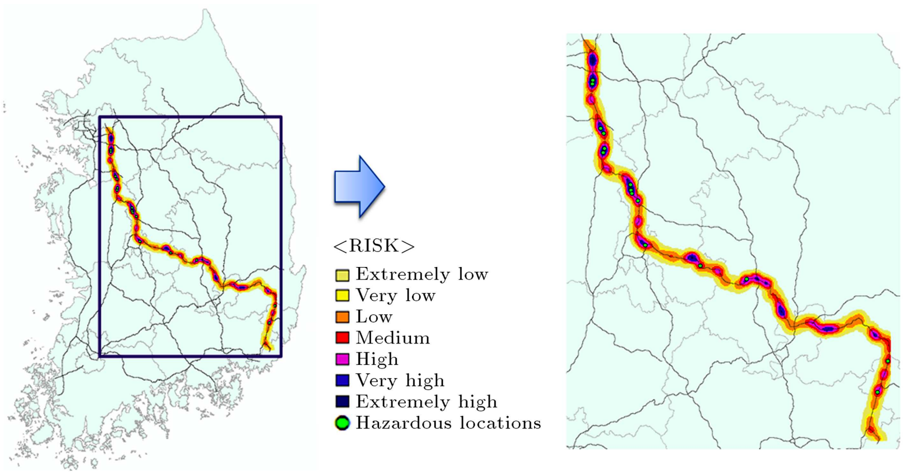

Figure 2. Kernel density map that overlaps hazardous locations identified by GWR.

Table 5. Selected locations by KDE method.

Priority	From (km)	To (km)	Length (m)	EPDO	Acc._cost (USD)	Unit accident cost (USD/m)
1	307.3	307.8	500	36	1,341,469	2,683
2	343.2	343.6	400	28	916,462	2,291
3	299.2	299.6	400	21	898,184	2,245
4	189.5	189.9	400	24	885,905	2,215
5	55.4	55.8	400	24	858,199	2,145
6	359.6	360.3	700	42	1,358,703	1,941
7	388.3	388.8	500	27	906,350	1,813
8	346	346.5	500	25	861,690	1,723
9	35.2	36	800	27	1,323,150	1,654
10	402.5	403.2	700	41	1,069,763	1,528
11	269.1	269.8	700	29	944,522	1,349
12	233.2	233.9	700	30	910,410	1,301
13	356	356.4	400	22	519,349	1,298
14	389.8	390.4	600	31	592,205	987
15	386.8	387.3	500	19	486,179	972

Table 6. Top 12 commonly selected segments by both the GWR and KDE models.

GWR					KDE				
Rank	From (km)	To (km)	Length (m)	Unit accident cost (USD/m)	From (km)	To (km)	Length (m)	Unit accident cost (USD/m)	Rank
1	307.380	307.891	511	2,625	307.3	307.8	500	2,683	1
2	310.978	311.380	402	2,096					
3	343.015	343.720	705	1,300	343.2	343.6	400	2,291	2
4	268.780	269.720	940	978	269.1	269.8	700	1,349	11
5	55.270	55.760	490	943	55.4	55.8	400	2,145	5
6	389.877	390.720	843	699	389.8	390.4	600	987	14
7	344.960	346.900	1,940	687	346.0	346.5	500	1723	8
8	188.710	190.110	1,400	652	189.5	189.9	400	2,215	4
9	226.320	227.820	1,500	599	233.2	233.9	700	1,301	12
10	386.920	388.510	1,590	579	388.3	388.8	500	1,813	7
					386.8	387.3	500	972	15
11	298.716	300.375	1,659	567	299.2	299.6	400	2245	3
12	353.940	356.340	2,400	389	356.0	356.4	400	1,298	13
					359.6	360.3	700	1941	6
					35.2	36	800	1654	9
					402.5	403.2	700	1528	10

location of each segment in the same row lies on a similar position in the expressway. As a result, 12 out of 15 segments identified by the KDE method are matched to 11 out of 12 locations selected by the GWR method. Here, highway segments are divided by horizontal curve attributes in the GWR method, and the lengths may vary from hundreds of meters to over 10 kilometers. In this analysis, most GWR segments are longer than KDE segments, and thus, the unit accident costs of GWR segments are generally lower than those of KDE segments. This implies that the GWR segments may include relatively lower risk sections, but at the same time, this may reduce false negatives by covering the high-risk sections with a relatively wide range of roadway segments. On the other hand, with tighter and more regular segment lengths, higher unit accident costs of KDE may increase the resource effectiveness.

5. Conclusion and future study

Numerous efforts have been made to identify factors influencing crash occurrences and roadway safety by developing crash prediction models based on explanatory variables, such as the geometric features of the site, traffic volume, and other environmental features. However, since these models assume that every crash occurs independently and that the rate at which crashes occur within a segment remains constant, they fail to fully reflect the interaction between adjacent sections of a continuous roadway.

In this study, the GWR and KDE methods were used to identify hazardous highway sections by considering the effect of spatial dependency and spatial heterogeneity on the outbreak of traffic accidents. The GWR method shows better analytical outcomes of spatial data than the conventional OLS method because it estimates different local regression equations depending on spatial characteristics, so the significance of variables and their influence can be considered separately at each highway segment. The KDE method can help identify crash-clustered areas through the use of a visualized kernel density map, and thus, it was used in this study to confirm the effectiveness of the GWR method.

For suggested approaches, the EPDO variable was tested as a dependent variable for a better evaluation of the seriousness of accidents, whereas most of the existing crash-frequency-based studies have considered all crashes to be identical, regardless of their level of severity.

In the case study of Gyeongbu Expressway in Korea, we verified that highway accidents are spatially correlated and that the suggested spatial analysis model can explain the distribution and relationship of crashes with better goodness-of-fit with the actual crash data. Most of the segments selected by the GWR and KDE methods overlapped each other. Commonly selected segments are prioritized by the unit accident cost and recommended as candidate sites for future safety improvement projects.

Although the procedure suggested in this study has reduced the theoretical limits of the assumptions of existing crash prediction models, further research on the GWR model can enhance the performance of the procedure. More studies should be performed to determine the proper criteria for segmenting highways for GWR and the modifiable areal unit problem, which may cause altered analysis results by the level of aggregation of the spatial data. Furthermore, more elaborate models, such as the Poisson-gamma model with random effects, should be developed and applied to the procedure to enhance its effectiveness in identifying hazardous locations.

In addition, before-and-after studies of safety improvement projects for suggested sites would be helpful for evaluating the effectiveness of suggested methodologies. The enhanced safety level and the economic and social benefits will validate the importance of considering the spatial relationship of accidents in identifying hazardous locations on highways.

Acknowledgment

This work was supported by Jungseok Logistics Foundation Grant and National Research Foundation of Korea (2009-0075811, 2010-0029446, and 2015R1C1A1A02037285).

Abbreviations

GWR	Geographically Weighted Regression
GIS	Geographic Information Systems
EPDO	Equivalent Property Damage Only
PDO	Property Damage Only
KDE	Kernel Density Estimation
SPF	Safety Performance Function
OLS	Ordinary Least Square
AIC	Akaike Information Criterion
K-BP	Koenker's studentized Breusch-Pagan

References

1. Anderson, T.K. "Kernel density estimation and K-means clustering to profile road accident hotspots", *Accid. Anal. and Prev.*, **41**, pp. 359-364 (2009).
2. Jang, K., Park, S.H., Kang, S., Song, K.H., Kang, S. and Chung, S. "Evaluation of pedestrian safety: Pedestrian crash hot spots and risk factors for injury severity", *Transp. Res. Rec.: J. of the Transp. Res. Board*, **2393**, pp. 104-116 (2013).
3. Persaud, B., Lyon, C. and Nguyen, T. "Empirical Bayes procedure for ranking sites for safety investigation by potential for improvement", *Transp. Res. Rec.: J. of the Transp. Res. Board*, **1665**, pp. 7-12 (1999).
4. Hauer, E., Allery, B.K., Kononov, J. and Griffith, M.S. "How best to rank sites with promise", *Transp. Res. Rec.: J. of the Transp. Res. Board*, **1879**, pp. 48-54 (2004).
5. El-Basyouny, K. and Sayed, T. "Comparison of two negative binomial regression techniques in developing accident prediction models", *Transp. Res. Rec.: J. of the Transp. Res. Board*, **1950**, pp. 9-16 (2006).
6. OH, J., Washington, S.P. and Lee, D. "Property damage crash equivalency factors to solve crash frequency-severity dilemma: Case study on South Korean rural roads", *Transp. Res. Rec.: J. of the Transp. Res. Board*, **2148**, pp. 83-92 (2010).

7. Fotheringham, A.S., Brunsdon, C. and Charlton, M., *Geographically Weighted Regression-the Analysis of Spatially Varying Relationships*, John Wiley & Sons Ltd., Chichester, UK (2002).

8. Flahaut, B., Mouchart, M., Martin, E.S. and Thomas, I. "The local spatial autocorrelation and the kernel method for identifying black zones: A comparative approach", *Accid. Anal. and Prev.*, **35**, pp. 991-1004 (2003).

9. Erdogan, S., Yilmaz, I., Baybura, T. and Gullu, M. "Geographical information systems aided traffic accident analysis system case study: City of Afyonkarahisar", *Accid. Anal. and Prev.*, **40**, pp. 174-181 (2008).

10. Tobler, W. "A computer movie simulating urban growth in the Detroit region", *Econ. Geogr.*, **46**(2), pp. 234-240 (1970).

11. Goodchild, M.F., Haining, R., Wise, S. and 12 others. "Integrating GIS and spatial data analysis: Problems and possibilities", *Int. J. of Geogr. Inform. Sci.*, **6**(5), pp. 407-423 (1992).

12. ESRI, ArcGIS version 9.3 User Manual, Environmental Systems Research Institute, Redlands, CA (2008).

13. De Smith, M.J., Goodchild, M.F. and Longley, P.A., *Geospatial Analysis: A Comprehensive Guide to Principles, Techniques and Software Tools*, Matador, Leicester, UK (2008).

14. Zhong, L.D., Sun, X.D., He, Y.L., Zhong, X.M. and Chen, Y.S. "Safety performance function for freeway in China", *88th Annual Meeting of the Transp. Res. Board*, Washington, D.C., USA (2009).

15. Jang, K., Kang, S., Seo, J. and Chan, C. "Cross-section designs for the safety performance of buffer-separated high-occupancy vehicle lane", *J. of Transp. Eng.*, **139**(3), pp. 247-254 (2012).

16. Chung, K., Ragland, D.R., Madanat, S. and Oh, S.M. "The continuous risk profile approach for the identification of high collision concentration locations on congested highways", In *Transportation and Traffic Theory 2009: Golden Jubilee*, W.H.K. Lam, S.C. Wong, and H.K. Lo, Eds., pp. 463-480, Springer (2009).

17. Chung, K., Jang, K., Madanat, S. and Washington S. "Proactive detection of high collision concentration locations on highways", *Transp. Res., Part A: Policy and Pract.*, **45**(9), pp. 927-934 (2011).

18. Miaou, S.P. and Lum, H. "Modeling vehicle accidents and highway geometric design relationships", *Accid. Anal. and Prev.*, **25**(6), pp. 689-709 (1993).

19. Shankar, V., Mannering, F. and Barfield, W. "Effect of roadway geometrics and environmental factors on rural freeway accident frequencies", *Accid. Anal. and Prev.*, **27**(3), pp. 371-389 (1995).

20. Tamburri, T.N. and Smith, R.N. "The safety index: Method of evaluating and rating safety benefits", *Highw. Re. Rec.*, **332**, pp. 28-39 (1970).

21. Washington, S., Haque, M.M., OH, J. and Lee, D. "Applying quantile regression for modeling equivalent property damage only crashes to identify accident blackspots", *Accid. Anal. and Prev.*, **66**, pp. 136-146 (2014).

22. Zhao, F. and Park, N. "Using geographically weighted regression models to estimate annual average daily traffic", *Transp. Res. Rec.: J. of the Transp. Res. Board*, **1879**, pp. 99-107 (2004).

23. Charlton, M. and Fotheringham, A.S. "Geographically weighted regression: A white paper". Available at: URL:http://ncg.nuim.ie/ncg/GWR/GWR_WhitePaper.pdf. (2009).

24. Jorgensen, N.O. "Statistical detection of accident blackspots", *Presented at OTAPIARC 11th Int. Study Week in Transp. and Saf.*, Brussels, Belgium (1972).

25. McGuigan, D.R.D. "The use of relationships between road accidents and traffic flow in 'black-spot' identification", *Traffic Eng. and Control*, **22**, pp. 448-453 (1981).

26. California Department of Transportation. "Table C task force: Summary report of task force's findings and recommendations", Available at URL: http://www.dot.ca.gov/newtech/researchreports/2002-2006/2002/table_c_overview- 2003_final.pdf., (2002).

27. Sung, N., *A Study on Evaluation of Safety at Highways Using Traffic Crash Prediction Models*, Korea Transport Institute, Korea (2003).

28. Kononov, J. and Allery, B. "Level of service of safety", *Transp. Res. Rec.: J. of the Transp. Res. Board*, **1840**, pp. 57-66 (2003).

29. Park, S.H., Bigham, J.M., Kho, S.Y., Kang, S. and Kim, D.K. "Geocoding vehicle collisions on Korean expressways based on postmile referencing", *KSCE J. of Civ. Eng.*, **15**(8), pp. 1435-1441 (2011).

30. Jang, K., Shim, J., Chung, S. and Park, S.H. "Effectiveness of managed lanes on south Korean expressways", *93rd Annual Meeting of the Transp. Res. Board*, Washington, D.C. (2014).

31. Kwak, H., Kim, D.K., Park, S.H. and Kho, S.Y. "Development of a safety performance function for Korean expressways", *Selected Proceedings of the 12th World Conf. on Transp. Res. Soc.*, Lisbon, Portugal (2010).

32. Park, S., Park, S.H., Kim, D.K. and Chon, K.S. "Factors that influence the level of accident severity in vehicle crashes: A case study of accidents on Korean expressways", *Selected Proceedings of the 12th World Conf. on Transp. Res. Soc.*, Lisbon, Portugal (2010).

33. Moran, P.A.P. "Notes on continuous stochastic phenomena", *Biom.*, **37**, pp. 17-33 (1950).

34. Kang, S., Spiller, M, Jang, K., Bigham, J. and Seom, J. "Spatiotemporal analysis of macroscopic patterns of urbanization and traffic safety: case study in Sacramento County, California", *Transp. Res. Rec.: J. of the Transp. Res. Board*, **2318**, pp.45-51 (2012).

35. Koenker, R. and Bassett, Jr., G. "Robust tests for heteroscedasticity based on regression quantiles", *Econom.*, **50**(1), pp. 43-61 (1982).

36. Akaike, H.A. "New look at the statistical model identification", *IEEE Trans. on Autom. Control*, **19**(6), pp. 716-723 (1974).

37. Pulugurtha, S.S., Krishnakumar, V.K. and Nambisan, S.S. "New methods to identify and rank high pedestrian crash zones: An illustration", *Accid. Anal. and Prev.*, **39**, pp. 800-811 (2007).

38. Silverman, B.W., *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, New York, USA (1986).

Biographies

Shin Hyoung Park received his PhD degree from Seoul National University, Seoul, South Korea in 2010, and is now an Assistant Professor of the Department of Transportation Engineering at Keimyung University in South Korea. His research interests include GIS-based spatial analysis for high collision concentration locations, traffic accident data analysis, road safety audit, and traffic management plan.

Kitae Jang received the BS degree in Civil Engineering from Hanyang University, Seoul, Republic of Korea, in 2003 and the MS and PhD degrees in Civil and Environmental Engineering from the University of California, Berkeley, U.S., in 2006 and 2011, respectively. Since 2012, he has been an Assistant Professor with the Cho Chun Shik Graduate School for Green Transportation, Korea Advanced Institute of Science and Technology, Daejeon. His research interests include traffic flow theory and transportation operations, traffic safety and transportation economics.

Dong-Kyu Kim received his PhD degree from Seoul National University in South Korea in 2006. He is currently an Assistant Professor of the Department of Civil and Environmental Engineering at Seoul National University in South Korea. His research interests include transportation planning, logistics, network modelling, and transportation safety.

Seung-Young Kho received his PhD degree from University of California at Berkeley in 1989. He has been a Professor of the Department of Civil and Environmental Engineering at Seoul National University in South Korea. His research interests include transportation safety, traffic operation, Intelligent Transportation System, logistics, and public transportation.

Seungmo Kang received the BS and MS degrees in Civil Engineering from Seoul National University, South Korea, in 1994 and 1998, respectively, and PhD degree in Civil Engineering from University of Illinois at Urbana-Champaign in 2008. He is an associate professor in School of Civil, Environmental and Architectural Engineering at Korea University. His research interest lies in developing mathematical models to address challenging problems that arise in the fields of transportation planning, safety and logistics.