EXPLORATIVE SPATIAL ANALYSIS OF TRAFFIC ACCIDENTS USING GWPR MODEL FOR URBAN SAFETY PLANNING

Ali Akbar Matkan
Associate Professor of RS&GIS center, Faculty of Earth Sciences,
Shahid Beheshti University,Tehran, Iran, E-mail: a-matkan@sbu.ac.ir

Afshin Shariat Mohaymany
Assistant Professor of Civil Engineering, Faculty of Civil Engineering,
Iran University of Science & Technology, Tehran, Iran, E-mail: shariat@iust.ac.ir

Babak Mirbagheri
Lecturer of RS&GIS center, Faculty of Earth Sciences,
Shahid Beheshti University, Tehran, Iran, E-mail: b_mirbagheri@sbu.ac.ir

Matin Shahri
MSc Student of RS&GIS center, Faculty of Earth Sciences,
Shahid Beheshti University, Tehran, Iran, E-mail: matinshahri@yahoo.com

Submitted to the 3rd International Conference on Road Safety and Simulation,
September 14-16, 2011, Indianapolis, USA

ABSTRACT

In recent years considerable studies have been carried out to investigate the relationships between crashes and related factors. Trip generator variables as the function of socio-economic and demographic characteristics and land use along with exposure variable, Vehicle Kilometers Travelled (VKT) might be appropriate variables to estimate the crashes at aggregated area-level. Generalized Linear Models (GLMs) as the most widely used global models employed in crash analyses, describe the relationships between the dependent and independent variables by estimating the fixed global coefficients. Since the crash occurrences are likely influenced by many spatial factors which vary over space, Geographically Weighted Poisson Regression (GWPR) is introduced to investigate the non-stationary effects in estimating the number of crashes per Traffic Analysis Zone (TAZ). In this paper, the GWPR was primarily employed on 253 TAZs in Mashhad, Iran, then Moran's I indicator was performed on model residuals to examine the spatial autocorrelation. GWPR shows an improvement in model performance as indicated by goodness-of-fit measures. The results also indicate the non-stationary state in the relationships between the number of crashes and independent variables. The positive sign of coefficients for most of the TAZs shows that although the selected variables have a positive influence on the number of accidents, their magnitude varies considerably over the space. Moreover, testing the GWPR residuals conveys that the model has successfully removed the spatial autocorrelation among residuals.

Keywords: Urban safety, Trip generation, Generalized Linear Models, non-stationary, Geographically Weighted Poisson Regression.

INTRODUCTION

The rapid expansion of road construction and ever-increasing growth of urbanization have led to increased number of vehicles. Although it has substantially contributed to socio-economic welfare benefits, broader communications and freight and passengers' high-speed transition, the traffic increase in cities has enormously augmented the number and severity of accidents. Hence, achieving a trip purpose without incurring personal harm or property damage, through being fully aware of the crash-affecting factors and obtaining reasonably accurate safety planning models with high predictive power have been the perpetual concerns of the traffic safety specialists in long-term transportation planning process. What magnifies the importance of this matter yet more is the conduct it provides for the safety officials to determine the appropriate strategies for immanent safety problems (Washington, Schalwyk, & Mitra, 2006).

There are many studies on the use of macro-level models for predicting crashes in urban areas, the variables of which have been aggregated at different levels e.g. TAZ (Hadayeghi, Shalaby, & Persaud, 2010; Lovegrove, 2008; Naderan & Shahi, 2010; Washington, et al., 2006) , census blocks (Levine, Kim, & Nitz, 1995; Wier, Weintraub, Humphrey, Seto, & Bhatia, 2009) or wards (Quddus, 2008).Also, various models have been developed to explain the relationships between the crash frequency and the number of explanatory variables such as network parameters (Noland, 2002),the traffic volume (Caliendo, Guida, & Parisi, 2007; Hadayeghi, et al., 2010; Noland & Quddus, 2004), land use (Hadayeghi, et al., 2010; Noland & Quddus, 2004) and demographic characteristics(de Guevara, Washington, & Oh, 2004; Noland & Lyoong, 2004).

The selected explanatory variables should be a true predictor of dependent variable rather than an extraneous one. In general, to any model, the more independent variable added to the equation, the greater the overall fit will be, yet this process must be followed with great caution; since adding too many variables to an equation would account for strange effects of overfitting (Levine, Lord, & Park, 2010). In order to handle this problem, one might apply diagnostic tests such as tolerance statistics or employ the surrogate variables to adequately model the dependent variable.

Travel demand estimation plays a key role as the fundamental step in long-term transportation planning process. Trip generation is the first step in travel forecasting in which the regression models are employed to estimate the number of trips made from or to each TAZ based on various variables such as socio-economic, demographic and land use characteristics. These models are calibrated using the concept of origin-destination and based on trip purposes over TAZs. Each trip generated from a TAZ can be subject to a risk of crash. In other words, travel demand in urban areas is a contributing factor in crash occurrence at TAZ level and the probable cause of a crash, could be the trip making itself (Naderan & Shahi, 2010). It is likely that the trip production and attraction, the first-hand results of trip generation models constructed based on socio-economic, demographic and land use characteristics, might be powerful enough to characterize the crash occurrences. Vehicle Kilometers Travelled (VKT) is the total kilometers travelled by vehicles on any particular road system during a given period of time and functions as an indicator commonly used to measure road safety.

The prevalent techniques employed in calibrating the safety planning models are the Generalized Linear Models (GLMs) with pre-assumed Negative Binomial (NB) or Poisson distribution for errors (Hadayeghi, et al., 2010). In such global techniques the fixed coefficients are estimated to describe the relationships between the dependent variable (commonly the crash frequency) and every one of the explanatory variables. Needless to say, actual spatial patterns will vary with local site conditions, a phenomenon referred to as spatial non-stationary process (Brundson, Fotheringham, & Charlton, 1998). The stationary assumption in GLMs hides some underlying spatial aspects affecting the crashes, thus the accuracy of such models to describe the relationships between the crashes and independent variables might be controversial. It is likely that while some explanatory factors have strong predictive power on the estimation of the number of crashes in one location, they might not be as powerful enough elsewhere. In addition, the GLMs assume that the model error terms are independent of each other, so if there is any spatial autocorrelation among the error terms, implementing GLMs might be questionable (Fotheringham, Charlton, & Brundson, 2002). To overcome these limitations, Geographically Weighted Regression (GWR), a local regression technique, was proposed by Fotheringham. This technique brings about the calibration of multiple regression models for the non-stationary process (Fotheringham, et al., 2002).A developed form adapted to model the count data is known as Geographically Weighted Poisson Regression (GWPR). The most important advantage of GWR or GWPR is that the model can estimate regression coefficients at any spatial location and produce a better predictive performance for the response variable. In addition, the residuals of such local models have more desirable spatial randomness than those derived from global models (Zhang, Gove, & Heath, 2005).

Over the past years, GWR technique has mainly been applied to health, economic and urban studies (Brundson, et al., 1998; Gao & Li, 2010; Nakaya, Fotheringham, Brundson, & Charlton, 2005). Among the studies which developed this technique in the transportation engineering application (Clark, 2007; Du & Mulley, 2006; N. Park, 2004; Zhao, Chow, Li, & Liu, 2005), very few cases have directly led towards safety particularly crash analysis. Much of the statistical analysis of traffic crashes has evaluated the influence of conventional explanatory variables on crash occurrences based on the global regression models.

In a recent study, Hadayeghi et al. employed GWR to investigate the relationship between crashes and the explanatory variables such as socio-economic and demographic characteristics, network characteristics and traffic volume in 463 TAZs in Toronto. The results indicated that GWR function improved the overall prediction by 17% increase in measuring the correlation coefficient; however, the crashes assumed to be normally distributed which deemed to be inaccurate assumption concerning crash counts (Hadayeghi, Shalaby, & Persaud, 2003). Another research carried out by Cloutier et al. to analyze the child pedestrian accidents in Montréal, Canada revealed the non-stationary effects of relationships. In spite of the fact that some variables are not significant in the global model, they show local significances (Cloutier, Apparicio, & Thouez, 2007). The inter-province differences in traffic accidents and mortality rate to evaluate the road safety in Turkey was also shown in a research by Erdogan in 2009. This research seeks to compare the conventional Ordinary Least Square (OLS) and GWR methods to evaluate the relationship between the crash rate and several types of predictors. The results indicated the great improvement in AICc (Corrected Akaike Information Criterion) and correlation coefficient for GWR model (Erdogan, 2009). Having taken the intrinsic properties of

crash count data into account, Hadayeghi et al. studied the spatial variations in the relationship between the number of zonal collisions and potential transportation planning predictors for total and severe collisions in 481 TAZs in Toronto, Canada using GWPR. The results were compared with Global NB and Poisson regression models. The local Vehicle Kilometers Traveled (VKT) coefficients for total and severe collisions were mapped and the varying coefficient signs were analyzed. The output shows a great improvement in Pearson's product moment correlation coefficient and AICc (Hadayeghi, et al., 2010).

A review on the past studies reveals that the common methods applied to predict the crashes particularly in aggregated area-level employed the GLMs which result in estimating the global fixed coefficients by characterizing a stationary relationship between the dependent (commonly the crash frequencies) and the independent variables though very few studies focused on the local variation of relationships. The common predictors such as land use, socio-economic characteristics and network parameters were applied in calibrating the safety models. The only study based on the travel demand variables (Naderan & Shahi, 2010) has developed an aggregated crash prediction model using crash generation variables over 380 TAZs in Mashhad, Iran. The NB models were calibrated separately for crash production and attraction based on different activities which generated the trips, yet it led to estimating the global coefficients. Among those studies employing GWR in safety, the common potential predictors such as land use, network parameters, demographic characteristics and traffic volume were used in calibrating the safety model. Not only does developing these models require detailed data aggregating in spatial units, it might also lead to inserting too many variables causing the effects of overfitting. The use of trip production and attraction as surrogate variables is; therefore, justified. The main objective of this paper with a case study in Mashhad, Iran is developing the local safety model using the first-hand results of trip generation models namely trip production and attraction variables along with VKT as an exposure variable. Furthermore, this study attempts to examine the global and local models in terms of probable spatial autocorrelations among residuals which seem to have been neglected in the prior studies. The method applied in this paper is particularly useful for policy makers who can make decisions based on TAZs characteristics before generating crashes and identifying the hazardous zones which demand more attention.

METHODOLOGY

Data Collection/Process

This research is based on three types of databases; the first is the total number of crash data available for Mashhad's 253 TAZs in 2008, updated by Mashhad Transportation and Traffic Organization. The crash data was geocoded using the GIS tools and were grouped into TAZs. The second type of data includes the total number of trips produced or attracted in TAZs collected based on observations in order to update the transportation studies, construct and validate the travel demand models in Mashhad .Such models are evaluated based on different trip purposes. However; since we are planning to evaluate the crash risks, the total number of trips is employed, regardless of the explicit classification on trip purpose. The third type of data is VKT aggregated for TAZs for the same year. Figure 1 illustrates the spatial density distribution of trip production, attraction and VKT per TAZ in city of Mashhad.

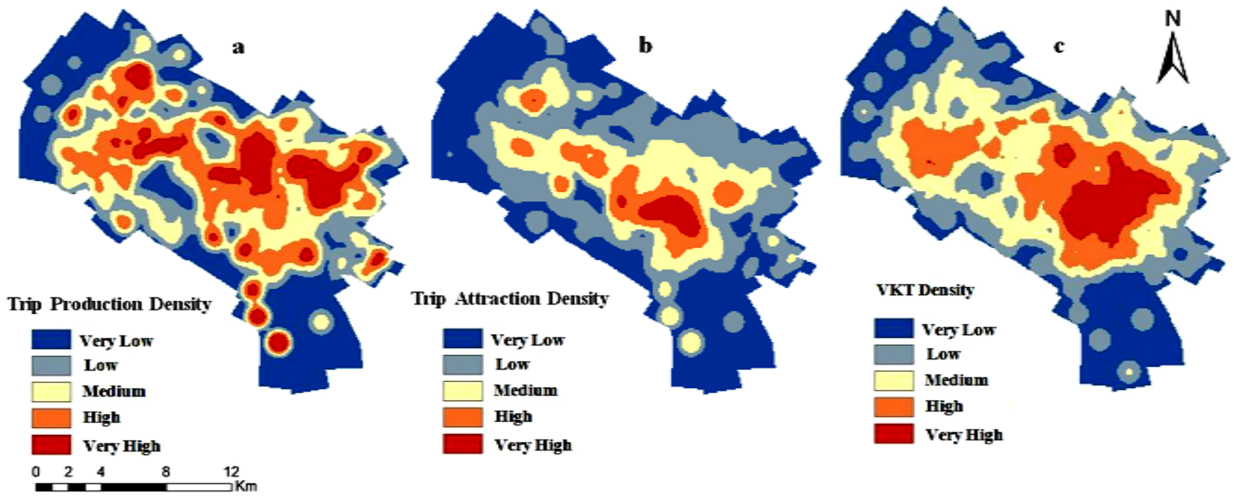

Figure 1 (a) Spatial density distribution of total Trip Production,
(b) Spatial density distribution of total Trip Attraction, (c) Spatial density distribution of VKT

Generalized Linear Models

The GLMs are known as the most common techniques used to predict crashes which assume that the error structure is either Poisson or NB (Hadayeghi, et al., 2003). In a Poisson regression model, the probability of ith TAZ having y_i number of crashes is given by (Cameron & Trivedi, 1998):

$$Prob(y_i) = \frac{e^{\lambda}\lambda^{y_i}}{y_i!} \tag{1}$$

where :
Prob(y_i) : the probability of y collisions occurring in ith TAZ
y_i : the number of crashes occurring in the ith TAZ
λ : the expected number of crashes

The conditional average of y_i can be specified as an exponential function of independent variables:

$$E(y_i|x_i) = \lambda_i = e^{{X_i}^T\beta} \tag{2}$$

where :
X : the set of independent variables
β : the set of estimated coefficients

The natural log of λ_i is assumed to be a linear function of k independent variables. That is:

$$Ln(\lambda_i) = \beta_0 + \sum_k \beta_k\, x_{ik} \tag{3}$$

The model coefficients in Equation 3 are estimated by the maximum likelihood method or ordinary least square (Cameron & Trivedi, 1998). The Poisson distribution has been shown to be reasonable for modeling collision data, but there is a limitation as the variance of the crash data is restrained to be equal to the mean. To handle this problem, NB model is suggested; although

Miaou indicated that both Poisson and NB models in safety studies produce approximately the same coefficients (Miaou, 1994).

Geographically Weighted Poisson Regression

Since the coefficients are assumed to be constant (stationary) across the study region, the model represented in Equation 3 is often referred to as global (or spatially stationary) model (Fotheringham, et al., 2002). When spatial non-stationary relationships prevail, the estimated coefficients will be the function of (u_i, v_i) implying the geographic coordinate of ith TAZ centroid, thus Equation 3 can be written as:

$$Ln(\lambda_i) = \beta_0(u_i, v_i) + \sum_k \beta_k (u_i, v_i) x_{ik} \tag{4}$$

In contrast to GLMs, this technique is carried out using localized points within geographic space to capture the spatial variations (Brundson, et al., 1998; Fotheringham, et al., 2002) .The idea behind GWR or GWPR states that the data points in close spatial proximity of regression point have more influence on estimating the coefficients rather than those in further distances. Such influence is determined by applying a weighting function. Consequently the estimated results depend not only on the observations received, but also on the choice of weighting function kernel and its bandwidth. Two types of weighting functions are known as followed:

Gaussian:

$$w_{ij} = \exp\left(-\frac{1}{2} \times \left(\frac{d_{ij}}{b_i}\right)^2\right) \tag{5}$$

Or Bi-Square:

$$w_{ij} = \begin{cases} [1 - (\frac{d_{ij}}{b_i})^2]^2 & if\ d_{ij} < b_i \\ 0 & otherwise \end{cases} \tag{6}$$

where:
w_{ij}: the spatial weight of jth TAZ on ith TAZ
d_{ij} : Euclidean distance between ith TAZ and jth TAZ
b_i: the kernel bandwidth.

The bandwidth controls the distance decay in weighting function and consequently the degree of locality. As the bandwidth increases, the local variations are missed and the estimated values approach the global estimations. The bandwidth might be fixed, which assumes that the bandwidth at each regression point is constant throughout the study area (Gaussian function) (Figure 2a); alternatively, an adaptive spatial kernel may be used, (Figure 2b) which adapts for the density of data at each regression location (TAZ). In an adaptive kernel, the optimal number of neighboring TAZs is selected to seek a specific number of nearest TAZs to ensure including constant number of local samples. The weights are then computed by using the specified kernel and setting the value for any TAZ based on Equation 6 (Bi-square function). For the TAZs whose distance is greater than the bandwidth the weight is set to zero and the TAZ will be excluded from the local calibration. In practice, the results obtain through GWR are not sensitive to the choice of weighting function type, but very sensitive to bandwidth. If the bandwidth is

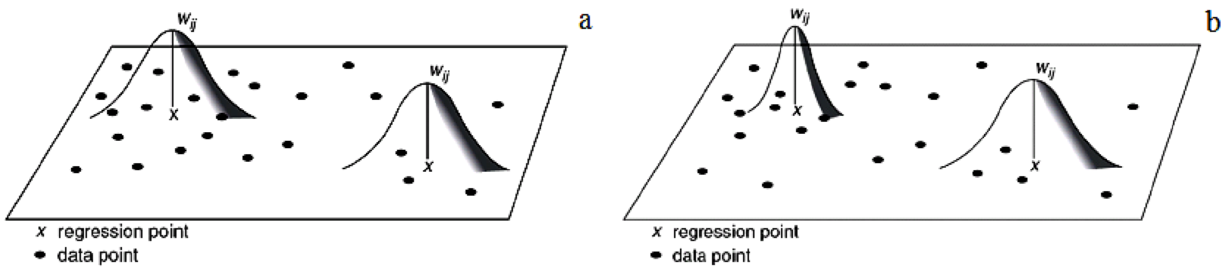

Figure 2 (a) GWR with fixed spatial kernel, (b) GWR with adaptive spatial kernel

known a priori (fixed), it is possible to apply it directly to the model, however if it is unknown (adaptive), AICc is used to optimize the bandwidth (Nakaya, et al., 2005):

$$AIC_c(b) = D(b) + 2K(b) + 2\frac{K(b)(K(b)+1)}{n-K(b)-1} \tag{7}$$

where
b : bandwidth
D : the deviance of parameters
K : the effective number of parameters in the model
n : the number of total TAZs

AICc estimation has the advantage of being generally applicable, meanwhile it can also be used to compare whether the results from GWR present a better fit than the global model by taking models' degrees of freedom into consideration as well (Fotheringham, et al., 2002). In this study, the adaptive kernel for development of model and the AICc selection process as highlighted above are employed to optimize the bandwidth.

Assessing the Goodness-of-fit

Measures of goodness-of-fit typically summarize the discrepancy between observed and predicted values under the model.

<u>Pearson's Correlation Coefficient</u>

Pearson's correlation of coefficient is a measure of the correlation between observed and the predicted values, giving a value between +1 and −1. Values near +1 and −1 imply the strong positive and negative relationships respectively. As the value approaches zero, it indicates that there is no correlation between variables.

<u>Mean Squared Error (MSE)</u>

MSE measures the average of the squares of the difference between observed and predicted number of crashes divided by the sample size (number of TAZs) and is defined as:

$$MSE = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n} \qquad (8)$$

Where:
y_i : the observed number of crashes per TAZ
$\hat{y}_i$: the predicted number of crashes per TAZ
n: the sample size (the number of TAZs)

The less the MSE, the better the model performance.

Testing the Spatial Autocorrelation of Residuals

Spatial autocorrelation is a term referring to the obvious fact that data from locations near one another in space is more likely to be similar than data from locations remote from one another (Anselin, 1995). It is assumed that estimated error for any observation cannot be related to the error for any other observation (Fotheringham, et al., 2002; Levine, et al., 2010). It is possible that much of the observed spatial autocorrelation in residuals that is frequently observed in the calibration of global models results from applying a global model to a non-stationary process, therefore when the non-stationary state prevails in relationships, GLMs might be misleading (Fotheringham, et al., 2002). The residuals resulted from global models must be examined to justify using global models. Moran's I is known as a common statistical index used to detect spatial autocorrelation and its values range from -1 to 1:

$$I = \frac{n\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(y_i - \bar{y})(y_j - \bar{y})}{\left(\sum_{i \neq j}^{n}\sum_{j=1}^{n} w_{ij}\right)\left(\sum_{i=1}(y_i - \bar{y})^2\right)} \qquad (9)$$

Where :
n : number of samples (the number of TAZs)
$\bar{y}$: the global mean value for number of crashes, calculated as a simple average value based on all data
$y_{i,}$: the number of crashes at ith TAZ
y_j : the number of crashes at jth TAZ
w_{ij} : a spatial weight matrix defined to determine the degree of locality

The larger the absolute value of Moran's I, the more significant the spatial autocorrelation and a value of zero means perfect spatial randomness (Anselin, 1995). In this research, we implement Moran's I to examine the local and global residuals in order to justify employing the local model.

RESULTS & DISCUSSIONS

Comparing the Goodness-of-fit Between GLMs and GWPR

Table 1 presents the goodness-of-fit measures for both global and local models. The average correlation coefficient obtained from GWPR compared with corresponding value obtained from global model shows that the local model has improved the crash prediction around 17%. Moreover, the MSE and AICc values for 2008 zonal crash prediction using GWPR model is

Table 1 Comparing the goodness-of-fit for Global Poisson model and GWPR

Model	Correlation Coefficient	MSE	AICc
Global Poisson	0.63	11763.25	12525.03
GWPR	0.80	7350.31	7869.20

lower than the equivalent values for global model indicating that the local model has been much more successful capturing the crash variations.

Results of Modeling

To highlight the performance of GWPR, the parameters were estimated using the "GWRx3.0" software for the total crashes per year as dependent variable and total trip production, attraction and VKT as explanatory variables. A summary of 5-number distribution for locally estimated coefficients and the corresponding results drawn from global model are presented in Table 2 and Table 3 respectively. As can be seen, opposed to fixed coefficients obtained by the global model, the estimated coefficients by the local model show a rather spatial non-stationary state. The degree of spatial non-stationary state in a relationship is detected by comparing the range of the local parameter estimates with a confidence interval around the global estimate of the equivalent parameter. Then the range of values of the local estimates between the lower and upper quartiles is compared with the range of values at ± 1 standard deviations of the global estimate. If the range of local estimates between the inter-quartile range is greater than that of ± 1 standard deviations of the global mean, this suggests the relationship might be non-stationary (Fotheringham, et al., 2002). Therefore based on Table 2 and Table 3, for the trip production variable, value of the local coefficients between the lower and upper quartiles is 0.000014 which is by large greater than the range of values at ± 1 standard deviations of global Poisson model corresponding to zero. The same analysis for trip attraction variable indicates that the local coefficients range equals to 0.000012 which is greater than the value at ± 1 standard deviations of global Poisson model equals as zero. The local coefficient range for VKT variable is 0.218881 which is significantly greater than the equivalent value at ± 1 standard deviations of corresponding global Poisson model equal to 0.011052 which reflects how the process is spatially non-stationary.

Table 2 Summary of coefficients results from GWPR model analysis

Variable	Minimum of coefficients	Lower Quartile of coefficients	Median of coefficients	Upper Quartile of coefficients	Maximum of coefficients
Trip Production	-0.000005	0.000006	0.000014	0.000020	0.000030
Trip Attraction	-0.000022	0.000005	0.000010	0.000017	0.000061
VKT	0.049078	0.322538	0.4457850	0.541419	0.704255

Table 3 Summary of results from Global Poisson model

Model	Variable	Coefficients	StdError	t-Statistics	P-Value
Global Poisson	Intercept	-0.195643	0.063186	-3.096301	0.010
	Trip Attraction	0.000013	0.000000	36.966679	0.001
	Trip Production	0.000011	0.000000	31.896622	0.001
	VKT	0.418720	0.005526	75.779064	0.001

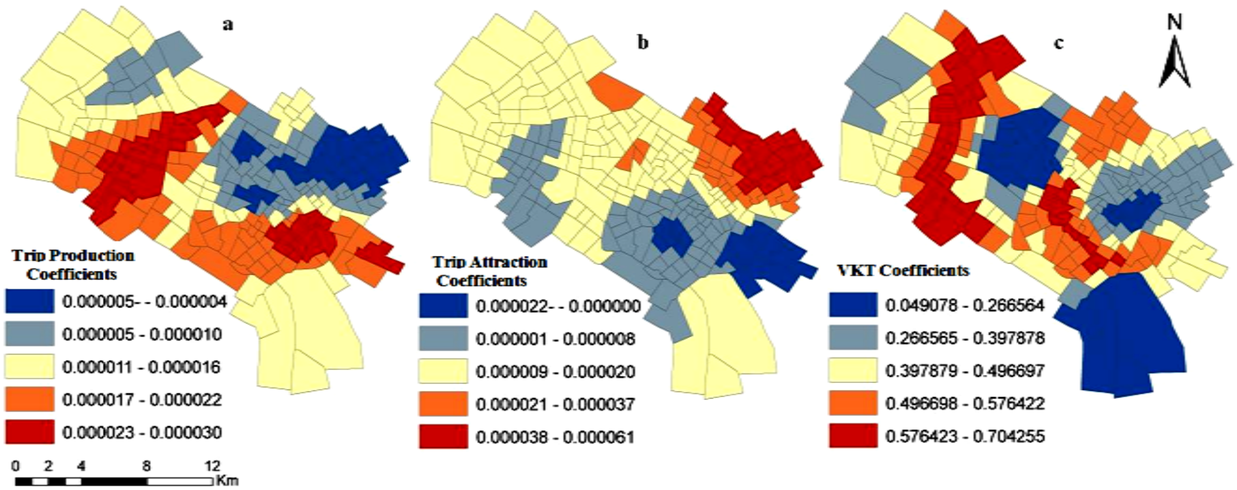

Figure 3 (a) Spatial variation of GWPR model (trip production estimates), (b) Spatial variation of GWPR model (trip attraction estimates), (c) Spatial variation of GWPR model (VKT estimates)

Figure 3 indicates that although the local parameter estimates are strongly positive for most of the TAZs, the intensity of relationships is not constant (Figure 3). It is evident that stronger positive relationships between crashes and trip production are found in the western parts of the study area near the core and some eastern TAZs of the city ; while the weaker but positive relationships are found in the northeast parts of the city (Figure 3a).Similarly, based on the maps produced for trip attraction coefficients, there are strong positive trends in trip attraction in the northeast parts of the city and the weaker positive relationships are detected in a vast area located in the southeast of the study area (Figure 3b). A brief review of past studies sheds light on the fact that the only study based on trip production and attraction variables was conducted over 380 TAZs in Mashhad in which the positive effects of the selected predictors on total and severe crash frequencies were reflected by developing NB models; however, as can be seen, according to Figure 3a and Figure 3b, the magnitude of the estimated local coefficients are not the same for all TAZs which could be an indicator for the non-stationary process in relationships. The results convey that the estimated local coefficients for VKT are positive over the study area which echoes its effect on crash occurrences. In an area located in northwest to southwest and some TAZs in southeast near the core of the city the stronger effects are highlighted (Figure 3c).In line with the findings of (Aguero-Valverde & Jovanis, 2006; D. Clark & Cushing, 2004; Quddus, 2008) VKT has been found to be an appropriate variable for predicting the crashes. However, the estimated coefficient for VKT variable has shown to be an average fixed value in previous studies while as depicted in Figure 3c the local estimated coefficients vary from 0.4 to 0.7.The similar findings reflecting the varying effects of VKT over space by employing GWPR model on TAZs in Toronto have been shown by Hadayeghi, et al., 2010 as well.

Evidently, the signs of local coefficients are not necessarily the same for all TAZs. Although the travel demand is expected to have a positive effect on the number of crashes, the TAZs with negative signs are also detected. One of the reasons for such counterintuitive signs as stated by Zhao et al. (2005) could be due to the multicollinearity among some variables or the collinearity

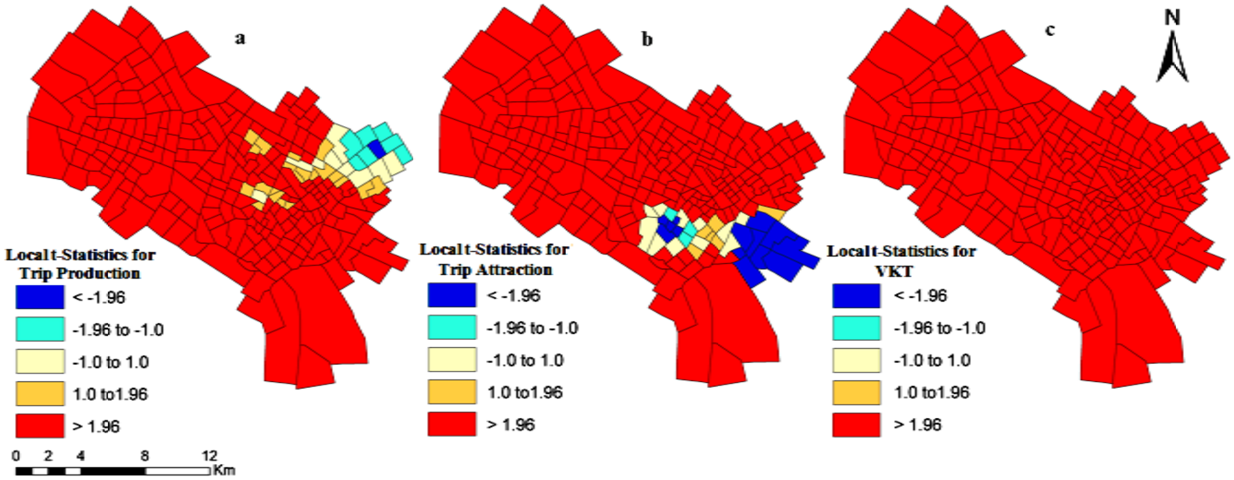

Figure 4 (a) t-Statistics of zonal trip production, (b) t-Statistics of zonal trip attraction, (c) t-Statistics of zonal VKT

in estimated coefficients. It is argued that the coefficients could be correlated even when there is no collinearity among variables (Griffith, 2008). Based on the data pre-analysis no serious multicollinearity is detected among explanatory variables, hence the existence of TAZs with negative signs could be due to collinearity in estimated coefficients. Although there is not a good diagnostic tool available in "GWRx3.0" to examine local multicollinearity, by employing the local t-statistics, one can determine where the relationships are significant and where they are not (Zhao, et al., 2005). Figure 4 depicts the results of t-statistics for explanatory variables. It demonstrates that for trip production and attraction, the t-values in most of the TAZs with negative signs are insignificant at 95% confidence level while the global model shows that these variables are significant in 90% confidence intervals. The same analysis reflects that although both local and global models are significant for VKT, the varying strength of its effect on crash occurrence can just be illustrated by local model which is depicted in Figure 3c. In spite of the controversies against producing unexpected signs for the coefficients, the researchers concur that GWR is reliable enough as an exploratory technique to understand how models may function differently across regions (Hadayeghi, 2009; Hadayeghi, et al., 2010; Ogneva-Himmelberger, Pearsall, & Rakshit, 2009).

Local R-Square obtained from GWPR along with estimated coefficients provides another way to detect the spatial varying relationships. Local variations of R-Square statistics can give an insight to how well a local model fits the observed data and the spatial heterogeneity in relationships between crashes and, trip generators and VKT. Unlike the global model, the spatial patterns of local R-Square in GWPR model represent a marked regional differentiation characterized by higher values (0.77) particularly found in southeast, center and some areas located in the west and some northern parts of the city indicating the ability of the model to locally fit the data (Figure 5). We might not expect the model calibrated in one location to replicate the data at other locations particularly well unless the processes being modeled are relatively stable. The values ranging from zero to 0.77 obtained from local model, give an intuition on the potential non-stationary process of variables.

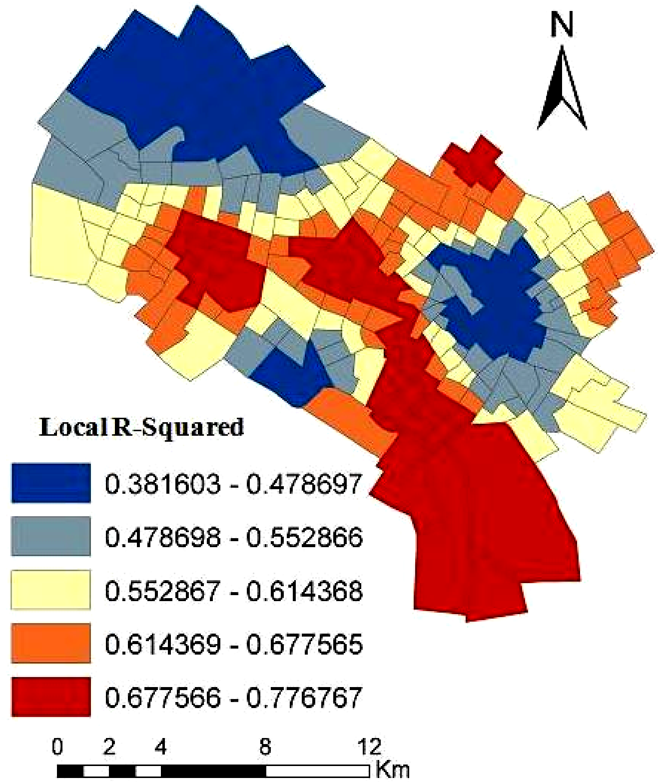

Figure 5 Results of local R-Squared for GWPR model

Comparison of Spatial Autocorrelation of Residuals between GLMs and GWPR

In order to compare the ability in addressing the spatial autocorrelation between global and local models, Moran's I index of residuals was computed using ArcGIS 9.3 tools.
As presented in Table 4, significant positive spatial autocorrelations are found for global Poisson model, characterized by Moran's I, corresponding z-score and p-value < 0.01. It was also found that no significant spatial autocorrelation was detected for the GWPR model specified respectively by Moran's I and z-score equal to -0.027 and -1.433, and p-value > 0.01. It reflects that with the same independent variables, GWPR improves the reliability of the relationships by removing the spatial autocorrelations in residuals. It could be another indication for spatially varying the trip generation variables which cannot be conveyed by employing the global models.

Table 4 Comparison of Moran's I of residuals between Global Poisson model and GWPR

Model	Moran's Index	Z Score	P-Value
Poisson	0.05	3	0.000
GWPR	-0.027	-1.433	0.1518

CONCLUSION

GLMs are known as the most prevalent techniques in calibrating the safety models with Poisson or NB assumption for error distribution which result to obtaining average fixed coefficients over the entire study area. However, the crash frequency is influenced by many factors which vary in space and demonstrate different spatial patterns.

This paper reveals incontestable spatial non-stationary relationships between the number of zonal crashes and trip generation variables along with VKT, using GWPR which allows the estimated coefficients to vary spatially. This represents a clear enhancement of the understanding offered by a global analysis by revealing certain aspects of the inter-relationships which do not emerge with traditional global specifications.

The model was developed based on trip production and attraction variables, the first-hand results of trip generation models as the probable cause of crashes in TAZs and VKT as an exposure risk variable. The estimated coefficients show significant differences. The magnitude of the regression coefficients are not the same in GWPR model, indicating that the explanatory power is stronger than GLMs. Additionally, our results based on the trip production , attraction and VKT reveal that GWPR not only represents a significant improvement of model performance over global model indicated by lower AICc, higher correlation coefficient values and the reduction of the spatial autocorrelation of residuals, but also prevents inserting the redundant explanatory variables which might lead to model overfitting. It is worth noting that the spatial varying effects of the selected predictors on crashes were investigated which was the main objective of this paper. In other words, we meant to convey how the relationships might be stronger somewhere but operate weaker elsewhere. Although the analyses of the TAZs which are affected more by a particular predictor still need further research. It is worth noting that in addition to investigating how travel demand in urban areas contributes to crash frequencies at TAZ level, which is the main purpose of this paper, the model can also be calibrated separately for each planning category such as land use or socio-economic variables. Such analyses are beneficial to urban planners and other analysts who deal with issues related to zoning and development of neighborhoods. Similarly, an exclusive model including network characteristics can be developed to help traffic engineers be informed of the varying local effects of individual crash predictors. The advantage of local models is highlighting the areas demanding more safety attention and consequently request assigning the resources such as budget and time, once they are included in detailed engineering study sites. Even though producing unexpected signs for the coefficients derived from GWPR pertaining to some TAZs for the reasons stated, can be considered as the model limitation, results convey that the GWPR models perform much better than the conventional Generalized Linear Models. Therefore, it can be concluded that the local model estimation capability of GWPR improves the safety studies based on its analyzing the effects of individual independent variables which might lead to urban safety improvement.

ACKNOWLEDGMENT

Data for this paper was provided by Tarhe Haftom consulting engineers. The writers wish to extend their very special thanks and appreciation to Dr.Ali Naderan for his relentless cooperation

in providing the data. Further appreciation goes to Mr. Mehrdad Hosseinpour for his interminably genuine collaboration in revising and commenting on the paper.

REFERENCES

Anselin, L. (1995). Local indicators of spatial association – LISA. *Geographical Analysis, 27*(2), 93-115.

Brundson, C., Fotheringham, S., & Charlton, M. (1998). Geographically weighted regression-modelling spatial non-stationarity. *Royal Statistical Society, 47*, 431-443.

Caliendo, C., Guida, M., & Parisi, A. (2007). A crash-prediction model for multilane roads. *Accident Analysis & Prevention, 39*(4), 657-670.

Cameron, C., & Trivedi, K., Parvin. (1998). *Regression Analysis of Count Data (Econometric Society Monographs)*: Cambridge University Press.

Clark, S. D. (2007). Estimating Local Car Ownership Models. *Transport Geography, 15*(3), 184-197.

Cloutier, M. S., Apparicio, P., & Thouez, J. P. (2007). GIS-based spatial analysis of child pedestrain accidents near primary schools in Montreal,Canada. *Applied GIS, 3*(4), 1-18.

de Guevara, F. L., Washington, S. P., & Oh, J. (2004). Forecasting Crashes at the Planning Level: a simultaneous negative binomial crash model applied in Tucson, Arizona. *Transportation Research Record,1897*, 191-199.

Du, H., & Mulley, C. (2006). Relationship Between Transport Accessibility and Land Value: Local Model Approach with Geographically Weighted Regression. *Transportation Research Record,1977*, 197-20.

Erdogan, S. (2009). Explorative spatial analysis of traffic accident statistics and road mortality among the provinces of Turkey. *Safety Research, 40*(5), 341–351.

Fotheringham, S., Charlton, M., & Brundson, C. (2002). *Geographically Weighted Regression*: John Wiley.

Gao, J., & Li, S. (2010). Detecting spatially non-stationary and scale-dependent relationships between urban landscape fragmentation and related factors using Geographically Weighted Regression. *Applied Geography*, *31*(1), 1-11.

Griffith, D. (2008). Spatial-filtering-based contributions to a critique of geographically weighted regression (GWR). *Environment and Planning A, 40*(11), 2751-2769.

Hadayeghi, A. (2009). *Use of advanced techniques to estimate zonal level safety planning models and examine their temporal transferability.* Toronto.

Hadayeghi, A., Shalaby, A. S., & Persaud, B. N. (2003). Macro-level accident prediction models for evaluating safety of urban transportation systems. *Transportation Research Record, 1840*, 87–95.

Hadayeghi, A., Shalaby, A. S., & Persaud, B. N. (2010). Development of planning level transportation safety tools using Geographically Weighted Regression. *Accident Analysis & Prevention, 42*(2), 676–688.

Levine, N., Kim, K. E., & Nitz, L. H. (1995). Spatial Analysis of Honolulu motor vehicle crashes:Zonal generators. *Accident Analysis & Prevention, 27*(5)*,* 675-685.

Levine, N., Lord, D., & Park, B.-J. (2010). *CrimeStat Version 3.3 Update Notes:Part 2: Regression Modeling*.

Lovegrove, G. R. (2008). Using Macro-Level Collision Prediction Models to Evaluate the Road Safety Effects of Mobility Management Strategies. *Transportation Research Board*.

Mei, C.-L., Wang, N., & Zhang, W.-X. (2006). Testing the importance of the explanatory variables in a mixed geographically weighted regression model. *Environment and Planning A, 38*(3), 587-598.

Miaou, S. P. (1994). The Relationship Between Truck Collisions and Geometric Design of Road Sections:Poisson versus Negative Binomial Regression. *Accident Analysis & Prevention, 26*, 471-482.

Naderan, A., & Shahi, J. (2010). Aggregate crash prediction models: Introducing crash generation concept. *Accident Analysis & Prevention, 42*(1), 339-346.

Nakaya, T., Fotheringham, S., Brundson, C., & Charlton, M. (2005). Geographically weighted Poisson regression for disease association mapping. *STATISTICS IN MEDICINE, 24*(17), 2695–2717.

Noland, R. (2002). *Traffic Fatalities and Injuries: The Effect of Changes in Infrastructure and Other Trends*. *Accident Analysis & Prevention, 35*(4)*,* 599-611.

Noland, R., & Lyoong, O. (2004). The effect of infrastructure and demographic change on traffic-related fatalities and crashes: a case study of Illinois county-level data. *Accident Analysis & Prevention, 36*(4), 525-532.

Noland, R., & Quddus, M. A. (2004). A spatially disaggregate analysis of road casualities in England. *Accident Analysis & Prevention, 36*(6), 937-984.

Ogneva-Himmelberger, Y., Pearsall, H., & Rakshit, R. (2009). Concrete evidence & geographically weighted regression: A regional analysis of wealth and the land cover in Massachusetts. *Applied Geography, 29*(4), 478-487.

Park, N. (2004). *Estimation of average annual daily traffic (AADT) using geographically weighted regression (GWR) method and geographic information system.* Florida International University, Miami,Florida.

Park, S. H., Kim, D.-K., Kho, S.-Y., & Rhee, S. (2010). *Idenfying hazardous locations based on severity scores of highway crashes.* Paper presented at the 12th WCTR, Lisbon.

Quddus, M. A. (2008). Modelling area-wide count outcomes with spatial correlation and heterogeneity: An analysis of London crash data. *Accident Analysis & Prevention, 40*(4), 1486–1497.

Washington, S., Schalwyk, I. V., & Mitra, S. (2006). *Incorporating Safety into Long-Range Transportation Planning*: National Cooperative Highway Research Program.

Wier, M., Weintraub, J., Humphrey, E. H., Seto, E., & Bhatia, R. (2009). An area-level model of vehicle-pedestrian injury collisions with implications for land use and transportation planning. *Accident Analysis & Prevention, 41*(1), 137–145.

Zhang, L., Gove, J. H., & Heath, L. S. (2005). Spatial residual analysis of six modeling techniques. *Ecological Modeling, 186*(2), 154–177.

Zhao, F., Chow, L.-F., Li, M.-T., & Liu, X. (2005). Transit Ridership Model Based on Geographically Weighted Regression. *transportation Research Record, 1972*, 105-114.