Accident Analysis and Prevention 42 (2010) 676–688

Contents lists available at ScienceDirect

Accident Analysis and Prevention

journal homepage: www.elsevier.com/locate/aap

Development of planning level transportation safety tools using Geographically Weighted Poisson Regression

Alireza Hadayeghi[a,*], Amer S. Shalaby[b,1], Bhagwant N. Persaud[c,2]

[a] CIMA+, 3380 South Service Road, Burlington, Ontario, Canada, L7N 3J5
[b] University of Toronto, 35 St. George Street, Toronto, Ontario, Canada, M5S 1A4
[c] Ryerson University, 350 Victoria Street, Toronto, Ontario, Canada, M5B 2K3

ARTICLE INFO

Article history:
Received 11 August 2008
Received in revised form 11 June 2009
Accepted 25 October 2009

Keywords:
Safety planning models
Model calibration
Geographically Weighted Poisson Regression
Local regression models
Generalized Linear Models
Spatial relationships

ABSTRACT

A common technique used for the calibration of collision prediction models is the Generalized Linear Modeling (GLM) procedure with the assumption of Negative Binomial or Poisson error distribution. In this technique, fixed coefficients that represent the average relationship between the dependent variable and each explanatory variable are estimated. However, the stationary relationship assumed may hide some important spatial factors of the number of collisions at a particular traffic analysis zone. Consequently, the accuracy of such models for explaining the relationship between the dependent variable and the explanatory variables may be suspected since collision frequency is likely influenced by many spatially defined factors such as land use, demographic characteristics, and traffic volume patterns. The primary objective of this study is to investigate the spatial variations in the relationship between the number of zonal collisions and potential transportation planning predictors, using the Geographically Weighted Poisson Regression modeling technique. The secondary objective is to build on knowledge comparing the accuracy of Geographically Weighted Poisson Regression models to that of Generalized Linear Models. The results show that the Geographically Weighted Poisson Regression models are useful for capturing spatially dependent relationships and generally perform better than the conventional Generalized Linear Models.

1. Introduction

Over the last few decades, the development of collision prediction models has enabled traffic engineers and road safety researchers to identify important factors related to the occurrence of collisions at intersections, on arterial road sections, on highways or on transportation networks. For transportation network applications, these models are usually developed at a macro-level, specifically at the traffic analysis zone (TAZ) level. The objective of these models is to capture the relationship between collision frequency and a number of explanatory variables such as traffic volume, road network characteristics, socioeconomic and demographic features, etc.

Several researchers have recently developed macro-level collision prediction models at the planning level in order to provide empirical tools for planners and engineers to conduct proactive road safety planning (Hadayeghi et al., 2007; Washington et al., 2006; Lovegrove and Sayed, 2005; de Guevara et al., 2004; Hadayeghi et al., 2003). The primary objective of those studies was to provide a safety planning decision-support tool, which facilitates a proactive approach to assess road safety implications of alternative network or land use planning initiatives and scenarios that are considered for medium to long-range implementation. In this sense, this approach is similar to how road designers make use of tools such as the Interactive Highway Safety Design Model (IHSDM) to assess the safety implications of design decisions. For this reason and others, it is desirable to obtain reasonably accurate safety planning models that have good predictive power.

A common technique used in the previous research for calibration of safety planning models is the Generalized Linear Modeling (GLM) procedure with the assumption of Negative Binomial error distribution. In this technique, fixed coefficients are estimated to represent the average relationship between the dependent variable, typically number of collisions per traffic analysis zone, and each explanatory variable. The model parameters are estimated globally for the entire study area with the implicit assumption that the relationship between the dependent variable and each independent variable does not vary across the geographic area. As such, a collision model calibrated with a constant coefficient

* Corresponding author. Tel.: +1 905 466 6672; fax: +1 289 288 0285.
E-mail addresses: alireza.hadayeghi@cima.ca (A. Hadayeghi), amer@ecf.utoronto.ca (A.S. Shalaby), bpersaud@ryerson.ca (B.N. Persaud).
[1] Tel.: +1 416 978 5907; fax: +1 416 978 5054.
[2] Tel.: +1 416 979 5000x6464; fax: +1 416 979 5122.

0001-4575/$ – see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.aap.2009.10.016

for the entire study area can be viewed as a "global" model. However, this stationary relationship may hide some important spatial factors affecting the number of collisions in a particular traffic analysis zone (TAZ). Consequently, the accuracy of such models for explaining the relationship between the dependent variable and the explanatory variables for specific TAZs may be suspected because collision frequency is likely influenced by many spatially defined factors such as land use, demographic characteristics, and traffic volume patterns. Some of these factors may have strong predictive power for estimating the number of collisions at certain locations but may be weak predictors at other locations.

To address this problem, several studies conducted in the past have developed techniques to consider the issue of spatial autocorrelation in multivariate regression models. Specifically, Geographically Weighted Regression (GWR) and Geographically Weighted Poisson Regression (GWPR) for count data have been developed by Fotheringham et al. (2002) to allow calibrating multivariate regression models of locally non-stationary processes. Through the explicit treatment of spatial coordinates, GWPR provides a set of local spatial parameters where the weights are linked to the distance between the observation and the location where independent variables in the models are measured. These local parameters are estimated using a geographically weighted likelihood principle described by Nakaya (2001).

The primary objective of study on which this paper is based was to develop GWPR models to investigate the local spatial variations in the relationship between the number of zonal collisions and potential transportation planning predictors such as traffic volume, road network characteristics, socioeconomic and demographic features, land use, dwelling unit, and employment type. As noted above, the main advantage of the GWPR models is that the independent variable coefficient estimates can vary spatially. This is a very essential feature of a model as it adds importance to the spatial location of data such as collision occurrence. It allows for a locally varying parameter representation of a particular point in space. The second objective of the paper is to compare the accuracy of GWPR models to that of GLMs with Negative Binomial and Poisson error structures, using a series of Goodness of fit measures.

Discussion related to application of the developed safety planning models in planning process is outside of scope of this paper. However, detailed discussion of this issue has been provided in a previous paper (Hadayeghi et al., 2007).

2. Literature review

Recently, a number of studies have been conducted to provide a regression framework in which spatial dependency is taken into account. These approaches are generally described as spatial regression models. These models recognize the fact that spatial data such as collision counts are not generally independent. Therefore, statistical inference based on the results of conventional normal regression analysis on spatial data may produce biased results. Conditional Autoregression (CAR) models, Simultaneous Autoregression (SAR) models, Spatial lag models, Generalized Estimating Equation (GEE) models, and Full-Bayesian Spatial models are some of techniques that are used for development of spatial models (see, e.g., Aguero-Valverde and Jovanis (2006), Miaou et al. (2003), Levine et al. (1995), LeSage (2001), Hadayeghi (2009) and Wang and Abdel-Aty (2006)).

While such models are generally not thought of as local models (i.e., GWR models), they do recognize the local nature of spatial data by relaxing the assumption that the error terms for each observation are independent. The output from these models consists of a set of global parameter estimates similar to conventional normal regression. However, spatial relationships are incorporated into the modeling framework through the covariance of the error terms (Fotheringham et al., 2002). In this case, these models can be thought of as "semi-local" as opposed to fully local as is the case for GWR models.

Several researchers have applied the GWR technique to the development of statistical models. However, this research was mostly in the areas of epidemiology and health science, and only very few models have been developed in transportation engineering applications, with none of these efforts directed towards road safety, or, specifically, road safety planning. In earlier research, (Hadayeghi et al., 2003) we used the GWR technique in order to explore the relationship between the number of zonal collisions and various explanatory planning variables. However, the model presented in that study was investigated just to test the hypothesis of how much improvement can be achieved by using the GWR model over conventional normal regression analysis. That study was exploratory in nature since it assumed a normally distributed error structure for collision data, which is not necessarily a correct assumption due to the non-negative and discrete properties of collision counts.

Of most relevance to the current effort was the work of Nakaya et al. (2005) who proposed the GWPR technique for relating the number of deaths in traffic analysis zones to socioeconomic covariates such as the proportion of elderly people, rate of house-ownership, unemployment rate, and proportion of professional and technical workers in each TAZ. The results indicated that there were significant spatial variations in the relationships between the dependent variable and some of the independent variables and that, consequently, the application of traditional global models would yield misleading results.

Zhao et al. (2005) is one of the few studies that applied the GWR technique in transportation engineering applications. That study investigated the spatial variations in the relationships between transit use and potential ridership variables including demographics, socioeconomics, land use, accessibility and transit supply features. The GWR models developed were compared with a global model estimated with the ordinary least squares method. The results indicated that the GWR models have better predictive power and provide an improved understanding of the spatial variations.

Park and Zhao (2004) applied the GWR technique to the estimation of the annual average daily traffic (AADT). The result of this study indicated that the GWR models are able to better explain the variation in the data and to predict AADT with smaller errors than the ordinary linear regression models. Additionally, GWR enabled the modeling of the spatial non-stationary data.

Du and Mulley (2006) used the GWR techniques to examine the relationship between transport accessibility and land value. They concluded that transport accessibility may have a positive effect on land value in some areas but in others a negative or no effect. The use of GWR allowed capturing such spatially varying relationships while a global model with an average or constant coefficient for land value was deemed inappropriate.

In summary, previous research reviewed suggests that GWR models perform much better than the global regression models in terms of the accuracy of the models. However, all of the transportation-related studies reviewed used GWR techniques that assume a normally distributed error structure in the calibration of regression models. Such an assumption is not optimal for calibrating regression models of count data (such as collision data).

As will be explained later in this paper, the GLM approach using Poisson and Negative Binomial regression provides a more appropriate basis for the collision dataset analysis than conventional linear regression, especially in entities where observed numbers of collisions are low. The use of Poisson regression in conjunction with GWR represents an advance in the calibration of safety planning models compared to previous studies which used the GLM approach for the calibration of such models. Although it would be beneficial to examine Geographically Weighted Negative Binomial regression, available GWR software, most notably "GWRx3.0", does not support the calibration of GWR models with Negative Binomial error structure, and as such these models are not calibrated in this study. It is worth noting that the use of Poisson regression instead of Negative Binomial regression does not produce inaccurate estimates from safety models in general since the model coefficients are quite similar for the two error distributions (Miaou, 1994). Likewise, Cameron and Trivedi (1998) showed that that a Poisson model with an overdispersion correction yields results that are equally as accurate as a Negative Binomial model.

The main reason for using Negative Binomial regression is that it facilitates the estimation of an overdispersion parameter that is necessary for applying the models in empirical Bayes analysis (Hauer, 1997). Given all of these considerations, it was deemed reasonable to adopt the GWPR technique for the development of collision models using the "GWRx3.0" software package.

3. Data

The data for this study was based on the data available for City of Toronto's 481 traffic analysis zones in year 2001. Several types of data were used for the completion of this study, as described next.

3.1. Collision data

The collision data were obtained from two main sources. The first dataset, consisting of the 2001 collision data for the City of Toronto, was provided electronically by the Traffic Data Centre (TDC) of the City of Toronto. The City's collision data are geocoded, which allows grouping of collisions by traffic analysis zone using GIS tools. The second dataset, consisting of the collision data for all the provincial highways within the City of Toronto, was obtained from the Ontario Ministry of Transportation (MTO). Since the collision data from the MTO were not geocoded, a GIS software (MapInfo) was used in order to geocode these data. Subsequently, the City's collision data were merged with the Ministry's data to produce a master collision database. A major issue with the City's collision data was that the coordinate systems assigned to some of the collisions were not accurate. Therefore, those collision data points did not overlay properly on the street files. To address this problem, a macro was developed in the GIS to assign those collisions to the nearest street.

3.2. Demographic, employment type and dwelling unit data

The second dataset was obtained from the 2001 Transportation Tomorrow Survey (TTS), a 1-day travel survey conducted in the Greater Toronto and surrounding areas (GTA) on approximately 5% of its population. The survey is conducted to obtain detailed demographic and travel information for the GTA population.

3.3. Traffic volume data

The traffic volume data were obtained from the results of a traffic assignment of the morning peak-period vehicle trips to the road network using the EMME/2 software package to estimate vehicle kilometers traveled (VKT) in each TAZ. This approximation was necessary because cordon counts, the only available source of traffic volume data collected consistently across the city, are made at a limited number of locations and therefore cannot be used directly to provide traffic flow estimates for each TAZ in the city. It is worth mentioning that it is a common practice in building travel demand and planning models to use traffic assignment model outputs.

3.4. Street network and land use data

GIS street network and land use data were obtained from Can-Map Street files from the University of Toronto Map Library.

The explanatory variables include demographic characteristics, employment type, dwelling unit type, traffic volume, street network characteristics and land use data. The definition of variables was based on previous research for similar studies and an iterative process in which various definitions were tried and the ones that gave the best results were ultimately selected. By combining the above data sources, various lists of possible explanatory variables were generated. Table 1 presents lists of variables and their descriptive statistics.

4. Model specification

As noted earlier both GWPR models and GLMs were calibrated, the latter for comparison purposes. These are described separately.

4.1. Generalized Linear Models

As noted above, the parameters in these models are estimated globally and do not change with locations where the models are being applied.

Following a review of various model forms used in the literature and applying selected model forms to the data, the following model form was used:

$$\ln(A) = \ln(\beta_0) + \beta_1 \ln(\mathrm{VKT}) + \beta_2 X_2 + \beta_3 X_3 + \cdots + \beta_p X_p \tag{1}$$

where $\ln(A)$, natural log of collision frequency per TAZ; VKT, vehicle kilometers travelled; X_j, j^{th} explanatory variables ($j = 2, \ldots, p$); β_j, jth model parameters ($j = 0, \ldots, p$).

The model form in Eq. (1) is consistent with the non-linear relationship between collisions and traffic intensity (measured in this case by VKT) which has been suggested by other researchers. Separate models were calibrated for total and severe (i.e., fatal plus injury) collisions.

4.2. Geographically Weighted Poisson Regression models

4.2.1. Geographically varying parameters in poisson regression

The model uses terms similar to Eq. (1) and is of the form

$$\ln(A) = \ln(\beta_0(\boldsymbol{u}_i)) + \beta_1(\boldsymbol{u}_i)\ln(\mathrm{VKT}) + \beta_2(\boldsymbol{u}_i)X_2 + \cdots + \beta_p(\boldsymbol{u}_i)X_p \tag{2}$$

here, unlike Eq. (1), β_j ($j = 0, 1, \ldots, p$) is a function of location $\mathbf{u_i}$(=(u_{xi}, u_{yi}))denoting the coordinates of the i^{th} point (TAZ centroid in this study) in space and is a vector of two dimensional coordinates describing the location of i with x and y coordinates.

As noted earlier, GWR is a technique developed to calibrate a multiple regression model that allows different relationships to exist at different points in space. The basic idea of GWR in general and GWPR as used in this study is that the observed data near point i have more of an influence on the estimation of the β_j (u_i)'s than data located farther from i. This influence around i is described by the weighting function.

The GWPR method attempts to capture spatial variation by fitting a regression model at each subject point (TAZ in this study) in the data, weighting all neighboring observations by a function of distance from that subject. It is evident that the spatial kernel function and window size, i.e., the bandwidth, used in the model fitting process impact the GWPR coefficient estimates. Therefore, the selection of the spatial kernel function and consequently, bandwidth, is an important step in the implementation of GWPR (Guo et al., 2008).

Table 1
Descriptive analysis of data.

Category	Variable	Avg	Min	Max	S.D.
Collisions	Number of all TAZ collisions in year 2001	171.04	14.00	880.00	123.44
	Number of TAZ fatal and injury collisions in year 2001	39.03	2.00	192.00	27.97
	Area of zone (1000 m^2)	1,317.88	59.37	8,735.03	937.70
Land use	Commercial (1000 m^2)	45.03	0.00	1,726.21	109.02
	Government and institutional (1000 m^2)	97.10	0.00	2,829.58	221.89
	Residential (1000 m^2)	669.53	0.00	2,816.70	535.05
	Open area (1000 m^2)	132.17	0.00	6,510.81	435.76
	Parks and recreational (1000 m^2)	142.01	0.00	2,278.35	257.98
	Resource and industrial (1000 m^2)	220.62	0.00	2,445.53	449.97
	Water body (1000 m^2)	11.04	0.00	612.93	42.66
Street network	Number of rail stations	0.20	0.00	4.00	0.51
	Total rail kilometer	0.43	0.00	4.31	0.72
	Number of schools	2.20	0.00	23.00	1.99
	Total arterial road kilometers	2.36	0.00	10.00	1.47
	Total expressway kilometers	0.74	0.00	9.96	1.62
	Total collector kilometers	1.42	0.00	7.99	1.38
	Total laneway kilometers	0.53	0.00	9.47	1.16
	Total local road kilometers	7.19	0.00	25.44	4.99
	Total ramp kilometers	0.51	0.00	10.01	1.16
	Number of 4-legged signalized intersections	2.52	0.00	10.00	1.91
	Number of 3-legged signalized intersections	1.00	0.00	8.00	1.17
Traffic	Average 85% operational speed	37.13	26.65	53.35	5.02
	Average posted speed	48.80	40.00	67.78	6.53
	Vehicle kilometer traveled	7365.64	100.51	61321.0	9112.29
	Average volume over capacity	0.49	0.00	1.00	0.17
Demographic	Number of females	2541.54	0.00	11194.0	1922.15
	Number of males	2386.20	0.00	105210	1804.79
	Number of people not possessing a driver's license	1945.03	0.00	11097	1731.61
	Number of people possessing a driver's license	2986.19	0.00	12499.00	2125.87
	Number of people	4931.60	0.00	21764	3720.28
	Number of people younger than 17-year-old	1026.86	0.00	5429	911.87
	Number of people older than 65-year-old	685.88	0.00	3504	542.88
	Number of people not holding a transit pass	4535.12	0.00	19812	3383.22
	Number of people holding a transit pass	394.47	0.00	3422	412.67
Dwelling units	Number of single houses	2621.29	0.00	13273	2097.01
	Number of apartments	2106.29	0.00	19535	2739.20
	Number of townhouses	204.00	0.00	2765	356.83
Employment types	Employed in: general office	304.77	0.00	1539	246.32
	Employed in: management	1150.07	0.00	5867	888.34
	Employed in: sales/service	513.33	0.00	2934	426.12
	Employed in: manufacturing/construction/trades	509.81	0.00	3242	518.70
	Number of full time employees	2062.65	0.00	8971	1523.69
	Number of part time employees	420.88	0.00	2351	335.34
	Number of not employed	2445.98	0.00	11548	1961.92

Number of traffic analysis zones = 481.

The bandwidth is the number of observations around each subject point and controls the distance decay in the weighting function. For the special case of GWPR models, the Gaussian and bi-square functions are commonly used to produce the weighting scheme for each data point as follows:

Gaussian:

$$w_{ij} = \exp\left(-\frac{1}{2} \times \frac{||\boldsymbol{u}_i - \boldsymbol{u}_j||}{G}\right) \quad (3)$$

Bi-square:

$$w_{ij} = \begin{cases} [1 - (||\mathbf{u}_i - \mathbf{u}_j||/G_i^2)]^2 & \text{if } ||\mathbf{u}_i - \mathbf{u}_j|| \prec G_i \\ 0 & \text{otherwise} \end{cases}$$

The parameter G_i (called the bandwidth) manages the kernel size. The results of GWPR models are sensitive to the bandwidth of a given weighting function. The bandwidth might be constant (fixed kernel), as in the Gaussian function, or variable (adaptive kernel) as shown in the bi-square function.

The shape and magnitude of the weighting functions will be the same for every sample point over space when the fixed kernel weighting scheme is used. As a result, the fixed kernel weighting scheme can produce inaccurate results for the estimation of model parameters since bandwidths may be larger for regions with dense data points and smaller for others. On the other hand, an adaptive kernel allows the kernel to vary spatially, i.e., to be larger for regions with scarce sample data and smaller for regions with more substantial data (Zhao et al., 2005; Silverman, 1986; Levine, 2004). Computational details for calibration procedures of GWPR, specifically local scoring procedures and calculation of standard error of parameters in GWPR models, can be found in Nakaya et al. (2005).

4.2.2. Calibration of collision GWPR models

For calibration of collision GWPR models, both fixed and adaptive kernels were examined. To do that, the Akaike Information Criterion (AIC) was used as an indicator not only for selection between fixed and adaptive kernels but also for bandwidth selection in adaptive kernels. Nakaya et al. (2005) showed that the AIC of GWPR models with bandwidth G can be defined as

$$\mathrm{AIC}(G) = D(G) + 2K(G) \tag{4}$$

where D and K are the deviance and the effective number of parameters in the model, respectively. The best GWRP model is the one with the lowest AIC. However, Nakaya et al. (2005) argue that since the degrees of freedom for GWPR models are likely to be small, small sample bias adjustment in the AIC definition should be done. They introduced the Corrected Akaike information Criterion (AICc) to address this bias. AICc is defined as follows:

$$\mathrm{AICc}(G) = D(G) + 2K(G) + 2\frac{K(G)(K(G)+1)}{N-K(G)-1} \tag{5}$$

If the effective number of parameters, K, is small relative to the number of observations, N, then the difference between AIC and AICc is negligible.

For this study, both fixed and adaptive kernels are applied for the development of each model. The AICc selection process, as highlighted above, is used to determine the optimal TAZ size required for GWPR modeling. For all cases, the models with the adaptive kernels provide lower values of AICc, which is an indicator of a better fit model. However, for the selection of bandwidth for the adaptive kernels, the "GWRx3.0" software package does not provide an optimal model with the best selected bandwidth. The solution was to manually insert a series of bandwidths, examine results of each, and finally, select a model which performs best.

5. Measures of goodness of fit

The GWPR model performance was assessed relative to the developed GLM models from a previous study (Hadayeghi et al., 2007) using the same database. The following goodness of fit measures was used to conduct the performance assessment of both models. These measures are similar to those used by Oh et al. (2003) and the authors' previous study (Hadayeghi et al., 2006).

5.1. Pearson's product moment correlation coefficient

As indicated by Oh et al. (2003), Pearson's product moment correlation coefficient is a measure of the linear relationship between two variables. A coefficient of exactly 1 for a model indicates that the model predicts observed data perfectly. It is defined as:

$$r = \frac{\sum (Y_i - Y_{\mathrm{avg}})(\hat{Y}_i - \hat{Y}_{\mathrm{avg}})}{\left[\sum (Y_i - Y_{\mathrm{avg}})^2 \sum (\hat{Y}_i - \hat{Y}_{\mathrm{avg}})^2\right]^{0.5}} \tag{6}$$

$\hat{Y}_i$, predicted number of collisions at zone i, Y_i, observed number of collisions at zone i, $\hat{Y}_{\mathrm{avg}}$, mean predicted number of collisions per TAZ, Y_{avg}, mean observed number of collisions per TAZ

5.1.1. Mean squared prediction error and mean squared error

The mean squared prediction error (MSPE) is the sum of squared differences between observed and predicted accidents divided by the sample size. It is defined as

$$\mathrm{MSPE} = \frac{\sum_{i=1}^{n} (\hat{Y}_i - Y_i)^2}{n} \tag{7}$$

The mean squared error (MSE) is the sum of squared differences between observed and predicted accidents divided by the sample size minus the number of model parameters. It is defined as:

$$\mathrm{MSE} = \frac{\sum_{i=1}^{n} (\hat{Y}_i - Y_i)^2}{(n-p)} \tag{8}$$

where p is number of model parameters, n is data sample size.

A comparison of MSPE and MSE shows potential over or under fitting of the models to the estimation data. If the MSE of a model is higher than the MSPE of another, this indicates that the first model may have been over fitted to the estimation data and that some of the observed relationships may have not been correctly shown in the model.

6. Results and discussion

Similar to our previous study (Hadayeghi et al., 2007) in which GLMs were estimated, several collision prediction models were developed with the GWPR technique to explore the relationship between collision frequency and zonal characteristics such as traffic intensity, road network, land use and socioeconomic and demographic characteristics. The procedure used in this study for the selection of explanatory variables in safety planning GLMs is a forward procedure. In this procedure, a simple model with only an intercept term is used as a starting point and the explanatory variables added to the model one by one. Prior to incorporating variables into the basic model, a correlation matrix was set up to examine whether or not variables of interest were highly correlated with other variables. If two explanatory variables were substantially correlated, inserting them simultaneously into the same model was avoided. Also, the correlation matrix was used for selection of the most important variable in the model.

For each model, the exposure variable (i.e., VKT) was initially considered due to its dominant prediction influence on collision data. Then, additional candidate variables were analytically selected from the lists in Table 1.

Separate GWPR models were developed for each planning category (different categories of the planning variables are shown in Table 1) and for combinations of categories in order to ensure applicability according to the availability of planning data.

In this section, the results of GWPR collision prediction models and their performance are summarized and presented in the first section. In the second section, the goodness of fit measures resulting from GWPR models are compared to those of the models estimated in our earlier research using GLM models with Negative Binomial and Poisson error distributions.

6.1. Results of modeling

The dependent variable of each developed model is the number of zonal collisions per year. The model parameters were estimated based on the maximum likelihood method for GWPR using the "GWRx3.0" package.

6.1.1. Model with traffic intensity as independent variable (Model 1 for total collisions and Model 1S for severe collisions)

As identified above, GWPR provides the ability to examine spatial relationships that can be hidden in a global GLM model. Therefore, the results of the GWPR models are a set of local parameters for each independent variable. In this case, the results of the traffic intensity-based GWPR models for total and severe (fatal and injury) collisions are different parameter values for VKT for each TAZ. The local parameter estimates are described by 5-number summaries and presented for total and severe collisions in Tables 2 and 3, respectively. The 5-number summary of a distri-

Table 2
Macro-level GWPR collision prediction models based on traffic and road network variables, total collisions.

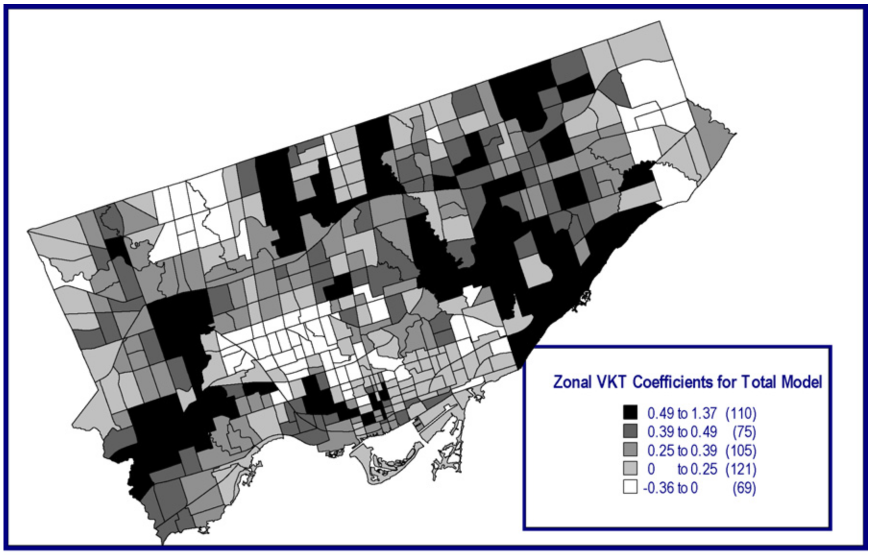
Parameters	GWPR model number							
	#1	#2	#3	#4	#5	#6	#7	#8
ln(*A*)	−6.4, 8.10 (1.09, 2.4,8.1)	1.06, 4.60 (3.31,3.84,4.14)	−3.23, 6.92 (1.2,2.3,3.3)	−6.42, 7.41 (1.32,2.53,3.74)	−4.58, 7.31 (1.55,2.67,3.74)	−3.81, 5.88 (1.50,2.56,3.47)	−3.97, 6.23 (1.41,2.42,3.47)	−2.839, 5.526 (1.49,2.50,3.40)
ln(VKT)	−0.36, 1.37 (0.12,0.33,0.47)	−0.009, 0.38 (0.03,0.07,0.13)	−0.33, 0.89 (0.14,0.27,0.38)	−0.32, 1.298 (0.06,0.23,0.38)	−0.26,1.17 (0.07,0.22,0.35)	−0.224, 0.979 (0.08,0.21,0.34)	−0.22, 0.99 (0.09,0.21,0.34)	−0.165, 0.857 (0.10,0.21,0.34)
Total arterial road kilometers		0.07, 0.33 (0.17,0.22,0.26)						
Total expressway kilometers		−0.15, 0.23 (0.05,0.08,0.12)						
Total collector kilometers		−0.02, 0.17 (0.05,0.09,0.12)						
Total laneway kilometers		−0.023, 0.48 (0.04,0.06,0.10)						
Total local road kilometers		−0.048, 0.016 (−0.02,−0.009,0.002)						
Total ramp kilometers		−0.29, 0.26 (0.01,0.07,0.13)						
Total road kilometers			−0.04, 0.16 (0.03,0.04,0.06)			−0.060, 0.175 (0.01,0.03,0.05)	−0.043, 0.156 (0.01,0.03,0.05)	−0.038, 0.118 (0.01,0.03,0.05)
Number of 4-legged signalized intersections				−0.165, 0.487 (0.08,0.14,0.21)		−0.083, 0.372 (0.05,0.10,0.17)	−0.098, 0.350 (0.05,0.10,0.16)	−0.069, 0.347 (0.05,0.09,0.15)
Number of 3-legged signalized intersections				−0.514, 0.564 (0.04,0.15,0.23)		−0.195, 0.543 (0.06,0.13,0.19)	−0.23, 0.474 (0.06,0.12,0.18)	−0.269, 0.419 (0.06,0.13,0.19)
Total number of signalized intersections					−0.067, 0.348 (0.09,0.13,0.18)			
Total rail kilometers								−0.56, 1.12 (−0.11,0.01,0.15)
Number of schools							−0.26, 0.24 (−0.06,0.00,0.05)	
GWPR AICc	16,304	13,404	14,602	8,137	9,910	8,220	7,629	8,032
Global AICc	27,410	20,431	24,373	20,041	20,161	18,805	18,634	18,779

Minimum, Maximum (Lower Quartile, Median, Upper Quartile).

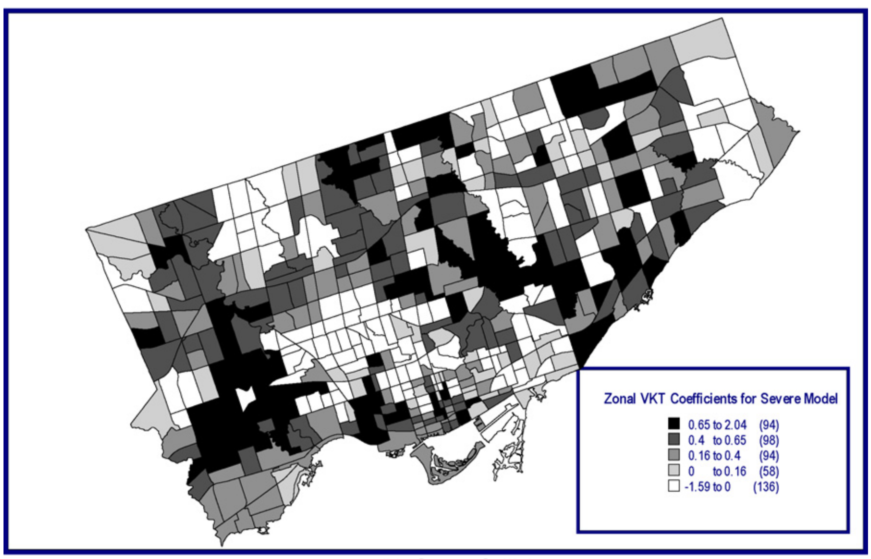
Table 3
Macro-level GWPR collision prediction models based on traffic and road network variables, severe collisions.

Parameters	GWPR model number							
	#1S	#2S	#3S	#4S	#5S	#6S	#7S	#8S
ln(*A*)	−13.50, 15.17 (−1.2,1.14,3.8)	−0.488, 3.761 (2.0,2.47,2.80)	−6.800, 7.130 (−0.31,0.93,2.35)	−7.703, 7.932 (−0.57,1.10,2.74)	−10.139, 11.700 (−0.52,1.38,2.81)	−4.886, 7.160 (−0.19,1.20,2.27)	−6.89, 8.275 (−0.38,1.17,2.37)	−3.561, 7.120 (0.02,1.2,2.1)
ln(VKT)	−1.585, 2.037 (0.03,0.28,0.57)	−0.089, 0.376 (0.01,0.04,0.1)	−0.459, 1.131 (0.08,0.24,0.38)	−0.846, 1.237 (0.02,0.21,0.41)	−1.075, 1.568 (0.00,0.18,0.41)	−0.438, 0.896 (0.05,0.19,0.34)	−0.562, 1.165 (0.04,0.18,0.36)	−0.490, 0.779 (0.07,0.19,0.31)
Total arterial road kilometers		0.087, 0.338 (0.20,0.25,0.27)						
Total expressway kilometers		−0.167, 0.282 (0.08,0.12,0.16)						
Total collector kilometers		−0.038, 0.148 (0.03,0.07,0.09)						
Total laneway kilometers		−8.11, 0.996 (0.02,0.06,0.09)						
Total local road kilometers		−0.043, 0.028 (−0.01,−0.00,0.00)						
Total ramp kilometers		−0.479, 0.280 (−0.03,0.02,0.12)						
Total road kilometers			−0.030, 0.241 (0.03,0.05,0.07)			−0.080, 0.228 (0.00,0.03,0.05)	−0.090, 0.213 (0.00,0.02,0.06)	−0.038, 0.118 (0.00,0.03,0.05)
Number of 4-legged signalized intersections				−0.229,0.736 (0.07,0.15,0.23)		−0.145, 0.478 (0.05,0.11,0.18)	−0.215, 0.473 (0.05,0.11,0.19)	−0.097, 0.392 (0.05,0.09,0.16)
Number of 3-legged signalized intersections				−0.700, 0.676 (0.02,0.14,0.25)		−0.255, 0.735 (0.03,0.11,0.20)	−0.336, 0.747 (0.03,0.12,0.19)	−0.356, 0.436 (0.05,0.12,0.19)
Total number of signalized intersections					−0.110, 0.599 (0.08,0.14,0.21)			
Total rail kilometers								−0.696, 0.991 (−0.11,0.01,0.14)
Number of schools							−0.308, 0.288 (0.007,0.01,0.07)	
GWPR AICc	3,195	3,228	3,251	2,492	2,404	2,295	2,275	2,358
Global AICc	6,742	4,861	5,848	4,885	4,905	4,492	4,448	4,485

Minimum, Maximum (Lower Quartile, Median, Upper Quartile).



(a) Total Collisions



(b) Severe Collisions

Fig. 1. Coefficients of Zonal VKT.

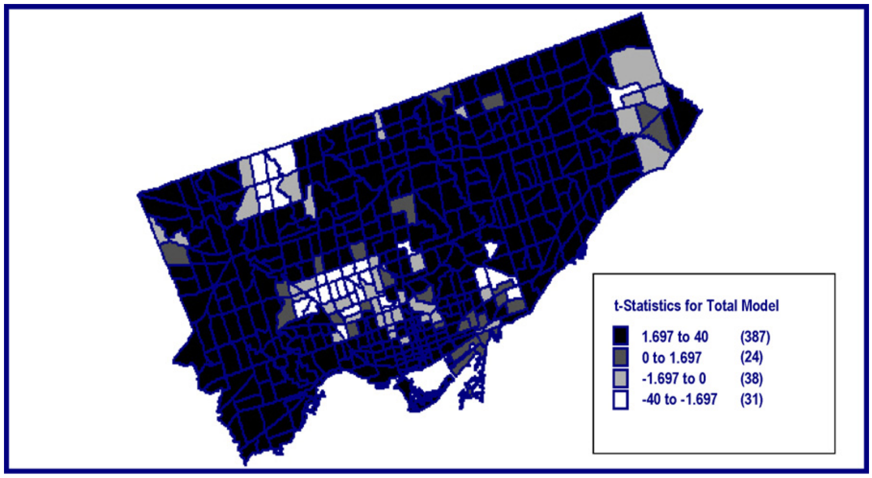
bution presents the median, upper and lower quartiles, along with the minimum and maximum values of the data.

As can be seen from Tables 2 and 3, the signs of coefficients of VKT for each TAZ are not always the same. Traffic Exposure (i.e., VKT) is expected to have a positive effect on the number of collisions in each TAZ; therefore, its coefficient should be positive. Fig. 1a and b depict the local parameters for both total and severe collision models for each TAZ, respectively. It is clear from the Figures that the parameters demonstrate spatial variation. The figures also show that in some of the TAZs the sign of VKT coefficient is negative. The numbers of TAZs with negative sign for the VKT coefficient are more for severe collision models than the one for total collision models.

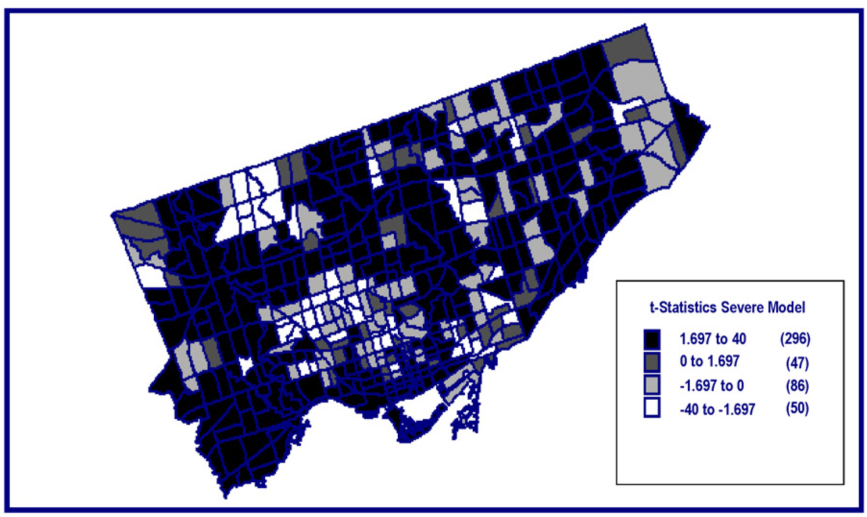
As indicated by Zhao et al. (2005) and Wheeler and Calder (2007), the problem with the counterintuitive signs is not uncommon in GWR or GWPR models. There are two possible reasons for this:

(1) There exists multicollinearity among some variables in the data set for some TAZs, or locations, due to correlation in the estimated coefficients. These variables may be correlated locally and not globally. Wheeler and Tiefelsdorf (2005) indicated that "while GWR coefficients can be correlated when there is no explanatory variable collinearity, the coefficient correlation increases systematically with increasingly more collinearity. The collinearity in explanatory variables can apparently be increased by the GWR spatial kernel weights, and moderate collinearity of locally weighted explanatory variables can lead to potentially strong dependence in the local estimated coefficients, which makes interpreting individual coefficients problematic".
(2) Some variables may be less significant or insignificant at all or some locations. This might be the case because of the method used for estimating the standard errors in the GWR models. These standard errors may only be rough estimates due to the reuse of the data at multiple locations (Congdon, 2003) and also because of using the data to estimate both the bandwidth and regression coefficients (Wheeler and Calder, 2007).

Both of the above phenomena can result in unexpected coefficient signs for some TAZs. A good diagnostics tool is not available to



(a) Total Collisions



(b) Severe Collisions

Fig. 2. *t*-Statistics of Zonal VKT.

examine multi-collinearity in GWRs. Therefore, this possibility cannot be examined. To examine the level of significance of each local variable, local *t*-statistics were computed in order to determine where relationships are significant and where they are not.

Fig. 2a and b shows the results of the *t*-statistics for both total and severe collision models for each TAZ, respectively. The results indicate that the *t*-values for most of the TAZs with negative coefficients are insignificant at the 90% confidence level (TAZs coloured in light grey in Fig. 2a and b). Another observation Fig. 2a and b is that the TAZs with insignificant *t*-values are in close proximity to each other in small pockets distributed throughout the study area.

Unexpected coefficient signs for the rest of TAZs with significant *t*-values could be due to missing or mis-specified explanatory variables in the model. For instance, VKT data were not available by type of road. So the model could not explain the possibility that a zone with more freeway kilometers and higher VKT can have fewer collisions than a neighbouring zone with lower VKT that is mostly on arterial roads that are known to be less safe than freeways, a possibility that can produce a counterintuitive sign for the VHT exponent for both zones.

Finally, the size of TAZs may also play some role for unexpected coefficient signs with significant *t*-values since the TAZs used for this study are of arbitrary size. To address this, further research needs to be conducted in order to explore the effects TAZs size on the GWPR model results.

The AICc associated with the total and severe GWPR models are 16304.6 and 3195.8, respectively. The AICc values for the total and severe global Poisson models are found to be 27410.2 and 6742.0, respectively. A comparison of the AICc values suggests that the GWPR models outperform the global models indicating that there is a spatial non-stationary feature in the relationships being examined.

6.1.2. Models with network characteristics as independent variables (Models 2–8 for total collisions and Models 2S–8S for severe collisions)

Similar to our previous study (Hadayeghi et al., 2007), several models that rely primarily on explanatory variables describing the zonal road network were developed. The 5-number summaries for these models are presented in Tables 2 and 3 for total and severe collisions, respectively. The coefficients of all the variables except

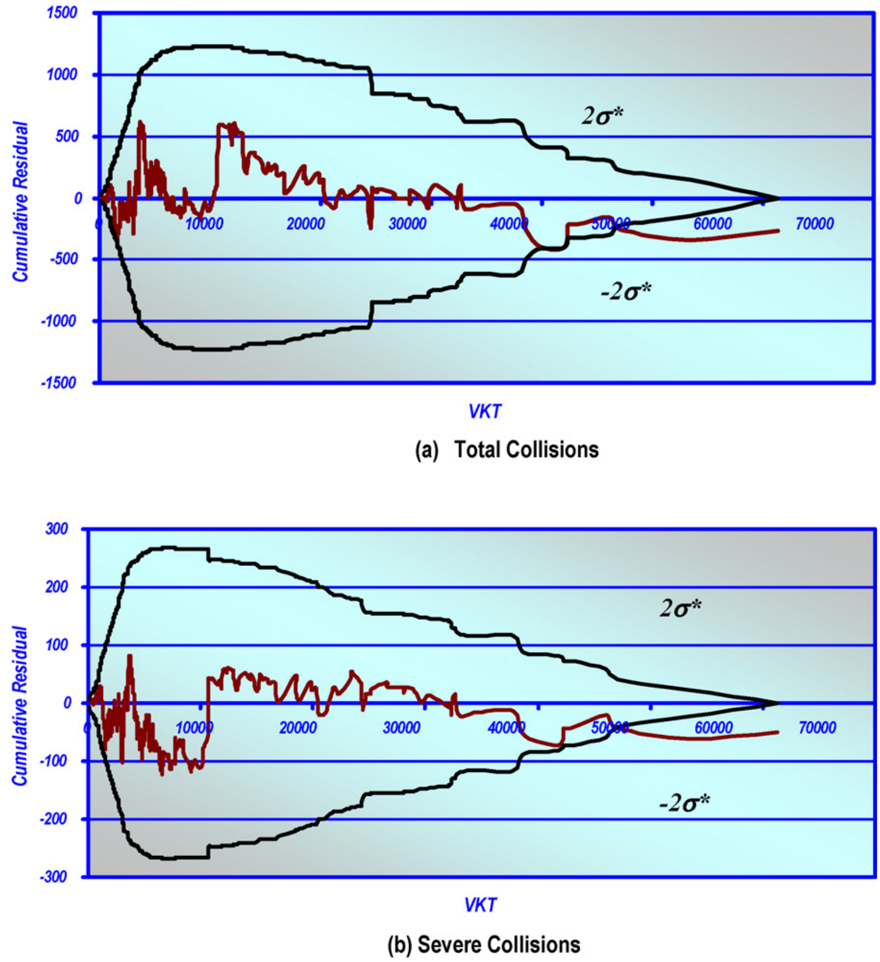


Fig. 3. CURE Plot for Comprehensive Models.

those of total arterial road kilometers in Models 2 and 2S vary from negative to positive values. For each variable in each model, similar analysis to the one for the coefficient of VKT in Models 1 and 1S in the previous section was conducted. The results of *t*-values indicated that most of the unexpected coefficient signs are insignificant at 90% confidence level. However, due to space limitations, the results are not shown in this paper.

The total rail kilometers and total local road kilometers were found to be inversely related to collision frequency in our previous study (Hadayeghi et al., 2007). Interestingly in this study, the lower quartiles for these two variables were found to be negative values, indicating that for these two variables most of the TAZs have negative corresponding coefficients.

The AICc values associated with GWPR models and GLMs are shown in Tables 2 and 3. A comparison of the AICc values suggests that the GWPR models outperform the global models in all cases, suggesting that there is a spatial non-stationary feature in the relationships being examined.

6.1.3. Models with land use, population and employment characteristics as independent variables (Models 9–11 for total collisions and Models 9S–10S for severe collisions)

The relationship between collision frequency and land use, population size and employment was explored by developing GWPR models. This type of model is beneficial to urban planners and other analysts who deal with issues related to zoning and development of neighbourhoods.

The 5-number summaries for the developed models are presented in Tables 4 and 5 for total and severe collisions, respectively. As can be seen from these Tables, the signs of coefficients in the developed models for all TAZ are not always the same since the minimum values are negative for all of the coefficients. However, the lower quartile values for most coefficients are positive except for “other” in Model 9, and number of townhouses in both Models 10 and 10S. The positive value of the lower quartile indicates that most TAZs have positive coefficients.

Models 10 and 10S, shown in Tables 4 and 5 respectively, explore the relationship between the number of collisions and dwelling units. For both total and severe collisions, these models show positive relationships for most the TAZs between collisions and the type of dwelling units, except for townhouses. As explained in the pervious section, one possible reason for a coefficient with a counterintuitive sign, in this case the coefficient for number of townhouses, is that the variable is insignificant in most TAZs. Indeed, in our previous study (Hadayeghi et al., 2007), although the coefficient for the number of townhouses was positive in the GLMs, it was statistically insignificant. The equivalent GLM coefficients for Models 10 and 10S from the previous study (Hadayeghi et al., 2007) are shown in Tables 4 and 5 respectively for comparison purposes.

The AICc values indicate that the developed GWPR models fit the data better than the global GLM models. In the next section, a detailed analysis will be presented in order to examine and compare the goodness of fit measures from the GWPR models with those from the Negative Binomial and Poisson regression GLMs.

Table 4
Macro-level GWPR collision prediction models based on traffic, land use, socioeconomic and demographic and road network variables, total collisions.

Parameters	GWPR model number				GLM #10
	#9	#10	#11	#12	
ln(A)	0.110, 4.842 (1.41,2.37,3.05)	−2.177, 6.450 (0.69,1.93,3.08)	0.345, 4.486 (1.32,2.11,3.04)	−0.181, 4.414 (1.57,2.44,3.18)	23.69
ln(VKT)	−0.038, 0.490 (0.20,0.26,0.35)	−0.202, 0.725 (0.21,0.34,0.47)	0.018, 0.477 (0.20,0.29,0.37)	−0.020, 0.502 (0.13,0.21,0.30)	0.3017
Total road kilometers				−0.015, 0.075 (0.02,0.03,0.04)	
Number of 4-legged signalized intersections				−0.089,0.209 (0.04,0.09,0.12)	
Number of 3-legged signalized intersections				−0.109, 0.251 (0.07,0.12,0.15)	
Commercial (1000 m^2)	−0.0004, 0.005 (0.001,0.002,0.003)		−0.0002, 0.005 (0.001,0.002,0.003)	−0.0007, 0.006 (0.001,0.002,0.003)	
Residential (1000 m^2)	0.000087, 0.0007 (0.0003,0.0003,0.0004)				
Resource and industrial (1000 m^2)	−0.00009, 0.0024 (0.0003,0.0004,0.0005)		−0.0002, 0.002 (0.0003,0.0004,0.0005)	−0.0012, 0.001 (0.0000,0.0003,0.0004)	
Other (1000 m^2)	−0.0012, 0.0009 (−0.0002,0.0001,0.0002)		−0.0008, 0.0008 (−0.0001,0.0001,0.0002)	−0.001, 0.0009 (−0.000,0.000,0.00001)	
Number of Houses × 10^{-3}		−0.119, 0.465 (0.03,0.07,0.11)			0.0314
Number of Apartments × 10^{-3}		−0.126, 0.164 (0.01,0.04,0.07)			0.0385
Number of Townhouses × 10^{-3}		−2.108, 2.311 (−0.19,0.08,0.34)			0.0714
Number of Populations × 10^{-3}				−0.091, 0.123 (−0.003,0.013,0.03)	
Number of Managements × 10^{-3}			−0.092, 0.294 (0.01,0.08,0.15)		
Number of Manufactures-constructions-trades × 10^{-3}			−0.134, 0.680 (0.06,0.27,0.44)		
GWPR AICc	17,667	17,103	17,065	9,288	
Global AICc	25,071	25,794	23,536	18,073	

Minimum, maximum (lower quartile, median, upper quartile).

Table 5
Macro-level GWPR collision prediction models based on traffic, land use, socioeconomic and demographic and road network variables, severe collisions.

Parameters	GWPR model number			GLM #10S
	#9S	#10S	#11S	
ln(A)	−1.191, 3.641 (0.20,0.85,1.84)	−4.007, 5.332 (−0.80,0.56,1.88)	−1.162, 3.297 (0.33,0.93,1.7)	0.9971
ln(VKT)	−0.080, 0.466 (0.18,0.26,0.33)	−0.259, 0.744 (0.17,0.33,0.46)	−0.053, 0.437 (0.13,0.21,0.28)	0.2835
Total road kilometers			−0.013, 0.069 (0.02,0.03,0.04)	
Number of 4-legged signalized intersections			−0.055, 0.182 (0.06,0.09,0.12)	
Number of 3-legged signalized intersections			0.003, 0.313 (0.10,0.13,0.16)	
Commercial (1000 m^2)	−0.0001, 0.004 (0.001,0.0014,0.0019)		−0.0017, 0.004 (0.000,0.000,0.0000)	
Residential (1000 m^2)	0.00000, 0.0008 (0.0003,0.0004,0.0005)			
Resource and industrial (1000 m^2)	−0.00016, 0.0018 (0.0003,0.0005,0.0006)		−0.00079, 0.001 (0.000,0.000,0.0000)	
Government and institutional (1000 m^2)	−0.0009, 0.002 (0.00002,0.0004,0.0009)		−0.0024, 0.0018 (0.000,0.000,0.0000)	
Other (1000 m^2)	−0.001, 0.001 (−0.000,0.0001,0.00003)		−0.0008, 0.001 (0.000,0.000,0.0000)	
Number of Houses × 10^{-3}		−0.125, 0.667 (0.04,0.08,0.12)		0.0442
Number of Apartments × 10^{-3}		−0.240, 0.192 (0.01,0.04,0.09)		0.0399
Number of Townhouses × 10^{-3}		−2.90, 2.71 (−0.22,0.12,0.42)		0.1251
Number of Populations × 10^{-3}			−0.044, 0.1067 (−0.0009,0.009,0.03)	
GWPR AICc	4,496	4,048	2,839	
Global AICc	6,087	6,162	4,338	

Minimum, maximum (lower quartile, median, upper quartile).

Table 6
Goodness of fit measures for total and severe collision models.

		Model number for total collisions											
		1	2	3	4	5	6	7	8	9	10	11	12
r12	NB	0.48	0.51	0.56	0.62	0.62	0.63	0.64	0.46	0.51	0.51	0.49	0.63
	Poisson	0.48	0.60	0.56	0.63	0.63	0.65	0.66	0.47	0.52	0.52	0.51	0.66
	GWPR	0.77	0.89	0.78	0.88	0.85	0.89	0.90	0.72	0.73	0.73	0.73	0.89
MSE	NB	11695	18400	10308	9562	9588	9687	9432	12363	11121	11121	11835	9684
	Poisson	11649	9771	10263	9061	9126	8645	8500	13310	11038	11038	11330	8622
	GWPR	6441	3171	6048	3436	4244	3282	3004	7211	7145	7145	7109	3171
MSPE	NB	11548	18170	10179	9442	9468	9566	9314	12208	10982	10982	11688	9563
	Poisson	11504	9649	10135	8948	9012	8537	8394	13143	10900	10900	11188	8514
	GWPR	6360	3132	5972	3393	4191	3241	2967	7121	7056	7056	7020	3132
		Model number for severe collisions											
		1S	2S	3S	4S	5S	6S	7S	8S	9S	10S	11S	
r12	NB	0.44	0.52	0.55	0.61	0.60	0.63	0.64	0.51	0.51	0.62	0.63	
	Poisson	0.44	0.60	0.55	0.62	0.62	0.65	0.66	0.44	0.52	0.66	0.65	
	GWPR	0.87	0.81	0.83	0.93	0.92	0.92	0.94	0.71	0.76	0.85	0.90	
MSE	NB	624	898	541	515	518	510	498	572	572	514	513	
	Poisson	623	510	538	480	484	449	443	751	569	439	450	
	GWPR	195	265	248	113	122	120	99	387	333	222	147	
MSPE	NB	617	887	534	509	511	503	492	565	565	508	506	
	Poisson	615	504	531	474	477	443	438	742	562	434	444	
	GWPR	193	261	245	112	121	119	98	382	329	219	145	

6.1.4. 6.1.4 Comprehensive models (Model 12 for total collisions and Model 11S for severe collisions)

These models were calibrated using all groups of planning variables—land use, network, traffic intensity and socioeconomic and demographic. Although these models include all available planning variables, their potential for use in some other regions might be limited due to the relatively large numbers of explanatory variables and the practical reality that data for some of these independent variables might not be available.

The 5-number summaries for the developed models in this section are presented in Tables 4 and 5 for total and severe collisions, respectively. Similar to the models presented in the previous section, unexpected signs for the coefficients pertaining to some TAZs were encountered.

The cumulative residuals were produced for model 12 for total collisions and model 11S for severe collisions. The cumulative residuals are presented for the total and severe collisions in Fig. 3a and b, respectively. The figures show that the cumulative residuals stay between the two standard deviation ($\pm 2\sigma^*$) boundaries for both models. However, both models tend to underestimate the collision counts for lower traffic flows, and overestimate them for higher traffic flows, suggesting that there might be a need to introduce additional explanatory variables which must be weighed against the disadvantages of over-fitting.

6.2. Comparison of results of GWPR models with GLM models

Table 6 presents the relative goodness of fit measures for predicting the number of collisions using GWPR models and GLM Negative Binomial and Poisson regression models for total and severe collisions, respectively. Goodness of fit statistics provide an ability to objectively assess the fit of a model to the data. Large differences in goodness of fit measures of subject models are indicative of their varying performance in predicting the number of collisions. The results of the linear correlation coefficients show the improved performance of all GWPR models. As indicated earlier, a comparison of the MSPE and the MSE shows potential for over or under fitting of the models to the estimation data. If the MSE of a model is higher than the MSPE of another, this indicates that the first model may have been over fitted to the estimation data. The MSE values for predicting the 2001 total and severe yearly zonal collisions using the GWPR models are lower than the MSPE values for the corresponding Negative Binomial models and Poisson GLMs, indicating that the variability in the 2001 collision frequency is better captured by the GWPR models than the GLMs.

7. Summary and conclusions

A common technique used for the calibration of safety planning models is the GLM procedure with the assumption of Negative Binomial or Poisson error distribution. In this technique, a fixed coefficient representing the average relationship between the dependent variable, typically number of collisions per traffic analysis zone, and each independent variable is calibrated under the assumption that this relationship does not vary across space. However, collision frequency is influenced by many spatially defined factors such as land use, demographic characteristics, and traffic volume patterns.

The primary objective of this study was to investigate the spatial variations in the relationship between the number of zonal collisions and potential transportation planning predictors through the development of GWPR models that allow the model coefficient estimates to vary spatially. This is a very essential trait of a model as it adds uniqueness to the spatial location of an event such as collision occurrence. The second objective of the paper was to compare the accuracy of GWPR models to the accuracy of GLMs with Negative Binomial and Poisson error distributions using a series of goodness of fit measures.

Twenty-three collision prediction models were developed using land use, traffic intensity, road network, and socioeconomic and demographic zonal characteristics as explanatory variables. A GWPR approach was employed for the development of models total collisions and for severe (fatal and injury) collisions.

Although the GWPR models produce unexpected signs for the coefficients pertaining to some TAZs for the reasons stated, the results show 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 can improve safety analysis of transportation networks. The research outcomes can help advance the state of the art in road safety research by enabling safety to be included among other traditional evaluation criteria in the strategic planning process for transportation systems.

Although the developed models seem to perform well in explaining the relationship of collisions with considered variables, there is still a need for further improvement of the safety planning models. The developed GWPR models are excellent tools for predicting the number of zonal collisions for a future planning year. These models can be temporally transferred. However, these models are not spatially transferable, since they produce a local coefficient for each TAZ in a specific geographic region. This is a relatively minor difficulty in that most jurisdictions tend to develop their own models.

References

Aguero-Valverde, J., Jovanis, P.P., 2006. Spatial analysis of fatal and injury crashes in Pennsylvania. Accident Analysis and Prevention 38 (3), 618–625.

Cameron, A.C., Trivedi, P.K., 1998. Regression Analysis of Count Data. Cambridge University Press.

Congdon, P., 2003. Modelling spatially varying impacts of socioeconomic predictors on mortality outcomes. Journal of Geographical Systems 5 (2), 161–184.

de Guevara, F.L., Washington, S.P., Oh, J., 2004. Forecasting crashes at the planning level: simultaneous negative binomial crash model applied in Tucson, Arizona. Transportation Research Record 1897, 191–199.

Du, H., Mulley, C., 2006. The relationship between transport accessibility and land value: a local model approach with geographically weighted regression. In: Proceedings of the 85th Annual Meeting of the Transportation Research Board. National Research Council, Washington, DC.

EMME/2 User's Guide Manual Release 9. 1998 INRO Consultants INC.

Fotheringham, A.S., Brunsdon, C., Charlton, M., 2002. Geographically Weighted Regression the analysis of spatially varying relationships. John Wiley & Sons, Ltd.

Guo, L., Ma, Z., Zhang, L., 2008. Comparison of bandwidth selection in application of geographically weighted regression: a case study. Canadian Journal of Forest Research 38 (9), 2526–2534.

Hadayeghi, A., Shalaby, A.S., Persaud, B., 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., Cheung, C., 2006. Temporal transferability and updating of zonal level accident prediction models. Accident Analysis and Prevention 38 (3), 579–589.

Hadayeghi, A., Shalaby, A.S., Persaud, B., 2007. Safety prediction models: proactive tool for safety evaluation in urban transportation planning applications. Transportation Research Record 2019, 225–236.

Hadayeghi, A., 2009. Use of advanced techniques to estimate zonal level safety planning models and examine their temporal transferability. Ph.D. Thesis. Department of Civil Engineering, University of Toronto.

Hauer, E., 1997. Observational Before-After Studies in Road Safety: Estimating the Effect of Highway and Traffic Engineering Measures on Road Safety. Pergamon, Elservier Science Ltd., Oxford, UK.

Interactive Highway Safety Design Model (IHSDM). Federal Highway Administration, US Department of Transportation. http://www.tfhrc.gov/safety/ihsdm/ihsdm.htm, accessed in May 2009.

LeSage, J.P., 2001. A Family of Geographically Weighted Regression Models. Department of Economics, University of Toledo. http://www.spatial-econometrics.com/html/bgwr.pdf, accessed in May 2009.

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

Levine, N., 2004. CrimeStat III: A Spatial Statistics Program for the Analysis of Crime Incident Locations (version 3.0). Ned Levine & Associates/National Institute of Justice, Houston, TX/Washington, DC.

Lovegrove, G.R., Sayed, T., 2005. Macro-level collision prediction models for evaluating neighbourhood traffic safety. In: Proceedings of Presented at 6th Transportation Specialty Conference. Canadian Society of Civil Engineering, Toronto, Ontario.

Miaou, S.P., 1994. The relationship between truck accidents and geometric design of road sections: poisson versus negative binomial regressions. Accident Analysis and Prevention 26 (4), 471–482.

Miaou, S.P., Song, J.J., Mallick, B.K., 2003. Roadway traffic crash mapping: a space-time modeling approach. Journal of Transportation Statistics, Bureau of Transportation Statistics 6 (1), 33–57.

Nakaya, T., 2001. Local spatial interaction modelling based on the geographically weighted regression approach. GeoJournal 53, 347–358.

Nakaya, T., Fotheringham, A.S., Brunsdon, C., Charlton, M., 2005. Geographically weighted poisson regression for disease association mapping. Statistics in Medicine 24 (17), 2695–2717.

Oh, J., Lyon, C., Washington, S., Persaud, B., Bared, J., 2003. Validation of FHWA crash models for rural intersection lessons learned. Transportation Research Record 1840, 41–49.

Park, N., Zhao, F., 2004. Estimation of AADT Using geographically weighted regression models. In: Proceedings of the 83rd Annual Meeting of the Transportation Research Board. National Research Council, Washington DC.

Silverman, B.W., 1986. Density Estimation for Statistics and Data Analysis. Statistics and Applied Probability. Chapman and Hall, London.

Wang, W., Abdel-Aty, M., 2006. Temporal and spatial analyses of rear-end crashes at signalized intersections. Accident Analysis and Prevention 38 (6), 1137–1150.

Washington, S., Schalkwyk, I.V., Meyer, M., Dumbaugh, E., Zoll, M., 2006. Incorporating safety into long-range transportation planning. NCHRP Report 546, TRB. National Cooperative Highway Research Program, Washington D.C.

Wheeler, D., Calder, C., 2007. An assessment of coefficient accuracy in linear regression models with spatially varying coefficients. Journal of Geographical Systems. 9 (2), 145–166.

Wheeler, D., Tiefelsdorf, M., 2005. Multicollinearity and correlation among local regression coefficients in geographically weighted regression. Journal of Geographical Systems. 7 (2), 161–187.

Zhao, F., Chow, L., Li, M., Liu, X., 2005. A Transit ridership model based on geographically weighted regression and service quality variables. Report DO97591, Lehman Center for Transportation Research, Department of Civil and Environmental Engineering, Florida International University.