Accident Analysis and Prevention 127 (2019) 118–133

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

How instantaneous driving behavior contributes to crashes at intersections: Extracting useful information from connected vehicle message data

Ramin Arvin, Mohsen Kamrani*, Asad J. Khattak

The University of Tennessee, Knoxville, United States

ARTICLE INFO

Keywords:
Volatility
Safety
Rear-end
Head-on
Crash frequency
Big data
Connected vehicles
Random parameter
Geographically Weighted Regression

ABSTRACT

Connected and automated vehicles have enabled researchers to use big data for development of new metrics that can enhance transportation safety. Emergence of such a big data coupled with computational power of modern computers have enabled us to obtain deeper understanding of instantaneous driving behavior by applying the concept of "driving volatility" to quantify variations in driving behavior. This paper brings in a methodology to quantify variations in vehicular movements utilizing longitudinal and lateral volatilities and proactively studies the impact of instantaneous driving behavior on type of crashes at intersections. More than 125 million Basic Safety Message data transmitted between more than 2800 connected vehicles were analyzed and integrated with historical crash and road inventory data at 167 intersections in Ann Arbor, Michigan, USA. Given that driving volatility represents the vehicular movement and control, it is expected that erratic longitudinal/lateral movements increase the risk of crash. In order to capture variations in vehicle control and movement, we quantified and used 30 measures of driving volatility by using speed, longitudinal and lateral acceleration, and yaw-rate. Rigorous statistical models including fixed parameter, random parameter, and geographically weighted Poisson regressions were developed. The results revealed that controlling for intersection geometry and traffic exposure, and accounting unobserved factors, variations in longitudinal control of the vehicle (longitudinal volatility) are highly correlated with the frequency of rear-end crashes. Intersections with high variations in longitudinal movement are prone to have higher rear-end crash rate. Referring to sideswipe and angle crashes, along with speed and longitudinal volatility, lateral volatility is substantially correlated with the frequency of crashes. When it comes to head-on crashes, speed, longitudinal and lateral acceleration volatilities are highly associated with the frequency of crashes. Intersections with high lateral volatility have higher risk of head-on collisions due to the risk of deviation from the centerline leading to head-on crash. The developed methodology and volatility measures can be used to proactively identify hotspot intersections where the frequency of crashes is low, but the longitudinal/lateral driving volatility is high. The reason that drivers exhibit higher levels of driving volatility when passing these intersections can be analyzed to come up with potential countermeasures that could reduce volatility and, consequently, crash risk.

1. Introduction

The need for a safer and more sustainable transportation system has pushed the public and private sectors to improve the performance of the network. Connected Vehicles (CV) provide enriched data such as instantaneous driving behavior, maneuvers, trajectory, individual origin and destination, and traffic data which previously were not obtainable. These data can be transmitted via vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communication which can be incorporated to gain precise information to monitor and evaluate the performance of the system (Ghiasi et al., 2017; Nezafat et al., 2018). The National Highway Traffic Safety Administration (NHTSA) has announced that communication between vehicles will become mandatory in the near future. In order to advance V2V and V2I technology, the U.S. Department of Transportation developed the Safety Pilot Model Deployment (SPMD) study. The SPMD is one of the most successful studies to implement V2V and V2I communication in the real-world environment (Henclewood et al., 2014), and is one of the largest vehicle communication test-bed by incorporating more than 2800 instrumented vehicles and more than 70 miles of roadway instrumented with Road Side Units (RSU) in Ann Arbor, MI (Henclewood et al., 2014). In this experiment, CVs and RSUs were capable of communicating via Dedicated Short-Range Communication (DSRC) at a frequency of 10 Hz (Henclewood et al., 2014). The emergence of Big Data provided by CVs,

* Corresponding author.
E-mail address: mkamrani@vols.utk.edu (M. Kamrani).

https://doi.org/10.1016/j.aap.2019.01.014
Received 20 August 2018; Received in revised form 9 January 2019; Accepted 10 January 2019
Available online 07 March 2019

RSUs, and other sources of information provides opportunities for researchers to innovate and implement new concepts aiming to increase safety, mobility and moves toward sustainability.

From the safety perspective, previous studies reveal that rear-end and sideswipe crashes are the most frequent type of crash at signalized intersections (Wang and Abdel-Aty, 2006). On average, rear-end and sideswipe crashes are the least dangerous type of collision, while head-on and angle crashes are the most dangerous ones (Paleti et al., 2010). According to U.S. traffic safety facts for the year 2015, while 4.1% of all crashes were head-on collisions, they contribute to 10.2% of fatal crashes (National Highway Traffic Safety Administration, 2001). As a result, researcher pay a great amount of attention to decrease the frequency and severity of head-on crashes.

Given the importance of type of crashes, this study explores the impact of instantaneous driving behavior on multiple crash types at intersections. The study utilizes “driving volatility”, a newly developed concept in transportation (Kamrani et al., 2018a, b, c; Wang et al., 2015) which captures variations in vehicular movements, as an indicator for driving behavior at intersections. This study extends the concept of driving volatility to longitudinal and lateral volatilities and explores the correlation between volatilities with rear-end, sideswipe, angle and head-on crashes. The main goals of this research are to:

1) Develop a framework for capturing and quantifying longitudinal and lateral driving volatilities using real-world instantaneous driving data.
2) Evaluate correlations between longitudinal and lateral volatilities with frequency of multiple crash types at intersections.
3) Account for unobserved heterogeneity by utilizing random parameter and semi-parametric geographically weighted Poisson regression models.

Since human-error contributes to 94 percent of crashes in the U.S (National Highway Traffic Safety Administration, 2008), findings from this study can help agencies proactively identify hazardous intersections where there is a substantial variations in driving behavior by utilizing the concept of driving volatility. Proactively countermeasures might apply to reduce driving volatilities to prevent future crashes.

2. Literature Review

Numerous studies have focused on capturing associations between crash frequency and the geometric characteristics and traffic factors at intersections or road segments. The most favorable method for finding relationships between these variables are statistical count models due to the non-negative, discrete, and randomness nature of crashes.

Focusing on modeling, various methods were utilized for capturing the impact of explanatory variables on crash frequencies, among which fixed parameter models are the simplest. In this approach, the estimated parameters are not allowed to vary across the data (e.g., the effect of Average Annual Daily Traffic (AADT) is constant across all the intersections). However, due to presence of unobserved variations among intersections, one might expect that some of the estimated coefficients vary across intersections, elaborating the model estimation process (Anastasopoulos and Mannering, 2009; Washington et al., 2010). To address this issue, different promising approaches were developed by researchers such as random-effect and random-parameter models that have been widely used in crash frequency modeling (Castro et al., 2012; El-Basyouny and Sayed, 2009; Wu et al., 2013). The main objective of these approaches is to handle temporal and spatial correlations and account for unobserved heterogeneity among observations (Wali et al., 2018a, b,c). However, models might not be transferrable to other datasets (Lord and Mannering, 2010). Geographically Weighted Poisson Regression (GWPR) is another method for capturing the spatial variations across observations. This method has the same spirit and methodology as local generalized linear regression method, but there is a different process for determining the weights (Loader, 2006). It has shown that GWPR models outperformed traditional statistical models (i.e. the Poisson model) in terms of capturing spatial variations among crash counts and independent variables (Fotheringham et al., 2003). In the literature, most papers consider spatial variations in all of the predicting factors, while in some cases, degrees of variation for some parameters might be negligible. Therefore, it is necessary to apply semi-parametric Geographically Weighted Regression (S-GWPR) models in which some of the factors are global (Xu and Huang, 2015). It should be noted that Random Parameter (RP) Poisson regression and GWPR methods are intrinsically different. The coefficients in RP Poisson models are drawn independently from a univariate distribution, disregarding the locations of the observations, while in GWPR the coefficients are derived from coordinates in the geographical space (Xu and Huang, 2015).

In the literature, while various location characteristic were considered in crash frequency modeling such as intersection density (Huang et al., 2010), skew angle (Nightingale et al., 2017), congestion and traffic flow (Stipancic et al., 2017; Wang et al., 2009), traffic patterns (Noland and Quddus, 2005), environmental and weather conditions (Ghasemzadeh and Ahmed, 2018a; Lee and Abdel-Aty, 2005), and signal characteristics (Agbelie and Roshandeh, 2015), driver behavior factors received less attention. In the U.S., more than 50 percent of all fatal crashes were caused by aggressive driving behaviors such as speeding, reckless driving, and failure to yield the right of way (American Automobile Association, 2009). In the literature, in order to quantify the variations in normal driving behavior, common vehicle kinematics are widely used (Ahmed and Ghasemzadeh, 2018; Ghasemzadeh and Ahmed, 2017, 2018b). Recently, the term “driving volatility” was introduced (Wang et al., 2015) which attempts to describe the driving behavior performance. In order to define volatility, researchers have applied different measurements to the kinematic features of vehicles such as speed (Arvin et al., 2019a, b; Kamrani et al., 2018a, b; Kamrani et al., 2019, 2017; Wang et al., 2015), acceleration (Arvin et al., 2019a, b; Kamrani et al., 2018b; Kamrani et al., 2019; Wang et al., 2015) and jerk (Kamrani et al., 2018b; Wang et al., 2015). Moreover, some studies (Kamrani et al., 2017) have looked at the impact of volatility on the safety performance of traffic networks. However, in the previous studies several gaps exist. First, the aforementioned studies ignored the variations in lateral movement of the vehicle and only focused on longitudinal volatility. Second, they modeled total number of crashes at intersections, while the impact of driving volatility might vary among different crash types. Finally, they ignored the spatial variations among intersections.

The contribution of this paper is addressing the aforementioned gaps by extending the concept of volatility to longitudinal and lateral volatilities in order to quantify the variations in longitudinal and lateral control of the vehicle. By incorporating large scale Basic Safety Messages (BSM) data transmitted between CVs in real-world environment, 30 measures of volatilities were developed to explore the impact of these measures on the frequency of rear-end, sideswipe, angle and head-on crashes. Our hypothesis is variations in longitudinal and lateral vehicle movement is associated with the frequency of various crash types, controlling for other variables (e.g. traffic exposure, number of legs, number of lanes, etc.). To address the unobserved heterogeneity and spatial correlation, the random parameter and S-GWPR model was employed, and the performance of the models were compared with the fixed parameter Poisson regression.

Considering the emergence of such a big data, this study is timely and original by bringing in the concept of longitudinal and lateral volatilities to quantify the variations in instantaneous driving behavior and explore its association with rear-end and head-on crashes.

3. Methodology

3.1. Modeling approach

Traditionally, to model the crash frequency, the count-data models such as Poisson, Negative Binomial and Zero Inflated Models are commonly utilized (Abdel-Aty and Radwan, 2000; Azizi and Sheikholeslami, 2012; Jamali and Wang, 2017) due to the fact that crash counts are non-negative integer values in a specific period of time (Anastasopoulos and Mannering, 2009). In this study, fixed parameter Poisson regression model, the random parameter Poisson regression model, and the geographically weighted Poisson regression model (GWPR) were used to model crash frequency.

3.1.1. Poisson model

In the Poisson regression model, the probability of occurrence of n crashes at intersection i can be written as (Greene, 2003):

$$P(n_i) = \frac{\lambda_i^{n_i}\exp(-\lambda_i)}{n_i!} \tag{1}$$

where λ_i (Poisson parameter) is the expected number of crashes for intersection i, $E(n_i)$. In order to fit the regression model, the Poisson parameter, λ_i, is written in the logarithm form (Greene, 2003):

$$\ln(\lambda_i) = \beta X_i \text{ (2)} \tag{2}$$

where X_i is the matrix of the independent variables and β is a vector of the estimated coefficients. The Poisson function defined in Eqs. (1) and (2) is maximized by the maximum likelihood with the following function (Washington et al., 2010):

$$L(\beta) = \prod_i \frac{\exp[-\exp(\beta X_i)][\exp(\beta X_i)]^n}{n_i!} \tag{3}$$

It should be noted that in cases where the mean and the variance of the dependent variable are not equal, applying the Poisson regression might lead to misleading results. Therefore, in order to test the over-dispersion existence in the Poisson model, the Lagrange multiplier method was performed (Greene, 2003). We can write:

$$LL = \left(\frac{\sum_{i=1}^{N}((y_i - \mu_i)^2 - y_i)}{2\sum_{i=1}^{N}\mu_i^2}\right)^2 \tag{4}$$

where y_i and μ_i are the observed and predicted crash frequency at the intersection i, and N is the number of intersections.

3.1.2. Random parameter Poisson model

In this approach, unobserved heterogeneity, arising from unobserved contributing factors, is addressed by developing a random parameter model using simulated maximum likelihood estimation (Greene, 2003). The RP Poisson regression model is an important method because it accounts for heterogeneity arising from factors relating to traffic characteristics, vehicle types, road geometry, pavement conditions, time of day and other unobserved factors (Anastasopoulos and Mannering, 2009). The formulation for estimating the coefficients of the RP Poisson model is (Greene, 2003):

$$\beta_i = \beta + \varphi_i \tag{5}$$

where φ_i is a randomly distributed term with a specified distribution. The log-likelihood function is (Anastasopoulos and Mannering, 2009):

$$LL = \sum_i ln\int_{\varphi_i}^{i} g(\varphi_i)P(n_i\ |\varphi_i)d\varphi_i \tag{6}$$

where g(.) is the pre-specified distribution of φ_i. In this study, the Halton draws simulation approach is utilized, which is the most popular simulation approach as it provides a more efficient distribution than other methods (Bhat, 2003; Train, 2000).

3.1.3. Geographically weighted Poisson regression model

The availability of geo-referenced crashes coupled with computational power has enabled researchers to develop rigorous geospatial models that account for spatial heterogeneity by allowing parameters to vary across space (Xu and Huang, 2015). The Geographically Weighted Poisson Regression (GWPR) can be used to test whether the relationship between the explanatory variables and the dependent variable substantially varies across space (Fotheringham et al., 2003; Liu et al., 2017). The model can be written as:

$$\ln(\lambda_i) = \beta_0(u_i,\ v_i) + \beta_1(u_i,\ v_i)\ln(E_{vi}) + \sum_{k=1}^{K}\beta_k(u_i,\ v_i)x_{ik} + \varepsilon_i \tag{7}$$

where $(u_i,\ v_i)$ denotes the coordinates of i. It should be noted that in GWPR, $\beta_k(u_i,\ v_i)$ is not randomly distributed, but rather is a function of the location i. The following equation can be used to estimate $\beta_k(u_i,\ v_i)$:

$$\hat{\beta}(u_i,\ v_i) = (X^TW(u_i,\ v_i)X)^{-1}X^TW(u_i,\ v_i)Y \tag{8}$$

where $\hat{\beta}(u_i,\ v_i)$ is the vector of estimated coefficients at location i, X is the matrix of independent variables, Y is the $n \times 1$ vector of the number of crashes at each intersection, and $W(u_i, v_i)$ is $n \times n$ spatial weight matrix:

$$W(u_i,\ v_i) = \begin{bmatrix} w_{i1} & 0 & \dots & 0 \\ 0 & w_{i2} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & \dots & \dots & w_{in} \end{bmatrix} \tag{9}$$

where w_{ij} is the weight of variable j at location i. In this approach, based on observations at nearby areas, a regression equation is estimated for each location. Based on the distance from the regression point each area is weighted (areas that are closer have a higher weight than ones that are farther). The W matrix can be estimated using the adaptive Gaussian Kernel function:

$$w_{ij} = \exp\left(-\frac{d_{ij}^2}{\theta_{i(N)}{}^2}\right) \tag{10}$$

where d_{ij} is the Euclidean distance between area i and j, $\theta_{i(N)}$ is the adaptive bandwidth defined by the N^{th} nearest neighbor. In this formulation, the Gaussian Kernel bandwidth is adaptive, meaning that the weight function magnitude varies across all intersections.

In this study, along with the adaptive Gaussian kernel, adaptive bi-square kernel was considered, which can be written as:

$$w_{ij}\begin{cases}\left(1-\left({}^{d_{ij}}/_{d_{iN}}\right)^2\right)^2 & if\ d_{ij} < d_{iN} \\ 0 & otherwise\end{cases} \tag{11}$$

where d_{iN} denotes the distance to the N^{th} nearest neighbor of intersection i.

It is worth mentioning that applying fixed bandwidth kernel, the local coefficients in areas with sparse intersections is estimated with limited points, leading to high standard error in estimation and unreliable results. Thus, in this study adaptive kernel was employed which tries to overcome this issue by letting the bandwidth vary based on the data's sparsity. To determine the bandwidth of the adaptive kernel, the corrected Akaike Information Criteria (AICc) (Hurvich et al., 1998) was used. The best model is the one with the lowest AICc score (Fotheringham et al., 2003; Hadayeghi et al., 2010).

As previously mentioned, there was a probability that some of the coefficients in the model do not significantly vary across space. In this case, the semi-parametric GWPR (S-GWPR) is ideal where some of the parameters vary spatially, while others are held fixed. We can write (Nakaya et al., 2005; Xu and Huang, 2015):

$$\ln(\lambda_i) = \beta_0(u_i,\ v_i) + \beta_1(u_i,\ v_i)\ln(E_{vi}) + \sum_{j=2}^{l}\beta_j x_{ij} + \sum_{k=1}^{k}\beta_k(u_i,\ v_i)x_{ik} + \varepsilon_i \tag{12}$$

where β_j is the j^{th} estimated global variable. In order to evaluate the existence of variation in the estimated coefficients across space (spatial variation), the non-stationarity test was performed. Given 167 intersections, the GWPR model suggests specific coefficients for each observation. The non-stationarity test calculates the difference between the upper and lower quartile of the estimated coefficients from GWPR and performs the evaluation. We can write:

$$Delta = \beta_{upper} - \beta_{lower} \tag{13}$$

$$\begin{cases} \text{Delta} > 1.96 & \text{Pass the test (}local\text{ coefficient)} \\ \quad {}^{*}SE \text{ and Delta} > \max(|t_i|) & \\ \qquad \text{if not} & \text{failed to pass (global coefficient)} \end{cases} \tag{14}$$

where *SE* is the standard error of the coefficient in the global Poisson model, and $|t_i|$ is the significance t-value of the GWPR model at intersection i which can be calculated as $\left|\frac{\beta(u_i, v_i)}{SE(u_i, v_i)}\right|$. If *Delta* is greater than $1.96{*}SE$ and max of $|t_i|$ is greater than 1.96, then the test is passed and there are substantial variations among the estimated coefficients across the space. Otherwise, the test failed and the coefficient is considered as the global coefficient. Obviously, if all the variables are estimated as local coefficients, the S-GWPR model is equivalent to the GWPR model. For further details regarding the S-GWPR calibration, please refer to (Nakaya et al., 2005).

It should be noted that GWPR provides a set of local coefficients at each intersection. To map the GWPR results across space, the Inverse Distance Weighted (IDW) method was applied (Bartier and Keller, 1996). The goal of this approach is to create a continuous coefficient surface that interpolates and maps the results across the space. IDW assigns the value to unknown locations based on the estimated coefficients for the nearby areas. The assigned value obtained by weighting the nearby coefficients based on their distance from the unknown point. We can write:

$$\hat{Z}(s_0) = \sum_{i=1}^{N} \lambda_i Z(s_i) \tag{15}$$

where $\hat{Z}(s_0)$ is the predicted coefficient at location s_0, N is number of known sample points surrounding the location s_0, λ_i are the assigned weights to each measured coefficient, and $Z(s_i)$ is the observed coefficient at location s_i. To determine the weights, we can write:

$$\lambda_i = \frac{d_{i0}^{-2}}{\sum_{i=1}^{N} d_{i0}^{-2}} \tag{16}$$

where d_{i0} represents the distance between point i and o. It can be inferred that by increasing the distance between the unknown coefficient and observed coefficient, the weight of the observed point will decrease.

In order to estimate S-GWPR model, GWR4.0 software which is developed by Nakaya et al. (Nakaya et al., 2012) was used.

3.2. Location-based volatility, measures and calculation

The concept of location-based volatility attempts to develop a meaningful process on instantaneous driving behavior and decisions in order to generate driving volatility measures at intersection/segment level (Wali et al., 2018a). These volatility measures can potentially be representative of the driving behavior of majority of drivers passing the study area (Wali et al., 2018a). Such volatility indices can be utilized to identify locations that driving behavior is different compared to driving behavior of same drivers at other locations. In addition, the correlation between volatility measures and frequency of various crash types can be investigated.

Multiple volatility measures were used by researchers to capture variations in longitudinal control of the vehicle (Arvin et al., 2019a, b; Kamrani, Arvin, et al., 2018b; Kamrani et al., 2019, 2017; Wang et al., 2015) and have been applied to speed, acceleration, and jerk. However, one of the main drawbacks of the previous studies is ignorance of lateral movement of vehicles which potentially could be contributing to crash frequency. Therefore, in this study, volatility functions were applied to speed, longitudinal, lateral acceleration, and yaw-rate at the level of intersections, which were available from connected vehicle BSM data from SPMD. The data is representative of 3–4 percent of total driving in Ann Arbor, MI (Shou and Di, 2018). The data provides high-resolution microscopic driving decisions and vehicle motions in terms of position, speed, acceleration, and yaw-rate with a frequency of 10 Hz. Given that, three groups of volatilities are identified and calculated for the selected 167 intersections in Ann Arbor, MI (discussed later in details):

1 Speed volatility
2 Longitudinal acceleration volatility
3 Lateral acceleration volatility.
4 Yaw-rate volatility

In order to process and calculate volatility indices at intersection level, 150-ft polygons were established from the center of intersections,[1] and BSM data was assigned to each intersection by processing more than 220 million BSM data. It should be noted that due to the difference in speed profile of vehicles at signalized and unsignalized intersections, and signal timing of signalized intersections, zero speeds were removed from the data prior to volatility calculation. For selected intersection, multiple volatility functions are applied on the speed, longitudinal and lateral acceleration, which presented in Table 1. For more details on volatility functions, please refer to (Kamrani et al., 2018b).

Finally, 30 measures of volatility at the aggregate level of intersections were defined among which six measures capture speed volatility, sixteen measures quantify longitudinal and lateral acceleration volatility, and 8 measures capture yaw-rate driving volatility.

3.3. Measures of goodness of fit

In order to evaluate and compare the performance of traditional Poisson regression, RP Poisson, and GWPR, four statistics were utilized to measure estimation accuracy.

1 *R-squared for Poisson model*: this statistic assesses the overall goodness of fit of model based on standardized residuals. Larger values of $R^2_{Poisson}$ (max is 1) indicate better fit. It is defined as (Cameron and Windmeijer, 1996):

$$R^2_{Poisson} = 1 - \frac{\sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2/\hat{Y}_i}{\sum_{i=1}^{n} (Y_i - \bar{Y})^2/\bar{Y}} \tag{17}$$

where Y_i and $\hat{Y}_i$ are the observed and predicted number of crashes at location i respectively, and $\bar{y}$ is the average number of crashes.

2 *AIC*: a lower AIC represents a better goodness of fit (Bozdogan, 1987). A three point decrease in an AIC value indicates a significant improvement in the goodness of fit (Bozdogan, 1987). We can write:

$$AIC = D + 2k \tag{18}$$

where D denotes the model deviance, and k is the number of

[1] Although 250-ft threshold from the center of intersection is a common threshold as an intersection influence area, in this study we chose 150-ft threshold due to two main reasons. First, the network of Ann Arbor city is dense, and intersections are close, and using 250-ft threshold leads to overlapping territories. Therefore, 150-ft represents the intersection influence area. Second, the crash and road inventory data of Ann Arbor, which obtained from MPO of Ann Arbor, is identified based on 150-ft threshold.

Table 1
Functions of volatility.

Measures of volatility	*Formulation*
Standard Deviation	$S_{dev} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$
Coefficient of Variation	$C_v = \frac{SD}{\bar{x}} * 100$
Mean Absolute Deviation	$D_{mean} = \frac{1}{n}\sum_{i=1}^{n}\lvert x_i - \bar{x}\rvert$
Quartile Coefficient of Variation	$Q_{cv} = \frac{Q_3 - Q_1}{Q_3 + Q_1} * 100$
Percent of extreme values	$\%T = \frac{c > Threshold}{n} * 100$ $Threshold = \bar{x} \pm z * s$

parameters. In the S-GWPR, due to the non-parametric framework of the model, the number of parameters is meaningless. Therefore, an effective number of parameters should be calculated which can be written as (Nakaya et al., 2005):

$$K = trace(S) \tag{19}$$

where S is the hat matrix. For more details, please see (Nakaya et al., 2005).

3 *Mean Absolute Deviation:* a smaller value of MAD implies a better model estimation. It can be defined as:

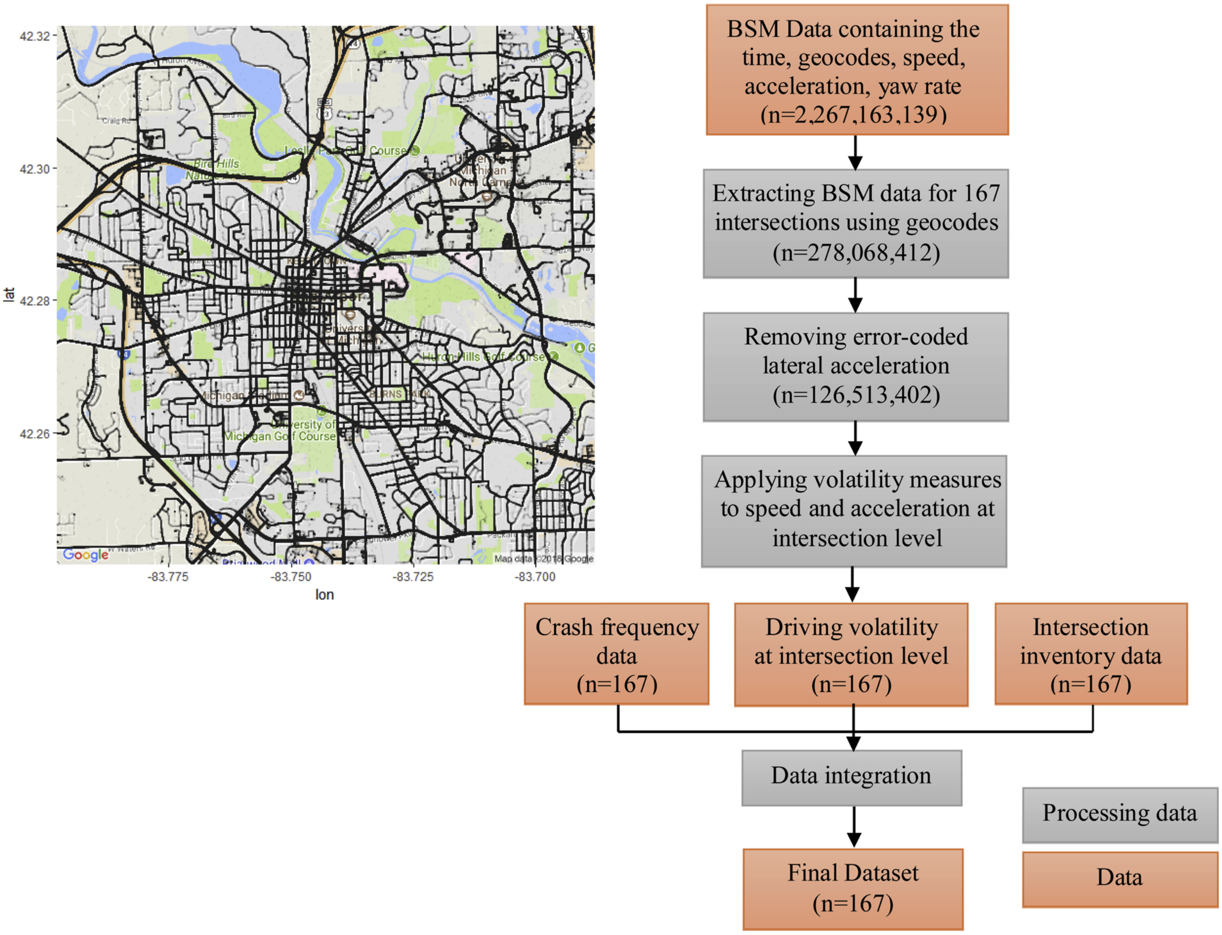
42.32
42.30
42.28
42.26
lat
-83.775
-83.750
-83.725
-83.700
lon
Google
Map data ©2018 Google
BSM Data containing the time, geocodes, speed, acceleration, yaw rate (n=2,267,163,139)
Extracting BSM data for 167 intersections using geocodes (n=278,068,412)
Removing error-coded lateral acceleration (n=126,513,402)
Applying volatility measures to speed and acceleration at intersection level
Crash frequency data (n=167)
Driving volatility at intersection level (n=167)
Intersection inventory data (n=167)
Data integration
Final Dataset (n=167)
Processing data
Data

Fig. 1. Created map from BSM data (left), Data preparation framework (right).

$$MAD = \frac{\sum_{i=1}^{n}\lvert\hat{Y}_i - Y_i\rvert}{N} \tag{20}$$

4 *Mean Squared Error:* assess the estimation accuracy of the model by measuring the distance between the observations and the estimated model. We can write:

$$MSE = \frac{\sum_{i=1}^{n}(\hat{Y}_i - Y_i)^2}{N} \tag{21}$$

where Y_i and $\hat{Y}_i$ are the actual number and estimated number of crashes, and N is the number of intersections. The MAD measure provides the average of misprediction in the method, while the MSE measure is used to assess the error associated in the estimation.

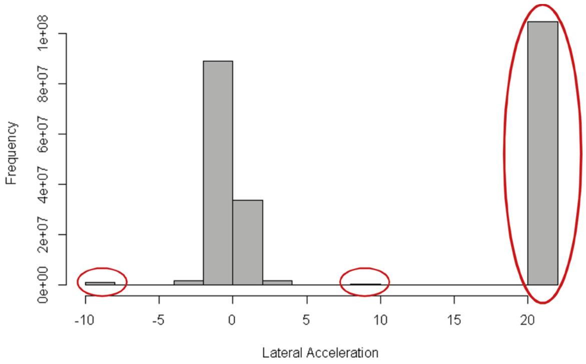

Fig. 2. Histogram of lateral acceleration.

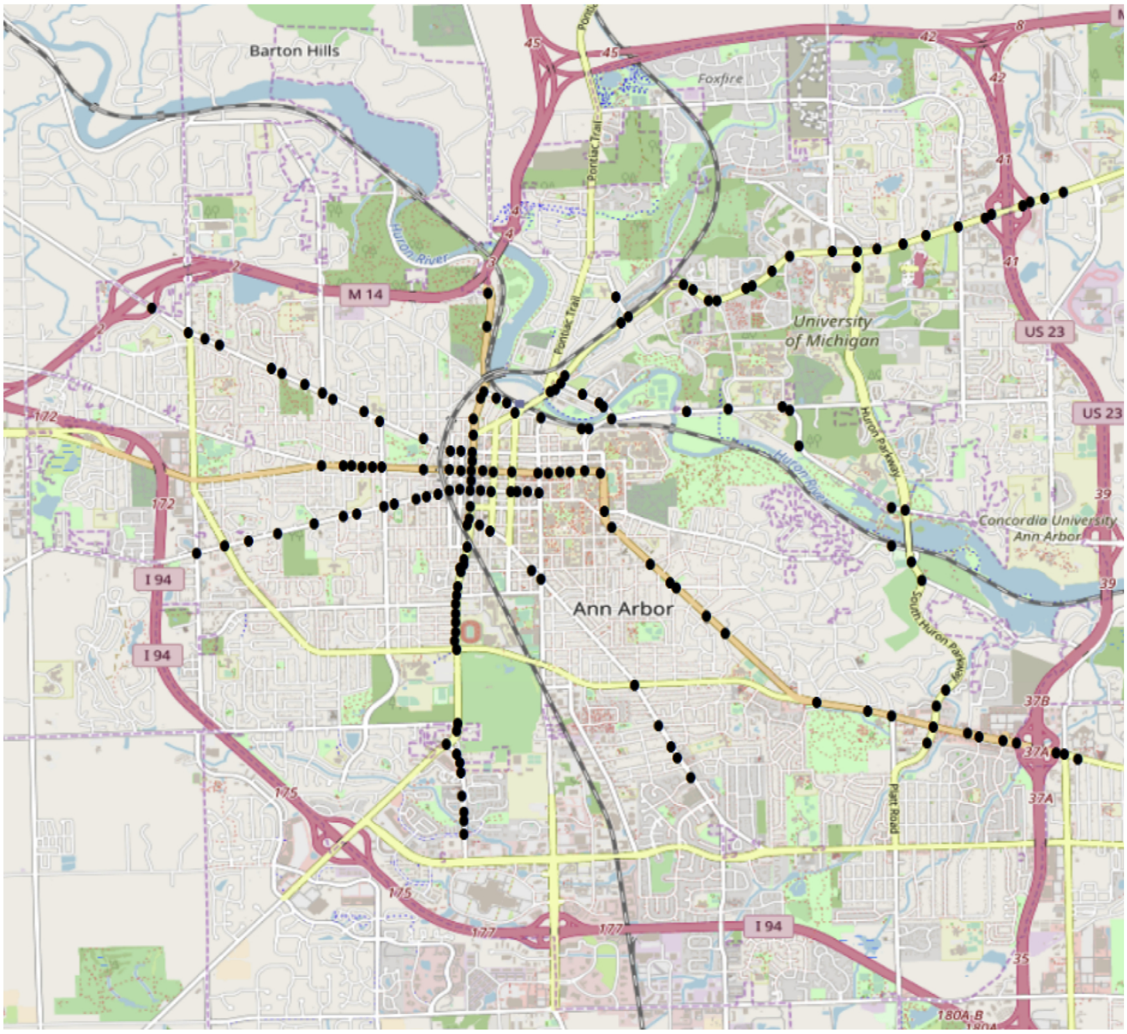

Fig. 3. Location of selected intersections (N = 167).

4. Data

In this study, three data sources were integrated: (1) Basic Safety Messages (BSM) data exchanged by connected vehicles obtained from the SPMD, (2) road inventory data and (3) historical crash data. Fig. 1 (right) shows the data process steps. BSM data were obtained via the Research Data Exchange website (https://www.its.dot.gov/data/). The data provides high-frequency information regarding vehicle location, motion, and driving context factors. Data were collected on October 2012 and April 2013 (N˜225 million observations) using standard protocols by UMTRI at university of Michigan. In this paper, full two-months of publicly available CV data is processed. Due to the error made by developers during data transfer process from DSRC devices to comma separated value (CSV) files, 45.4% of lateral acceleration data are stored as either -9.81, 9.81, and 19.62 m/s^2 which are equivalent to "$-g$", "g" and "$2g$" in the dataset (the histogram of the lateral acceleration is shown in Fig. 2). However, these values belong to 1048 vehicles out of the 2544 vehicles that passed the selected intersections. Therefore, we did not include the erroneous data (shown in Fig. 2 via red eclipses) and the final dataset contains the information of 1496 vehicles passing the intersections.

In order to evaluate the correlation of driving volatility and crashes, we should account for the effect of traffic and geometric characteristics of intersections. Therefore, significant effort was undertaken in order to obtain road inventory data including AADT for major and minor approaches, speed limit, number of lanes in each direction, etc. Data were collected from Google Maps and the Metropolitan Planning Organization Website (http://semcog.org/). Among the intersections in Ann Arbor, 167 intersections were selected (Fig. 3), considering AADT information availability and the availability of BSM data that can calculate 30 measures of driving volatility. To extract the BSM data at intersections, 150-ft. threshold from the center of intersections was established, and by processing 230 million BSMs, CV data for each intersection is extracted and linked to selected 167 intersections. Next, by applying volatility functions to extracted BSMs at aggregate intersection level, speed, longitudinal acceleration, lateral acceleration, and yaw-rate volatility measures are calculated for each intersection.

The historical crash data were obtained from the Metropolitan Planning Organization Website. One of the main challenges in this study is describing 2-month connected vehicle data with historical crash data. In this paper, we are assuming that drivers of CVs that passing the selected intersections are representative of the majority of drivers. The ideal approach is comparing the speed distribution of connected vehicles with the distribution of speed obtained from non-at-fault drivers at study area by conducting quasi-induced method and evaluate whether the difference is acceptable (Chandraratna and Stamatiadis, 2009; Lyles et al., 1991; Stamatiadis and Deacon, 1997). However, the speed of vehicles prior to crash involvement is not available in the crash data. To mitigate this issue, we filtered the historical crash data from October 2012 to 2013 (1-year period) to obtain accurate inference regarding the correlation of intersection volatility and frequency of crashes. It is worth noting that 2-month CV data lies

Table 2
Descriptive Statistics of dependents and key variables (N = 167 intersections).

Variable	Mean	S.D.	Min	Max
Dependent variables				
Rear end crashes	3.51	4.79	0	28
Sideswipe crashes	1.54	2.22	0	13
Angle crashes	1.38	1.77	0	9
Head-on crashes	0.61	1.23	0	6
Intersection related variables				
AADT major road (1000)	18.47	8.60	2.53	45.40
AADT minor road (1000)	8.85	3.87	1.10	27.40
Speed limit of major road (mph)	34.52	6.52	25	45
Speed limit of minor road (mph)	29.28	4.37	15	45
Signalized intersection (yes = 1)	0.49	0.50	0	1
4-legged intersection (yes = 1)	0.47	0.50	0	1
Total through lanes	4.25	1.38	2	8
Total left turn lanes	1.38	1.37	0	6
Total right turn lanes	0.84	0.80	0	4
Intersection-based volatility measures				
Speed Volatility measures				
$Speed - S_{dev}$ (m/s)	10.88	2.57	4.83	16.78
$Speed - C_v$ (%)	44.48	16.00	12.34	80.81
$Speed - Q_{cv}$ (%)	31.67	16.74	6.17	66.74
$Speed - D_{mean}$ (m/s)	7.56	2.07	3.18	12.33
$Speed - 1S_{dev}$ (%)	28.74	13.30	11.35	60.28
$Speed - 2S_{dev}$ (%)	3.63	2.90	0.00	11.31
Longitudinal acceleration volatility measures				
$AccDec_x - S_{dev}$ (m/s^2)	0.76	0.18	0.33	1.42
$Acceleration_x - C_v$ (%)	58.49	5.53	42.53	74.11
$Deceleration_x - C_v$ (%)	65.44	8.54	52.30	120.13
$Acceleration_x - Q_{cv}$ (%)	38.47	5.64	21.84	50.00
$Deceleration_x - Q_{cv}$ (%)	43.20	7.42	22.58	59.62
$AccDec_x - D_{mean}$ (m/s^2)	0.39	0.09	0.15	0.54
$AccDec_x - 1S_{dev}$ (%)	23.44	4.71	6.50	35.87
$AccDec_x - 2S_{dev}$ (%)	6.45	1.83	1.61	11.09
Lateral acceleration volatility measures				
$AccDec_y - S_{dev}$(m/s^2)	1.05	0.36	0.12	2.14
$Acceleration_y - C_v$ (%)	87.34	38.25	28.64	225.62
$Deceleration_y - C_v$ (%)	128.25	34.18	57.45	221.63
$Acceleration_y - Q_{cv}$ (%)	45.80	14.02	10.00	93.01
$Deceleration_y - Q_{cv}$ (%)	57.41	20.39	15.09	92.79
$AccDec_y - D_{mean}$ (m/s^2)	0.77	0.62	0.07	4.65
$AccDec_y - 1\,S_{dev}$ (%)	5.80	6.10	0.0	39.57
$AccDec_y{-}2S_{dev}$ (%)	1.92	2.17	0.0	13.16
Yaw rate volatility measures				
$YawRate - S_{dev}$ (degree/s)	3.64	1.45	0.418	8.88
$YawRate - C_v$ (%)	1.64	0.54	0.22	3.13
$YawRate - C_v$ (%)	1.64	0.51	0.33	3.21
$YawRate - Q_{cv}$ (%)	0.57	0.20	0.08	0.92
$YawRate - Q_{cv}$ (%)	0.55	0.20	0.10	0.92
$YawRate - D_{mean}$ (degree/s)	2.99	1.70	0.18	6.93
$YawRate{-}1S_{dev}$ (%)	0.10	0.08	0.00	0.40
$YawRate{-}2S_{dev}$ (%)	0.04	0.03	0.00	0.13

*S_{dev}: standard deviation; ($1S_{dev}$): % of extreme points beyond mean ± one standard deviation; ($2S_{dev}$): % of extreme points beyond mean ± two standard deviation; C_v: coefficient of variation; Q_{cv}: quartile coefficient of variation; D_{mean}: mean absolute deviation; $Acceleration_x$: longitudinal acceleration; $Deceleration_x$: longitudinal deceleration; $AccDec_x$: both longitudinal acceleration and deceleration; $Accleration_y$: lateral acceleration; $Deceleration_y$: lateral deceleration; $AccDec_y$: both lateral acceleration and deceleration;

between the selected period.

Finally, to ensure the accuracy of the manually collected data, 20% of the data was randomly checked and verified. In addition, the plot of the data in Fig. 1 (left) indicates the high precision of the BSM data.

5. Results

5.1. Descriptive statistics

In this section, the descriptive statistics of dependent variables, calculated intersection-based driving volatilities, and intersection related variables are shown in Table 2. As discussed before, the two-month CV data is used to calculate the volatility measures that attempt to capture the variations of speed, longitudinal/lateral acceleration, and yaw-rate. In order to help conceptualize the distribution of variables, the mean, standard deviation, minimum and maximum of the variables are provided for 167 intersections.

Table 3
Measures of goodness of fit for the fitted model.

	Goodness of fit	Fixed Parameter	Random Parameter	S-GWPR
Rear-End	$R^2_{poisson}$	0.674	0.908	0.754
	AIC	376.99	215.96	338.6
	MAD	2.014	0.885	1.723
	MSE	8.477	1.322	6.001
Sideswipe	$R^2_{poisson}$	0.473	0.791	0.582
	AIC	279.67	190.05	258.05
	MAD	1.144	0.685	1.039
	MSE	2.991	0.843	2.346
Angle	$R^2_{poisson}$	0.391	0.675	0.457
	AIC	247.59	198.37	237.61
	MAD	0.994	0.739	0.938
	MSE	1.851	0.949	1.644
Head-on	$R^2_{poisson}$	0.446	0.584	0.509
	AIC	169.46	152.51	165.68
	MAD	0.539	0.459	0.508
	MSE	0.798	0.538	0.711

5.2. Modeling results

According to the aforementioned methods the Poisson regression, RP Poisson regression, and GWPR models were developed to explain the observed variations in frequency of rear-end, sideswipe, angle and head-on collisions given road inventory and intersection-based driving volatility factors. Although the aim of this study is not to compare different methodological approaches for modeling crash counts, we provide a model comparison in the following sections to illustrate more insights regarding their performance. In the following, the models performance on estimating the frequency of various crash types is compared. Next, the developed models are presented and discussed.

5.2.1. Model comparison

In order to estimate the fixed parameter Poisson regression models, the intersection related factors and driving volatility measures were incorporated.

In order to estimate the RP Poisson regression, 200 Halton draws was applied considering multiple functional form of the coefficients such as normal, lognormal, triangular, and uniform. Similar to previous studies (Anastasopoulos and Mannering, 2009; El-Basyouny and Sayed, 2009; Kamrani et al., 2018b; Xu and Huang, 2015), for the random-held parameters the normal distribution had the best fit to the data, in all the crash type models.

To estimate the S-GWPR, the study considered bi-square and Gaussian fixed and adaptive kernels. In all crash type models, the adaptive bi-square kernel showed the best fit to the data based on their AIC score. In addition, all variables are significantly varied across the space for rear-end crashes, leading to the basic GWPR model. On the other hand, the S-GWPR model performed better for sideswipe, angle and head-on crashes by reducing the model complexity.

As discussed in the methodology section, to compare performance of the models, the $R^2_{poisson}$, AIC, MAD and MSE statistics are quantified. Table 3 shows the results for rear-end, sideswipe, angle, and head-on crashes. Based on the results, the RP Poisson regression outperformed the fixed parameter and GWPR models in all types of crashes. It should be noted that, both the RP and the S-GWPR models improved the fit for the fixed parameter models.

Table 4
Modeling results for rear-end crashes (N = 167 intersections).

Variable	Poisson Regression		Random Parameter							GWPR						
	β.[1]	ME	Mean	ME	Min	1st Q	Med	3rd Q	Max	Mean	Min	1st Q	Med	3rd Q	Max	Test[2]
Constant	-4.135^{***}	-	-4.837^{***}	-						−3.518	−5.126	−3.969	−3.453	−2.864	2.369	Yes
AADT MAJOR (1000)	0.054^{***}	0.19	0.053^{***}	0.1	0.052	0.059	0.062	0.066	0.079	0.063	0.021	0.059	0.071	0.073	0.075	Yes
Std. AADT Major			0.016^{***}	-												
AADT MINOR (1000)	0.057^{***}	0.2	0.06^{***}	0.11						0.049	0.036	0.041	0.046	0.051	0.097	Yes
SIGNALIZED (yes = 1)	0.464^{***}	1.46	0.468^{***}	0.88						0.529	0.02	0.217	0.299	0.495	1.16	Yes
4 legged intersection	0.630^{***}	2.15	0.494^{***}	0.91						0.401	0.143	0.43	0.538	0.596	1.087	Yes
Speed-2S_{dev} $Speed\text{-}2S_{dev}$	0.066^{***}	0.23	0.093^{***}	0.17						0.066	−0.008	0.054	0.059	0.079	0.156	Yes
$Speed\text{-}C_v$	0.009^{***}	0.03	0.015^{***}	0.03	0.0184	0.0186	0.0187	0.0188	0.0201	0.015	−0.012	0.007	0.020	0.022	0.029	Yes
Std. $Speed\text{-}C_v$			0.002^{***}	-												
$Accleration_x\text{-}Q_{cv}$	0.058^{***}	0.2	0.063^{***}	0.1	0.014	0.035	0.046	0.053	0.093	0.036	0.003	0.018	0.029	0.059	0.078	Yes
Std. $Accleration_{x\text{-}}Qcv$			0.011^{***}	-												
Null Deviance	878.86		878.86							878.86						
Model Deviance	360.99		193.96							294.2						
Explained Deviance	0.589		0.779							0.665						
AIC	376.99		215.96							338.6						

[1] Significance at *** 1%, ** 5%, and * 10%.
[2] Non-stationary test.

5.2.2. Model estimation

In order to estimate the fixed parameter models, intersection related variables were used, and the significant ones were kept in the model, then measures of driving volatility were added into the model. For model selection, the AIC, log-likelihood values, and variable significance were used. As discussed in the methodology section, the Lagrange Multiplier test was conducted to test for the over-dispersion existence (Greene, 2003). Based on the results, the LM values for rear-end, sideswipe, angle and head-on crashes were lower than the critical Chi-square value for the 95 percent confidence interval, which is 3.84. Therefore, for all the crash type models the null hypothesis failed to reject, and it is appropriate to use the Poisson regression models (Washington et al., 2010).

After developing the fixed parameter model, significant variables in the models were used to develop RP Poisson and GWPR models. The estimated parameters for the RP Poisson and S-GWPR are presented by the minimum, lower quartile, median, upper quartile, and maximum estimated coefficients. In order to check for the multicollinearity, a common rule of thumb suggests that if the variance of Inflation (VIF) is higher than 5, multicollinearity might be an issue. VIF values for included variables were checked and all of them were below 5. The following sections discuss modeling results for rear-end, sideswipe, angle and head-on collisions.

5.2.2.1. Rear-end crashes. The modeling result for frequency of rear-end crashes in the selected time period is shown in Table 4. As discussed before, it is evident that the RP Poisson model outperformed the fixed Poisson and GWPR. The models suggest that three measures of driving volatility are highly correlated with the number of rear-end crashes at intersections: Coefficients of variation in speed ($Speed\text{-}C_v$), number of speed points lying beyond two standard deviations ($Speed\text{-}2S_{dev}$), and coefficient of variation volatility of positive longitudinal acceleration ($Acceleration_x - Q_{cv}$). The fixed parameter Poisson model states that the associations of driving volatility on rear-end crashes are fixed across the intersections. However, based on the RP Poisson model results, the effects of coefficients of some volatilities significantly vary across intersections with normal distribution. The number of speed points lying beyond two standard deviations ($Speed\text{-}2S_{dev}$), are positively associated with number of rear-end crashes. They indicate that intersections with higher speed volatility are prone to have a higher number of rear-end crashes. Referring to partial effects, it can be observed that a one percent increase in Speed-C_v and Speed-$2S_{dev}$ increase the average number of rear-end crashes for 0.17 and 0.03, respectively. In addition, quartile coefficient of variation volatility of positive longitudinal acceleration ($Acceleration_x - Q_{cv}$) is significant in the model with positive sign reveals that increase in the variation of longitudinal control of the vehicle in terms of acceleration, increases the expected number of rear-end crashes. Controlling for other variables, a one percent increase in $Acceleration_x - Q_{cv}$ increases the rear-end number of crashes, on average, for 0.1. Considering the high variations in these volatilities, they have a substantial impact on the number of crashes. It is worth mentioning that all of the lateral volatilities are tested in the model but none of them was significant. From the model, it can be inferred that intersections with higher longitudinal volatility expected to have a higher number of rear end crashes. Based on intuition, we expect that failure in longitudinal control of the vehicle lead to rear-end crashes which is consistent with the results.

Other factors used in the model as control variables are significant and show the expected sign. According to the Table 4, a one thousand increase in AADT in major and minor streets contributes to a 0.1 and 0.11 increase in the number of rear-end crashes, respectively. Based on the results, on average, signalized intersections have 0.88 more rear-end crashes than un-signalized intersections. Four-legged intersections have 0.91 more crashes than T-intersections.

Referring to the GWPR model, as shown in Table 4, non-stationary test and the results show that there is a non-stationary spatial pattern and significant variation in all of the estimated coefficients across space. According to the results, volatility measures are positively correlated with the number of rear-end crashes in almost all locations. It should be noted that presence of over-dispersion in data could lead to negative coefficient signs at some intersections (Xu et al., 2015). In addition, volatilities with unexpected signs might be insignificant in the model. Focusing on volatility measures, an estimated coefficient for *Speed-* C_v varies from -0.012 to 0.029. Based on the results, 17 intersections have negative values among which none of them are significant at a 95% confidence level. The Estimated coefficients for $Speed\text{-}2S_{dev}$ vary from -0.008 to 0.156, and 11 intersections (6.5%) have negative signs. However, none of them was significant in the model. Along with volatilities, as shown in Table 4, intersection related variables vary across space significantly. Although the coefficients vary from negative values to positive, none of the negative estimates is significant in the

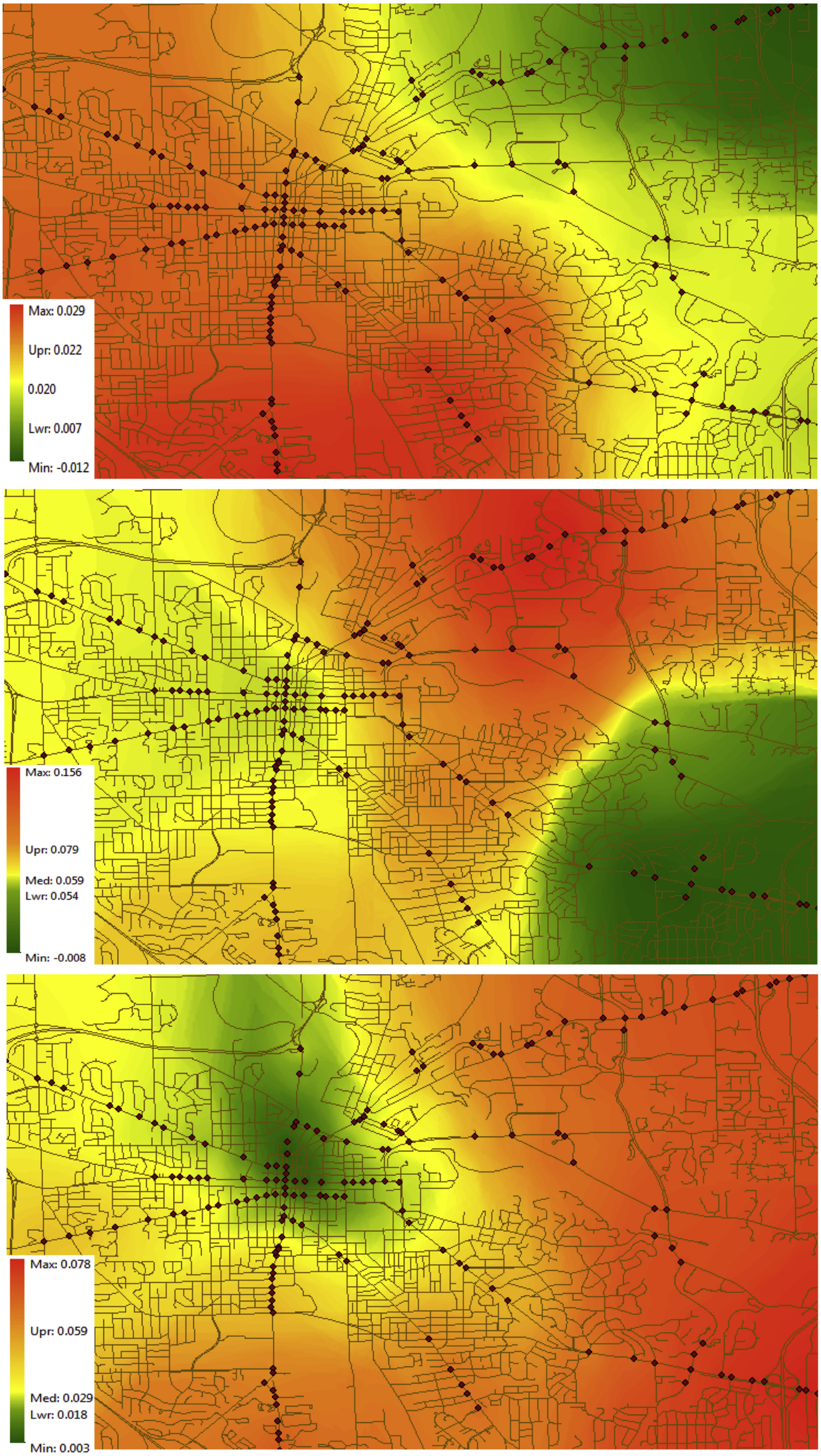

Fig. 4. Local estimation of *Speed-*C_v (top), *Speed-2*S_{dev} (middle), and $Acceleration_x - Q_{cv}$ (bottom) on rear-end crashes.

model. By applying IDW interpolation, the coefficients are mapped in the space and the results of GWPR model for the local estimation of volatility measures are shown in Fig. 4.

5.2.2.2. Sideswipe crashes. In this section, the association of intersection-based volatilities on the frequency of sideswipe crashes is discussed. Table 5 summarizes the modeling results for fixed parameter,

Table 5
Modeling results for sideswipe crashes (N = 167 intersections).

Variable	Poisson Regression		Random Parameter							GWPR						
	β.[1]	ME	Mean	ME	Min	1st Q	Med	3rd Q	Max	Mean	Min	1st Q	Med	3rd Q	Max	Test[2]
Constant	-6.869^{***}	-	-6.737^{***}	-						−4.971	−8.014	−6.459	−4.176	−3.945	−3.758	Yes
AADT MAJOR (1000)	0.002	0.00	0.004	0.00						0.001						
AADT MINOR (1000)	0.023^{*}	0.03	0.011	0.01						0.019	−0.054	−0.007	0.034	0.043	0.049	Yes
SIGNALIZED (yes = 1)	2.337^{***}	3.26	2.317^{***}	1.98						2.067						
$Speed-2S_{dev}$	0.104^{***}	0.16	0.103^{***}	0.1						0.079	0.026	0.055	0.068	0.108	0.146	Yes
$Acceleration_x-C_v$	0.062^{***}	0.09	0.061^{***}	0.06	0.063	0.106	0.109	0.112	0.171	0.05	0.029	0.033	0.037	0.074	0.1	Yes
Std. $Acceleration_x-C_v$			0.01^{***}	-												
$Deceleration_y-C_v$	0.009^{***}	0.01	0.008^{***}	0.01						0.004						
$AccDec_y\text{-}2S_{dev}$	0.114^{***}	0.17	0.11^{***}	0.10	0.051	0.057	0.059	0.063	0.085	0.071	−0.006	0.012	0.022	0.09	0.348	Yes
Std. 2 $AccDec_y\text{-}2S_{dev}$			0.07^{***}	-												
Null Deviance	461.86		461.86							461.86						
Model Deviance	263.67		170.05							228.76						
Explained Deviance	0.429		0.632							0.505						
AIC	279.67		190.05							258.05						

[1] Significance at *** 1%, ** 5%, and * 10%.
[2] Non-stationary test.

random parameter, and S-GWPR models on such crashes. In terms of goodness of fit, by capturing unobserved heterogeneity with RP Poisson and S-GWPR models, the model fits improved significantly. All the models suggest that intersection volatilities in terms of speed, longitudinal and lateral acceleration volatilities are highly associated with frequency of sideswipe crashes. That said, four intersection-based volatility measures are highly contributing to crash frequency: number of speed points lying beyond two standard deviations (*Speed-2S*$_{dev}$), coefficient of variation volatility of positive longitudinal acceleration ($Acceleration_x - C_v$), coefficient of variation of negative lateral acceleration ($Deceleration_y - C_v$), and number of lateral acceleration points lying beyond two standard deviations ($AccDec_y - 2S_{dev}$). However, RP Poisson and S-GWPR suggest that the impacts of intersection-based volatility measures are not fixed across the intersections.

The marginal effect of the RP Poisson model reveals that a one percent increase in *Speed-2S*$_{dev}$ is correlated with 0.16 increase in sideswipe crashes, on average. The model also indicates that the effect of $Acceleration_x - C_v$ is normally distributed across the intersections so that one percent increase in $Acceleration_x - C_v$, on average, is associated with a 0.06 increase in sideswipe crashes. Referring to lateral acceleration volatilities, a one percent increase in $Deceleration_y - C_v$ is correlated to 0.01 increase in the frequency of sideswipe crashes. In addition, the effect of $AccDec_y - 2S_{dev}$ on sideswipe crashes is normally distributed across the selected intersections contributing to 0.1 increase in the frequency of sideswipe crashes, on average, by one percent increase in its magnitude. It is worth noting that in order to control for intersection-related variables, traffic exposure and type of the signal is used in the model, which is summarized in Table 5.

Coming to S-GWPR model, the results suggest that along with $Acceleration_x - C_v$ and $AccDec_y - 2\ S_{dev}$ volatilities, the impact of *Speed-2S*$_{dev}$ is not fixed across the intersections. The distribution of estimated coefficients is shown in Table 5. One might notice that for some intersections, the estimation of $AccDec_y - 2\ S_{dev}$ is negative, while these observations are 5.9 percent of the intersections and they are not statistically significant. The local estimation plots of the volatility measures are shown in Fig. 5. The estimated local coefficients suggest that intersection-based volatilities are an issue in eastern region of the city, comparing to the west side.

5.2.2.3. Angle crashes. Modeling results for angle crashes are summarized in Table 6. Compared to the fixed parameter model, S-GWPR and RP Poisson models fit better, indicating that unobserved heterogeneity is captured. It can be observed that RP Poisson outperformed S-GWPR in terms of AIC value.

The developed models suggest that frequency of angle crashes is associated with four intersection-based volatility measures including speed, longitudinal, and lateral acceleration volatilities. In terms of speed volatility, quartile coefficient of variation in speed (*Speed-* Q_{cv}), is significantly correlated with angle crashes. On average, a one percent increase in *Speed-* Q_{cv} is associated with a 0.02 increase in angle crashes. Along with speed volatility, intersections with higher longitudinal volatilities experienced a higher number of angle crashes. One percent increase in the coefficient of variation of positive longitudinal accelerations ($Acceleration_x - C_v$) is associated with a 0.08 increase in number of angle crashes, on average. Intersections with higher lateral volatility are prone to have higher number of angle crashes. The coefficient of variation of negative lateral acceleration ($Deceleration_y - C_v$), and number of lateral acceleration points lying beyond two standard deviations ($AccDec_y - 2S_{dev}$) are statistically associated with angle crashes. In particular, the model suggests that coefficients of $Deceleration_y - C_v$ are normally disturbed across the intersections. On average, one percent increase in $Deceleration_y - C_v$ and $AccDec_y - 2S_{dev}$ is associated with a 0.01 and 0.11 increase in angle crashes respectively.

The S-GWPR model shows a better fit than fixed parameter model and its results suggest that there is a spatial variation regarding the impact of lateral volatility across the intersections. Fig. 6 depicts the heatmap of estimated coefficients for the lateral volatility ($Deceleration_y - C_v$). The estimated coefficients range from 0.009 to 0.018 and 85 percent of the estimated coefficients are statistically significant.

In terms of intersection-specific variables, the AADT of major road is contributing to frequency of angle crashes, while the AADT of minor approach is not statistically significant in the model. In addition, signalized and four-legged intersections have 0.55 and 0.67 higher angle crashes compared to unsignalized and three-legged intersections, on average.

5.2.2.4. Head-on crashes. Table 7 shows estimated fixed parameter Poisson, RP Poisson, and S-GWPR models for head-on crashes. Comparing goodness of fit, in terms of AIC value, the RP Poisson model outperformed the fixed and GWPR models. In the following, the estimated parameters in the RP Poisson model will be discussed.

Based on the results, four measures of driving volatility are significantly associated with the number of head-on crashes. The coefficient of variation for speed (*Speed-C*$_v$), which represents variations in vehicle speeds, is significant in the model, suggesting that intersections

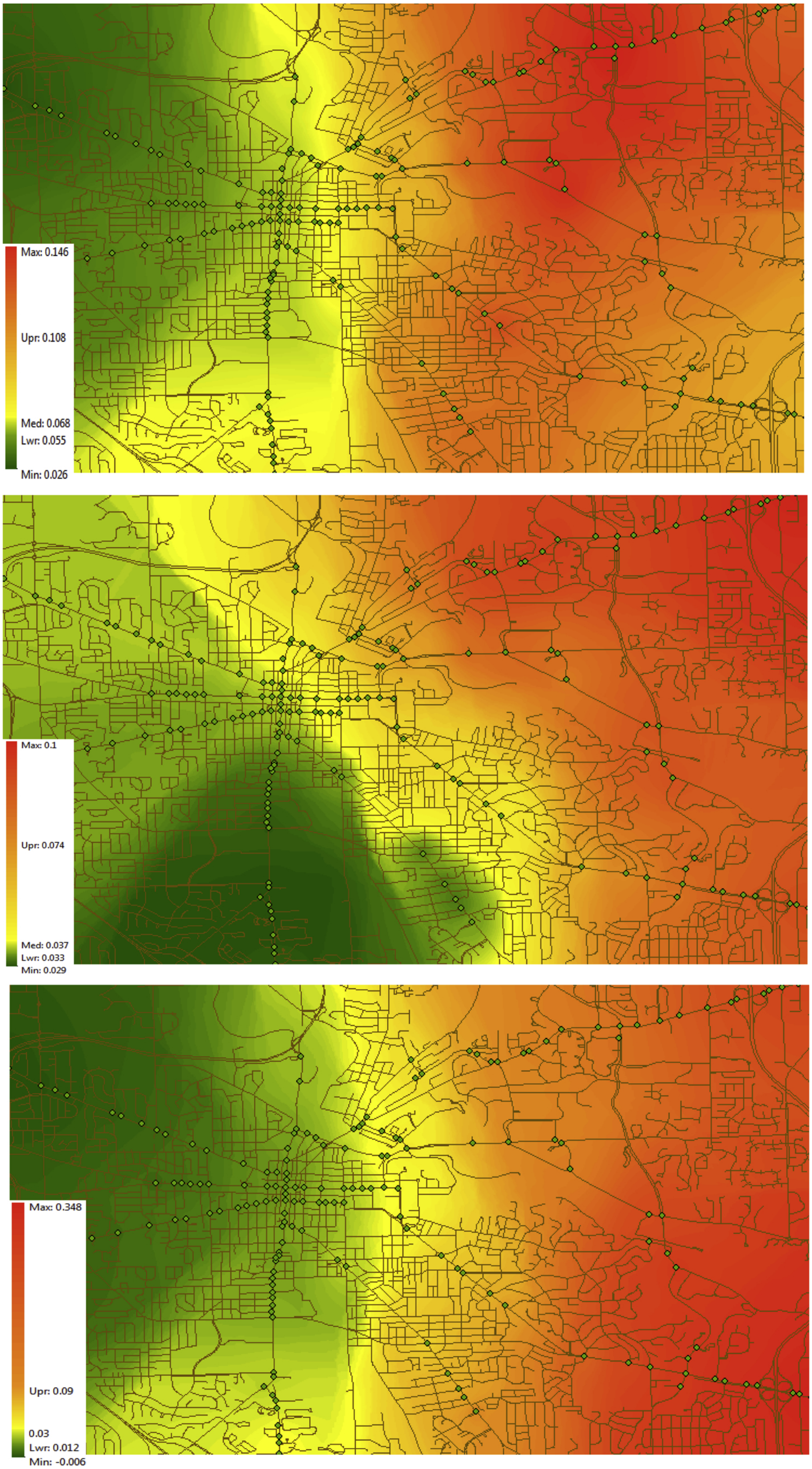

Fig. 5. Local estimation of *Speed-2*S_{dev} *(top)*, *Acceleration*$_x$ − C_v *(middle), and AccDec*$_y$− 2S_{dev} (bottom) on sideswipe crashes.

with higher speed volatility have higher numbers of head-on crashes. A one percent increase in *Speed*-C_v, on average, increases the number of head-on crashes by 0.03. The coefficient of variation for positive longitudinal acceleration (*Acceleration*$_x$ − C_v), which represents variations in longitudinal control of the vehicle, is significant in the model with a positive sign. Based on the model, on average, a one percent

Table 6
Modeling results for angle crashes (N = 167 intersections).

Variable	Poisson Regression		Random Parameter							S-GWPR						
	β.[1]	ME	Mean	ME	Min	1st Q	Med	3rd Q	Max	Mean	Min	1st Q	med	3rd Q	max	Test[2]
Constant	-8.794^{***}	–	-8.988^{***}	–						−8.891						
AADT MAJOR (1000)	0.024^{***}	0.03	0.026^{***}	0.02						0.036						
AADT MINOR (1000)	0.018	0.02	0.018	0.01						0.022						
SIGNALIZED (yes = 1)	0.726^{***}	0.90	0.651^{***}	0.55						0.602						
4 legged intersection	0.791^{***}	1.01	0.803^{***}	0.67						0.783						
Speed-Q_{cv}	0.022^{***}	0.03	0.024^{***}	0.02						0.0177						
$Acceleration_x - C_v$	0.094^{***}	0.13	0.096^{***}	0.08						0.078						
$Deceleration_y - C_v$	0.007^{**}	0.01	0.006^{**}	0.01	0.003	0.005	0.006	0.007	0.012	0.013	0.009	0.011	0.014	0.015	0.018	Yes
$Std.\ Deceleration_y - C_v$			0.003^{***}	–												
$AccDec_y{-}2S_{dev}$	0.133^{***}	0.18	0.129^{***}	0.11						0.178						
Null Deviance	368.50		368.50							368.5						
Model Deviance	229.59		178.37							208.45						
Explained Deviance	0.377		0.516							0.434						
AIC	247.59		198.37							237.61						

[1] Significance at *** 1%, ** 5%, and * 10%.
[2] Non-stationary test.

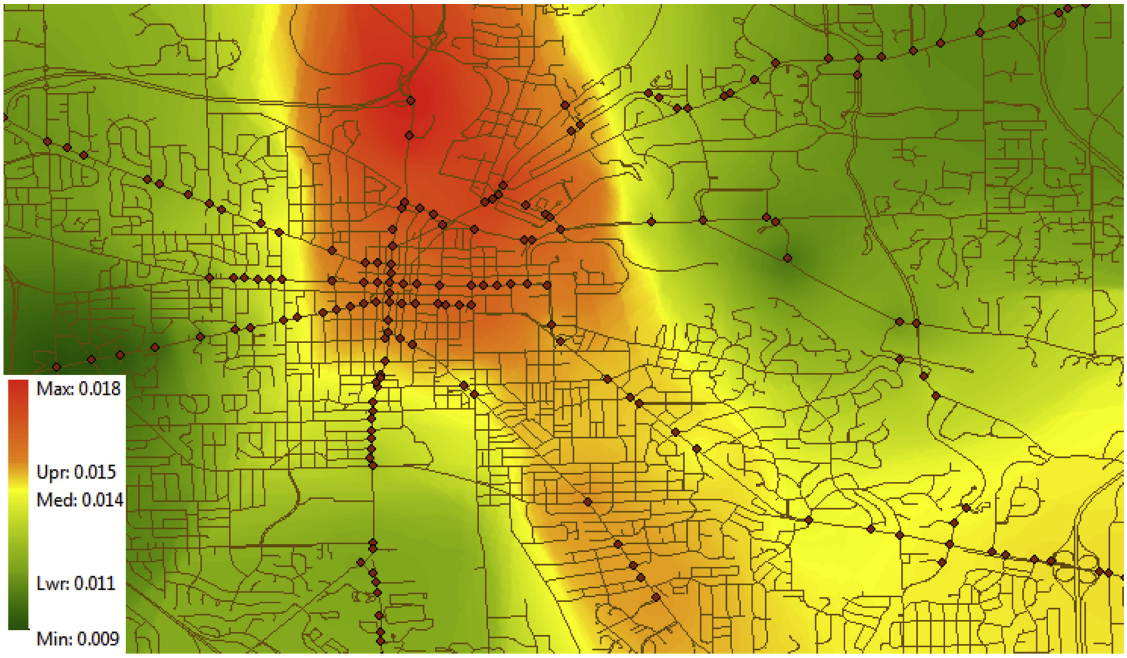

Fig. 6. Local estimation of $Deceleration_y - C_v$ in angle crashes.

increase in $Acceleration_x - C_v$ is associated with an increase in the number of head-on collisions for 0.02. Two volatility measures capturing the variation in lateral movement of the vehicle ($Acceleration_y - Q_{cv}$, and $Deceleration_y - C_v$) are significant with a positive sign. Controlling for other variables, a one percent increase in quartile coefficient of variation of positive lateral acceleration ($Acceleration_y - Q_{cv}$), and coefficient of variation of negative lateral acceleration ($Deceleration_y - C_v$) increases the number of head-on crashes by 0.01 and 0.005.

According to the results, not only variations in longitudinal control of the vehicle (in terms of speed and longitudinal acceleration) are positively significant but also intersections with greater lateral volatility are prone to experience a higher number of head-on crashes. Deviation from the centerline of the road is a major cause of head-on collisions (Gårder, 2006), which is more probable at intersections with greater variations in lateral acceleration. In addition, higher variations in the longitudinal control of a vehicle might lead to deviations from the lane in order to avoid a crash (e.g. rear-end), leading to head-on collisions with vehicles approaching from opposite direction.

Intersection related variables are used in the model as control variables. Based on the results, AADT in major approaches is not significant in the model. However, it was kept in the model as a control variable. Based on the results, a 1000 increase in AADT of minor approach increase the frequency of head-on crashes for 0.01. Controlling for other variables, signalized intersections have 0.22 more head-on crashes compared to un-signalized intersections. In terms of intersection geometry, four-legged intersections on average have 0.17 more head-on crashes than T-intersections.

Referring to S-GWPR model, by considering the spatial variation of the coefficients, the model improved the AIC and explained deviance compared to the fixed parameter Poisson model. As shown in Table 7, non-stationary test was conducted on all variables and those that failed to pass the test are considered a global variable in the model. In the final model, the signalized intersection and measures of lateral

Table 7
Modeling results for head-on crashes (N = 167 intersections).

Variable	Poisson Regression		Random Parameter							S-GWPR						
	β.[1]	ME	Mean	ME	Min	1st Q	Med	3rd Q	Max	Mean	Min	1st Q	med	3rd Q	max	Test[2]
Constant	-13.772^{***}	–	-13.100^{***}	–						−13.182	−14.852	−13.961	−13.127	−12.653	−11.773	No
AADT MAJOR (1000)	−0.006	0.00	−0.005	0.00												
AADT MINOR (1000)	0.042^{*}	0.03	0.047	0.01												
SIGNALIZED (yes = 1)	0.892^{**}	0.42	0.952^{**}	0.22						1.079	0.162	0.622	1.227	1.461	1.820	Yes
4 legged intersection	0.694^{***}	0.39	0.707^{***}	0.17												
Speed-Q_{cv}	0.072^{***}	0.05	0.065^{***}	0.03	0.061	0.064	0.065	0.066	0.076							
Std. Speed-Q_{cv}			0.007^{***}	–												
$Acceleration_x - C_v$	0.1^{***}	0.06	0.105^{**}	0.02												
$Acceleration_y - Q_{cv}$	0.023^{**}	0.02	0.018^{*}	0.01						0.024	0.002	0.015	0.02	0.033	0.056	Yes
$Deceleration_y - C_v$	0.009^{**}	0.01	0.008^{*}	0.005						0.008	0.001	0.005	0.008	0.012	0.016	Yes
Null Deviance	279.70		279.70							279.70						
Model Deviance	151.46		134.51							137.45						
Explained Deviance	0.458		0.519							0.508						
AIC	169.46		152.51							165.68						

[1] Significance at *** 1%, ** 5%, and * 10%.
[2] Non-stationary test.

acceleration volatilities ($Acceleration_y - Q_{cv}$, and $Deceleration_y - C_v$) passed the non-stationary test and significantly vary across the space. By applying IDW interpolation, we mapped estimated lateral acceleration volatilities. Fig. 7 displays the results. Focusing on measures of driving volatility, measures positively contribute to the number of head-on crashes in all areas. Results revealed that in downtown areas, the estimated coefficient of driving volatility measures have a lower correlation with the number of head-on crashes. In the east side of the city, lateral acceleration volatilities have a greater contribution in crashes.

6. Limitations and future work

Because of the error in decoding the CV data from DSRC to csv, around 45 percent of the lateral volatility of trips were voided. However, these errors came from specific device IDs, which were removed from the dataset during the cleaning process. In addition, during the intersection selection procedure, there might be a sample selection issue due to the unavailability of AADT and speed limit information for minor roads in the city. Furthermore, drivers in the study might not represent the population. Also, vehicles whose data was used to obtain driving volatilities might not be representative of the ones who were involved in crashes at intersections. Finally, although the data was error-checked, some errors might still remain from the data collection process.

This study investigates the association of longitudinal and lateral volatilities on the frequency of rear-end and head-on crashes at intersections. The future study would explore the impact of volatility on other crash types, such as sideswipe, angle, and single-vehicle crashes extending the model to multivariate random parameter and geographically weighted Poisson regression models. Furthermore, future studies should investigate contributing factors such as geometric design, traffic conditions, signal timing, etc., that might increase driving volatility at intersections.

While in the literature, there are multiple surrogate safety measures such as time-to-collision (TTC), exposed time-to-collision, time integrated time-to-collision (TIT), and rear-end crash risk index (RCRI) aiming to quantify the crash risk (Essa and Sayed, 2018; Rahman and Abdel-Aty, 2018; Rahman et al., 2018), calculation of such measures need relative distance and kinematic information of front vehicle, which was not available in the data. In future, with higher penetration rate of CVs and availability of data, this information can be integrated in the model.

7. Conclusion

This study evaluated the impact of variations in longitudinal and lateral vehicular control on the frequencies of rear-end, sideswipe, angle and head-on crashes at intersections using the driving volatility which quantifies the degree of variations in instantaneous driving behavior. The goal of this study is to develop a fundamental framework to conceptualize and quantify variations in longitudinal and lateral control of vehicles (using speed, longitudinal/lateral acceleration, and yaw-rate volatilities), and explore the association of volatilities with type of crash.

To reach these goals, the Basic Safety Messages (BSMs) data exchanged by connected vehicles in real-world environments obtained from the Safety Pilot Model Deployment (SPMD) study conducted by the US Department of Transportation in Ann Arbor, MI is used. Such a big and precise dataset is available, which could be incorporated with historical crash data in order to understand the safety performance of the system prior to crash occurrences. This study creates a unique dataset by integrating BSM data, historical crash, and road inventory data. More than 2,225,000,000 BSMs obtained from two months of experiments in Michigan is processed and observations (n ˜ 125,000,000) from 167 intersections are extracted. In order to capture the variations

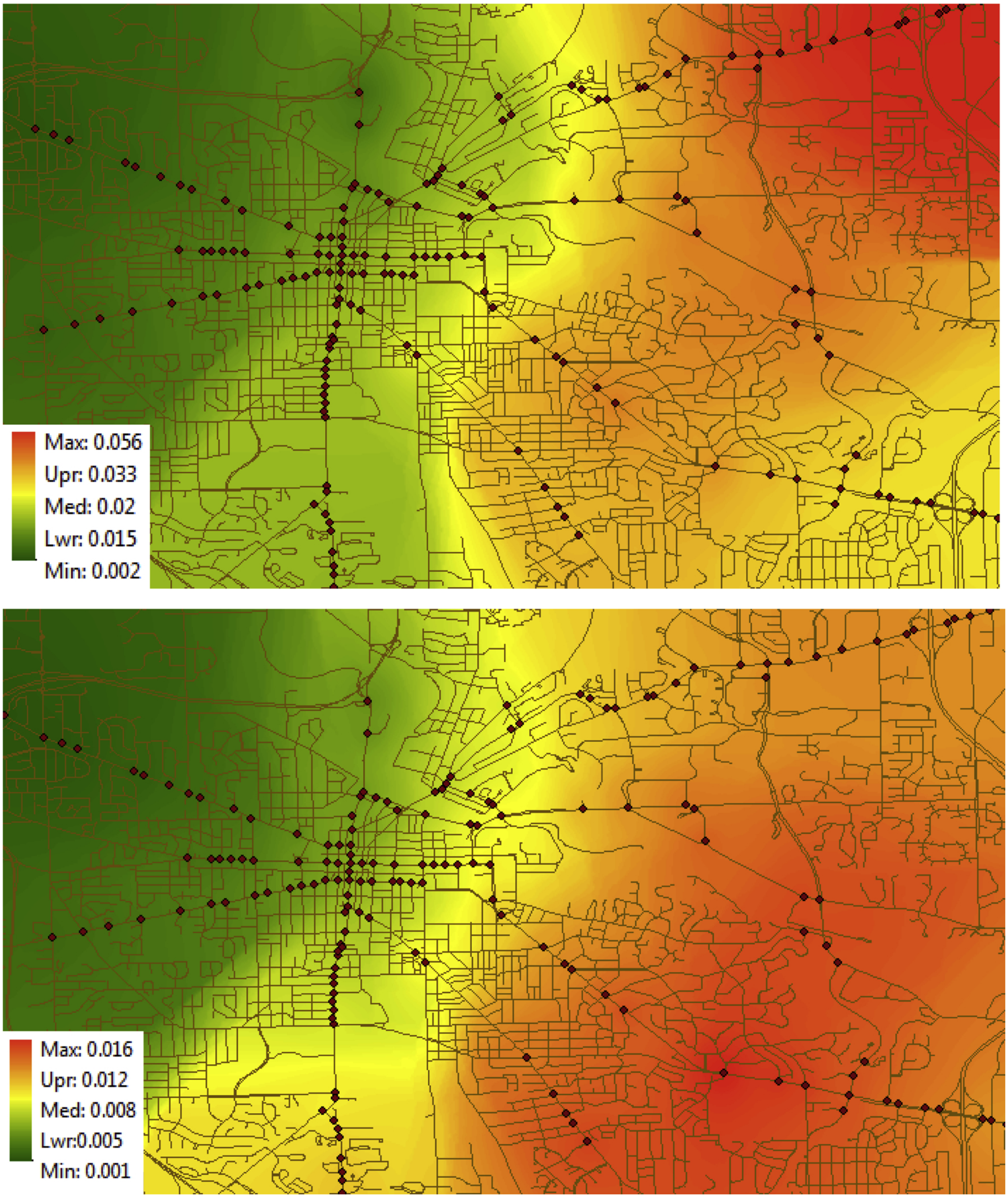

Fig. 7. Local estimation of $Acceleration_y - Q_{cv}$ (top), and $Deceleration_y - C_v$ in head-on crashes.

in vehicle control, 30 measures of driving volatility at the intersection level are calculated using speed, longitudinal/lateral acceleration and yaw-rate. Crash data from October 2012 to 2013 is linked with road inventory data including AADT of major and minor approaches, speed limits, and number of lanes, integrated with BSM data. Significant efforts were made to clean, process, and link the datasets.

From a methodological standpoint, rigorous modeling techniques including fixed parameter, random parameter (RP), and semi-parametric geographically weighted Poisson regression (S-GWPR) are developed to explore the impact of the measured volatilities on the frequency of several crash types. RP Poisson and S-GWPR allows us to consider the unobserved heterogeneity in the data coming from multiple unobserved factors. It is worth noting that RP Poisson model outperformed S-GWPR and Fixed Poisson models in all of the crash type models.

Referring to rear-end crashes, the RP Poisson model fitted better to the data compared to the fixed parameter and GWPR. Based on the random parameter and GWPR results, variations in longitudinal control of the vehicle in terms of longitudinal acceleration and speed are highly correlated with the number of rear-end crashes, and the estimated coefficients significantly vary across intersections. None of the lateral volatilities is significant in the model.

Focusing on sideswipe and angle crashes, modeling results suggest that along with longitudinal volatilities, in terms of longitudinal acceleration and speed, lateral volatility is highly associated with the frequency of frequency of such crashes. The results indicate that there is a substantial variation among the estimated coefficients for volatility indices at intersections level.

When it comes to head-on crashes, both longitudinal and lateral volatilities are positively associated with the number of crashes. Based on the results, variations in speed and longitudinal and lateral acceleration are statistically significant and increase the frequency of head-on crashes. Deviation from the centerline of the road is the main reason of head-on crashes, and vehicles passing the intersections with higher lateral volatility are prone to deviate from their lane leading to head-on collision. In the S-GWPR model, contributions of lateral acceleration volatility vary across space. In downtown areas, it has a lower contribution while in the east side of the city the association is higher.

Given the calculated measures of volatilities, researchers can proactively identify hotspot intersections where crashes are waiting to happen. These hotspots are intersections where the frequency of crashes is low while the driving volatility is high (Kamrani et al., 2017). We can identify at-risk intersections where the driving behavior differs compared to other intersections by evaluating the driving volatility measures. In order to treat the intersection proactively, further examinations are needed to identify the contributing factors that increase the volatility of intersections such as inappropriate geometric designs, traffic conflicts, limited sight distances, inappropriate signal timing, etc. In addition, utilizing V2I communication, proactive warnings could be generated and transmitted by RSUs at these locations that inform drivers about potential hazards. This information could potentially enhance drivers' situational awareness, leading to a decrease in their driving volatility (Arvin et al., 2018).

Acknowledgments

This paper is supported by National Science Foundation (Award number: 1538139). Additional support is provided by the US Department of Transportation through the Collaborative Sciences Center for Road Safety, a consortium led by The University of North Carolina at Chapel Hill in partnership with The University of Tennessee. Any opinions, findings, and conclusions expressed in this paper are those of the authors and do not necessarily reflect views of the sponsors.

References

Abdel-Aty, M.A., Radwan, A.E., 2000. Modeling traffic accident occurrence and involvement. Accid. Anal. Prev. 32 (5), 633–642.

Agbelie, B.R., Roshandeh, A.M., 2015. Impacts of signal-related characteristics on crash frequency at urban signalized intersections. J. Transp. Saf. Secur. 7 (3), 199–207.

Ahmed, M.M., Ghasemzadeh, A., 2018. The impacts of heavy rain on speed and headway behaviors: an investigation using the SHRP2 naturalistic driving study data. Transp. Res. Part C Emerg. Technol. 91, 371–384.

American Automobile Association. (2009). Aggressive driving: Research update. *American Automobile Association Foundation for Traffic Safety*, 202–638.

Anastasopoulos, P.C., Mannering, F.L., 2009. A note on modeling vehicle accident frequencies with random-parameters count models. Accid. Anal. Prev. 41 (1), 153–159.

Arvin, R., Kamrani, M., Khattak, A.J., Rios-Torres, J., 2018. Safety impacts of automated vehicles in mixed traffic. Paper Presented at the Transportation Research Board 97th Annual Meeting, Washington DC.

Arvin, R., Kamrani, M., Khattak, A.J., 2019a. Examining the role of speed and driving stability on crash severity using shrp2 naturalistic driving study data. Paper Presented at the Transportation Research Board 98th Annual Meeting, Washington DC.

Arvin, R., Khattak, A.J., Rios-Torres, J., 2019b. Evaluating safety with automated vehicles at signalized intersections: application of adaptive cruise control in mixed traffic. Paper Presented at the Transportation Research Board 98th Annual Meeting, Washington DC.

Azizi, L., Sheikholeslami, A., 2012. Safety effect of U-Turn conversions in Tehran: empirical Bayes observational before-and-After study and crash prediction models. J. Transp. Eng. 139 (1), 101–108.

Bartier, P.M., Keller, C.P., 1996. Multivariate interpolation to incorporate thematic surface data using inverse distance weighting (IDW). Comput. Geosci. 22 (7), 795–799.

Bhat, C.R., 2003. Simulation estimation of mixed discrete choice models using randomized and scrambled Halton sequences. Transp. Res. Part B Methodol. 37 (9), 837–855.

Bozdogan, H., 1987. Model selection and Akaike's information criterion (AIC): the general theory and its analytical extensions. Psychometrika 52 (3), 345–370.

Cameron, A.C., Windmeijer, F.A., 1996. R-squared measures for count data regression models with applications to health-care utilization. J. Bus. Econ. Stat. 14 (2), 209–220.

Castro, M., Paleti, R., Bhat, C.R., 2012. A latent variable representation of count data models to accommodate spatial and temporal dependence: application to predicting crash frequency at intersections. Transp. Res. Part B Methodol. 46 (1), 253–272.

Chandraratna, S., Stamatiadis, N., 2009. Quasi-induced exposure method: evaluation of not-at-fault assumption. Accid. Anal. Prev. 41 (2), 308–313.

El-Basyouny, K., Sayed, T., 2009. Accident prediction models with random corridor parameters. Accid. Anal. Prev. 41 (5), 1118–1123.

Essa, M., Sayed, T., 2018. Traffic conflict models to evaluate the safety of signalized intersections at the cycle level. Transp. Res. Part C Emerg. Technol. 89, 289–302.

Fotheringham, A.S., Brunsdon, C., Charlton, M., 2003. Geographically Weighted Regression: the Analysis of Spatially Varying Relationships. John Wiley & Sons.

Gårder, P., 2006. Segment characteristics and severity of head-on crashes on two-lane rural highways in Maine. Accid. Anal. Prev. 38 (4), 652–661.

Ghasemzadeh, A., Ahmed, M.M., 2017. Drivers' lane-keeping ability in heavy rain: preliminary investigation using SHRP 2 naturalistic driving study data. Transp. Res. Rec. (2663), 99–108.

Ghasemzadeh, A., Ahmed, M.M., 2018a. Exploring Factors Contributing to Injury Severity at Work Zones Considering Adverse Weather Conditions. IATSS Research.

Ghasemzadeh, A., Ahmed, M.M., 2018b. Utilizing naturalistic driving data for in-depth analysis of driver lane-keeping behavior in rain: non-parametric MARS and parametric logistic regression modeling approaches. Transp. Res. Part C Emerg. Technol. 90, 379–392.

Ghiasi, A., Hussain, O., Qian, Z.S., Li, X., 2017. A mixed traffic capacity analysis and lane management model for connected automated vehicles: a Markov chain method. Transp. Res. Part B Methodol. 106, 266–292.

Greene, W.H., 2003. Econometric Analysis. Pearson Education, India.

Hadayeghi, A., Shalaby, A., Persaud, B., 2010. Development of planning-level transportation safety models using full Bayesian semiparametric additive techniques. J. Transp. Saf. Secur. 2 (1), 45–68.

Henclewood, D., Abramovich, M., Yelchuru, B., 2014. Safety Pilot Model Deployment-one Day Sample Data Environment Data Handbook. USDOT Research and Technology Innovation Administrations, pp. 1.

Huang, H., Abdel-Aty, M., Darwiche, A., 2010. County-level crash risk analysis in Florida: bayesian spatial modeling. Transp. Res. Rec. (2148), 27–37.

Hurvich, C.M., Simonoff, J.S., Tsai, C.L., 1998. Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion. J. R. Stat. Soc. Series B Stat. Methodol. 60 (2), 271–293.

Jamali, A., Wang, Y., 2017. Estimating pedestrian exposure for small urban and rural areas. Transp. Res. Rec. (2661), 84–94.

Kamrani, M., Wali, B., Khattak, A.J., 2017. Can data generated by connected vehicles enhance safety? Proactive approach to intersection safety management. Transp. Res. Rec. (2659), 80–90. https://doi.org/10.3141/2659-09.

Kamrani, M., Arvin, R., Khattak, A.J., 2018a. Analyzing highly volatile driving trips taken by alternative fuel vehicles. Paper Presented at the Transportation Research Board 97th Annual Meeting Washington DC, United States.

Kamrani, M., Arvin, R., Khattak, A.J., 2018b. Extracting useful information from Basic Safety Message Data: an empirical study of driving volatility measures and crash frequency at intersections. Transp. Res. Rec., 0361198118773869.

Kamrani, M., Khattak, A.J., Li, T., 2018c. A framework to process and analyze driver, vehicle and Road infrastructure volatilities in Real-time. Paper Presented at the Transportation Research Board 97th Annual Meeting.

Kamrani, M., Arvin, R., Khattak, A.J., 2019. The role of aggressive driving and speeding in road safety: insights from shrp2 naturalistic driving study data. Paper Presented at the Transportation Research Board 98th Annual Meeting, Washington DC.

Lee, C., Abdel-Aty, M., 2005. Comprehensive analysis of vehicle–pedestrian crashes at intersections in Florida. Accid. Anal. Prev. 37 (4), 775–786.

Liu, J., Khattak, A.J., Wali, B., 2017. Do safety performance functions used for predicting crash frequency vary across space? Applying geographically weighted regressions to account for spatial heterogeneity. Accid. Anal. Prev. 109, 132–142.

Loader, C., 2006. Local Regression and Likelihood. Springer Science & Business Media.

Lord, D., Mannering, F., 2010. The statistical analysis of crash-frequency data: a review and assessment of methodological alternatives. Transp. Res. Part A Policy Pract. 44 (5), 291–305.

Lyles, R.W., Stamatiadis, P., Lighthizer, D.R., 1991. Quasi-induced exposure revisited. Accid. Anal. Prev. 23 (4), 275–285.

Nakaya, T., Fotheringham, A.S., Brunsdon, C., Charlton, M., 2005. Geographically weighted Poisson regression for disease association mapping. Stat. Med. 24 (17), 2695–2717.

Nakaya, T., Charlton, M., Lewis, P., Fortheringham, S., Brunsdon, C., 2012. Windows Application for Geographically Weighted Regression Modeling. Ritsumeikan University, Kyoto, Japan.

National Highway Traffic Safety Administration, 2001. A compilation of motor vehicle crash data from the fatality analysis reporting system and the general estimates system. Traffic Safety Facts.

National Highway Traffic Safety Administration. (2008). National motor vehicle crash causation survey: Report to congress. *National Highway Traffic Safety Administration Technical Report DOT HS*, *811*, 059.

Nezafat, R.V., Beheshtitabar, E., Cetin, M., Williams, E., List, G.F., 2018. Modeling and evaluating traffic flow at sag curves when imposing variable speed limits on connected vehicles. Transp. Res. Rec., 0361198118784169.

Nightingale, E., Parvin, N., Seiberlich, C., Savolainen, P.T., Pawlovich, M., 2017. Investigation of skew angle and other factors influencing crash frequency at high-speed rural intersections. Transp. Res. Rec. (2636), 9–14.

Noland, R.B., Quddus, M.A., 2005. Congestion and safety: a spatial analysis of London. Transp. Res. Part A Policy Pract. 39 (7-9), 737–754.

Paleti, R., Eluru, N., Bhat, C.R., 2010. Examining the influence of aggressive driving behavior on driver injury severity in traffic crashes. Accid. Anal. Prev. 42 (6), 1839–1854.

Rahman, M.S., Abdel-Aty, M., 2018. Longitudinal safety evaluation of connected vehicles' platooning on expressways. Accid. Anal. Prev. 117, 381–391.

Rahman, M.S., Abdel-Aty, M., Wang, L., Lee, J., 2018. Understanding the highway safety benefits of different approaches of connected vehicles in reduced visibility conditions. Transp. Res. Rec., 0361198118776113.

Shou, Z., Di, X., 2018. Similarity analysis of frequent sequential activity pattern mining. Transp. Res. Part C Emerg. Technol. 96, 122–143.

Stamatiadis, N., Deacon, J.A., 1997. Quasi-induced exposure: methodology and insight. Accid. Anal. Prev. 29 (1), 37–52.

Stipancic, J., Miranda-Moreno, L., Saunier, N., 2017. Impact of congestion and traffic flow on crash frequency and severity: application of smartphone-collected GPS travel data. Transp. Res. Rec. (2659), 43–54.

Train, K., 2000. Halton Sequences for Mixed Logit. Department of Economics, UCB.

Wali, B., Khattak, A.J., Bozdogan, H., Kamrani, M., 2018a. How is driving volatility related to intersection safety? A Bayesian heterogeneity-based analysis of instrumented vehicles data. Transp. Res. Part C Emerg. Technol. 92, 504–524.

Wali, B., Khattak, A.J., Khattak, A.J., 2018b. A heterogeneity based case-control analysis of motorcyclist's injury crashes: evidence from motorcycle crash causation study. Accid. Anal. Prev. 119, 202–214.

Wali, B., Khattak, A.J., Waters, J., Chimba, D., Li, X., 2018c. Development of safety performance functions: incorporating unobserved heterogeneity and functional form analysis. Transp. Res. Rec., 0361198118767409.

Wang, X., Abdel-Aty, M., 2006. Temporal and spatial analyses of rear-end crashes at signalized intersections. Accid. Anal. Prev. 38 (6), 1137–1150.

Wang, X., Abdel-Aty, M., Almonte, A., Darwiche, A., 2009. Incorporating traffic operation measures in safety analysis at signalized intersections. Transp. Res. Rec. (2103), 98–107.

Wang, X., Khattak, A.J., Liu, J., Masghati-Amoli, G., Son, S., 2015. What is the level of volatility in instantaneous driving decisions? Transp. Res. Part C Emerg. Technol. 58, 413–427.

Washington, S.P., Karlaftis, M.G., Mannering, F., 2010. Statistical and Econometric Methods for Transportation Data Analysis. CRC press.

Wu, Z., Sharma, A., Mannering, F.L., Wang, S., 2013. Safety impacts of signal-warning flashers and speed control at high-speed signalized intersections. Accid. Anal. Prev. 54, 90–98.

Xu, P., Huang, H., 2015. Modeling crash spatial heterogeneity: random parameter versus geographically weighting. Accid. Anal. Prev. 75, 16–25.