Accident Analysis and Prevention 146 (2020) 105735

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

Developing Bicycle-Vehicle Crash-Specific Safety Performance Functions in Alabama Using Different Techniques

Niloufar Shirani-bidabadi[a], Naveen Mallipaddi[a], Kirolos Haleem[b,*], Michael Anderson[a]

[a] Department of Civil and Environmental Engineering University of Alabama in Huntsville, Huntsville, AL, 35899
[b] School of Engineering & Applied Sciences Western Kentucky University 1906 College Heights Blvd, EBS 2122, Bowling Green, KY 42101

ARTICLE INFO

Keywords:
Bicycle-Vehicle Crashes
Multivariate Adaptive Regression Splines
MARS
Safety Performance Functions
Conway-Maxwell-Poisson
COM-Poisson

ABSTRACT

This study develops bicycle-vehicle safety performance functions (SPFs) for five facilities in the Highway Safety Manual (HSM). These are urban two-lane undivided segments (U2U), urban four-lane divided/undivided segments (U4DU), rural two-lane undivided segments (R2U), urban four-leg and three-leg signalized intersections (USG), and urban four-leg and three-leg stop-controlled intersections (UST). Two modeling techniques were explored, the Conway-Maxwell-Poisson (COM-Poisson) model (to accommodate bicycle-vehicle crash under-dispersion) and a machine learning technique, the multivariate adaptive regression splines (MARS). MARS is a non-black-box model and can effectively handle non-linear crash predictors and interactions. A total of 1,311 bicycle-vehicle crashes from 2011 through 2015 in Alabama were collected and their respective police reports were reviewed in details. Results from the SPFs for roadway segments using COM-Poisson showed that bicycle-vehicle crash frequencies were reduced along curved and downgrade/upgrade stretches and when having heavy traffic flow (along U2U segments). For urban signalized (USG) intersections, the absence of right-turn lanes on minor roads, the presence of bus stops, and the increase in the major road annual average daily traffic (AADT) were significant factors contributing to the increase in the number of bicycle-vehicle crashes. However, the presence of divided medians on major approaches was found to reduce bicycle-vehicle crashes at USG and UST intersections. MARS outperformed the corresponding COM-Poisson models for all five facilities based on mean absolute deviance (MAD), mean square prediction error (MSPE), and generalized R-square. MARS is recommended as a promising technique for effectively predicting bicycle-vehicle crashes on segments and intersections.

1. Introduction

Cycling is one of the rapidly-growing commuting modes in the U.S., with an increase in work commutes from 488,000 in 2000 to 786,000 in 2008-2012, which is a significant increase compared to any other commuting mode (American Community Survey Reports, 2018). However, bicyclists are at a greater risk of injuries and deaths when involved in a crash with motor vehicles National Highway Traffic Safety Administration 2019. In 2015, the number of bicycle fatalities and injuries in the U.S. were higher when compared to the previous ten years, with over 1,000 bicycle fatalities and 467,000 bicycle-related injuries (Centre for Disease Control and Prevention 2019). In the State of Alabama, 1,311 bicycle-vehicle crashes were reported from 2011 to 2015. During such years, 78% increase in bicycle fatal crashes per 10 million population was observed. This shows that bicyclist safety is of concern in Alabama.

This study develops safety performance functions (SPFs) for bicycle-vehicle crashes at intersections and segments (midblocks) in urban and rural areas of Alabama. The Highway Safety Manual (HSM) defines SPFs as regression models (or crash prediction models) for estimating the predicted crash frequency on roadway segments or intersections (American Association of State Highways and Transportation Officials 2010). Bicycle crash count data in Alabama exhibit under-dispersion (i.e., variance less than the mean). Although traditional Poisson and negative binomial (NB) have been widely used for crash data modeling, they have limitations to model under-dispersed data. One promising technique that can handle under-dispersed data is the Conway-Maxwell-Poisson (COM-Poisson) that is introduced by Conway and Maxwell (1962).

This study has two main objectives. The first one is to develop

* Corresponding author.
E-mail addresses: nb0040@uah.edu (N. Shirani-bidabadi), vm0027@uah.edu (N. Mallipaddi), kirolos60@hotmail.com (K. Haleem), andersmd@uah.edu (M. Anderson).

https://doi.org/10.1016/j.aap.2020.105735
Received 25 May 2020; Received in revised form 5 August 2020; Accepted 10 August 2020
Available online 21 August 2020
0001-4575/

bicycle-vehicle SPFs (or bicycle-vehicle crash prediction models) for segment and intersection facilities in the HSM to identify those significant factors affecting bicyclist safety at these two locations. The study develops SPFs for various facilities in the HSM using the COM-Poisson model. These include urban two-lane undivided segments (U2U), urban four-lane divided/undivided segments (U4DU), rural two-lane undivided segments (R2U), urban four-leg and three-leg signalized intersections (USG), and urban four-leg and three-leg unsignalized (or stop-controlled) intersections (UST). The second objective is to improve the prediction of bicycle-vehicle crashes to help state transportation agencies propose the appropriate countermeasures to improve bicyclist safety. For this, the study compares the COM-Poisson models' predictions to a promising machine learning technique, the multivariate adaptive regression splines (MARS), that has been recently used to enhance crash predictions (Abdel-Aty and Haleem, 2011; Gaweesh et al., 2019; Yue et al., 2019). MARS is a non-parametric and non-black-box regression model that accounts for nonlinearities and interactions in the predictor variables.

Previous studies on bicycle crash SPF development did not consider rural segment facilities (see for example, Gates et al., 2016; Mukoko and Pulugurtha, 2020). Recognizing this gap, the contributions of this study to the field of traffic safety are as follows: (1) developing bicycle crash SPFs for rural two-lane undivided segments (i.e., R2U) in Alabama; (2) collecting and incorporating explanatory variables that have been rarely considered in bicycle crash SPFs, such as presence of bus and bicycle signs/lanes, presence of parking, and number of turn lanes; (3) using the COM-Poisson model to account for bicycle-related crash under-dispersion to highlight the significant predictors affecting bicycle crash frequency; and (4) using reliable bicycle crash data through a detailed review of police reports to confirm that the crash was indeed bicycle-related. To achieve the aforementioned study objectives, 1,311 bicycle-vehicle crashes in Alabama (representing five-year crashes from 2011 to 2015) were used and their respective police reports were reviewed in details.

2. Prior Research

In an attempt to identify the contributing factors affecting pedestrian and bicycle crash counts, Chimba et al. (2014) employed the NB model and found that the number of lanes crossed by pedestrians or bicyclists, posted speed limit, and presence of special speed zones significantly influenced the frequency of pedestrian and bicycle crashes. Along the same context, Cai et al. (2020) investigated the effects of bicycle facilities and other contributing features on bicycle crashes at intersections. They found that adding bicycle lanes, wide shared paths, and the less distance between the shared path and edge of the roadway were positively related to bicycle crash counts. There have been other studies that investigated bicycle-vehicle crashes along roads with bicycle lanes (e.g., Lusk et al., 2015), street parking lots (e.g., Chen, 2015), and highly motorized city areas (e.g., Yao and Loo, 2016).

Investigating the relationship between built environment factors and bicycle-vehicle crashes in Seattle, Chen (2015) used a Poisson lognormal random effects model and found that zones with more road signals, street parking signs, and greater automobile trips were likely to have more bicycle-vehicle crashes. Amoh-Gyimah et al. (2016) explored the impact of built environment factors associated with bicycle crashes using a random parameter negative binomial (RPNB) model. The results showed that vehicle kilometers traveled, population, mixed type of lane use, and percentage of commuters cycling had a significant positive correlation with the number of bicycle-vehicle crashes. Hamann and Peek-Asa (2013) examined the effects of environmental, and demographic characteristics on bicycle crashes at intersections and non-intersections in Iowa between 2007 and 2010. Obscured vision, rural areas, and poor lightening were found associated with bicycle crash frequency at non-intersection locations.

Developing bicycle-vehicle crash SPFs has been the emphasis of other research efforts over the years (e.g., Dolatsara, 2014; Nordback et al., 2014; Pulugurtha and Thakur, 2015; Gates et al., 2016; Kwigizile et al., 2016; Thomas et al., 2017; Chen et al., 2020; Mukoko and Pulugurtha, 2020). For instance, focusing on urban areas, Gates et al. (2016) developed bicycle SPFs along urban segments and intersections. Crash data for a ten-year period were used. In the study, vehicle annual average daily traffic (AADT) data were used since bicycle volume data were unavailable. It was found that two-lane undivided and four-lane divided urban segments had lower proportions of bicycle crashes, while intersections had more crash frequencies than road segments. Most recently, Mukoko and Pulugurtha (2020) developed SPFs for bicycle-vehicle crashes in the urban area of Mecklenburg County in North Carolina using NB log-link distribution-based models. Similar to previous studies, the results showed that crashes occurred more frequently in areas with no bicycle lanes. Moreover, the absence of traffic lights and high speed limit segments of 45 mph showed positive correlation with bicycle-vehicle crash frequency increase.

From the review of previous studies, it was found that most studies that developed SPFs for exploring the factors affecting crashes at segments and intersections used the NB regression model and its extensions (see for example, Bauer and Harwood, 2000; Srinivasan and Carter, 2010; Mehta and Lou, 2013; Donnell et al., 2014; Xie and Chen, 2016; Rista et al., 2018). In addition, the COM-Poisson model was mostly used for developing SPFs for vehicle-vehicle crashes (see Lord et al., 2008, Lord et al., 2010; Geedipally and Lord, 2011). However, interestingly, developing SPFs for bicycle-vehicle crashes using the COM-Poisson model was not previously researched.

Some safety researchers examined the use of machine learning techniques in crash analysis. Examples include: MARS (Abdel-Aty and Haleem, 2011; Gaweesh et al., 2019; Yue at al., 2019; Park et al., 2019), support vector machine (SVM) (Li et al., 2008), and neural networks (Bargegol et al., 2016; Ulak et al., 2020; Kononov et al., 2008). Regarding studies adopting the MARS model, Gaweesh et al. (2019) developed three SPFs of NB, MARS, and spatial autoregressive (SAR) models to quantify the impact of connected vehicles in the full 402-mile of I-80 corridor in Wyoming. The results showed the overperformance of the MARS model compared to the other two models. In another recent study, Yue et al. (2019) employed both MARS and generalized linear model (GLM) with four years of Florida crash data to quantify the safety impact of signalization at Florida's rural three-leg stop-controlled intersections. They found that MARS had a better fit than the GLM. Park et al. (2019) used MARS model to evaluate the safety effects of various installed freeway designed features. The performance of the MARS model was compared to NB and the results indicated the overperformance of the MARS model.

Interestingly, a few studies used machine learning techniques to analyze non-motorized safety (see for example, Zangenehpour et al., 2015). However, there is a lack of studies that aimed at improving prediction accuracy of specifically bicycle-vehicle crashes using advanced machine learning methods.

This study fills this gap by developing bicycle-vehicle SPFs on both intersections and segments (including both urban and rural segments) in Alabama using both non-machine learning (i.e., COM-Poisson) and machine learning (i.e., MARS) models. Extensive review of police reports has been conducted and rarely-explored variables have been used, e.g., presence of bus and bicycle signs, presence of parking, number of turn lanes, and median type (as discussed in the next section).

3. Data Collection and Preparation

Five-year bicycle-related crashes (2011 to 2015) were retrieved from the Critical Analysis Reporting Environment (CARE), which represent the most recent crash data available. A total of 1,311 crashes were obtained and the associated police reports were reviewed in detail. Fig. 1 shows the distribution of the 1,311 bicycle-vehicle crashes that occurred in a five-year span (2011-2015) in the state of Alabama.

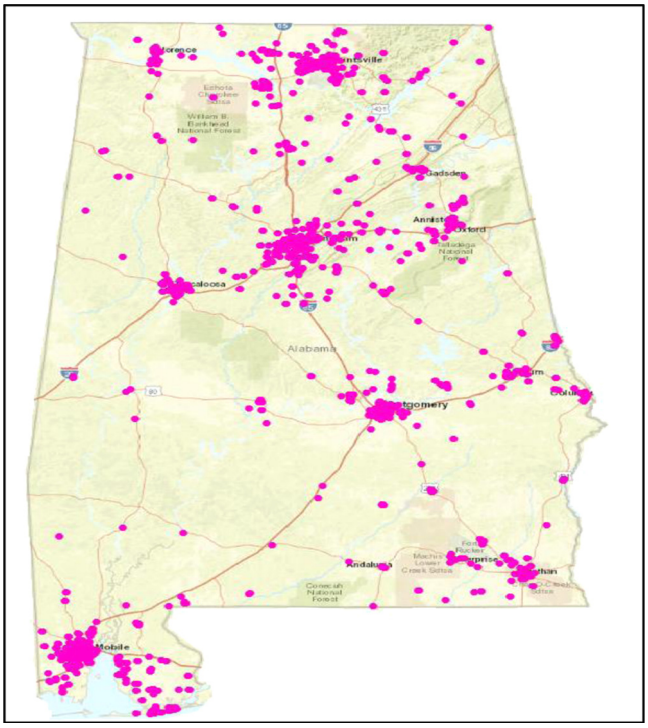

Fig. 1. Statewide Bicycle-Vehicle Crash Cluster Map in Alabama

The research team reviewed in details the police reports to make sure that the crash was indeed bicycle-related and to collect relevant information from the police reports. Specifically, the police report's narrative, detailed descriptions, and sketch were beneficial for that purpose and to collect the necessary information, e.g., bicycle crash manner "how the crash happened and the specific crash cause" and at-fault road user "either bicyclist or vehicle driver". Furthermore, the police report provided the latitude/longitude of each crash, which helped geo-locate the crash on Google Maps 2017, then collect the relevant roadway/geometric and traffic features at the crash location, specifically those rarely-explored in previous bicycle safety studies. Examples include: number of turn lanes on both major and minor roads, type of control on minor road (stop, yield, no control, or signalized), skewness angle, median type on major road, median presence on minor road, bus stop/sign presence on major road, and parking presence on major and minor roads. It should be noted that if a crash was within 250 feet from the center of the intersection, it was considered as an intersection-related crash; otherwise, considered as a segment-related crash.

3.1. Intersection Data Preparation

Intersection SPFs have included both signalized and unsignalized (or stop-controlled) intersections. Traffic data were available from the Alabama Department of Transportation (ALDOT) for both major and minor roads. However, the AADT data were not available for the minor approaches of the three-leg and four-leg unsignalized intersections since most of these approaches were located on non-state-maintained roads.

Relevant geometric and operational data for both major and minor approaches were collected using Google Earth Pro (2018). Later, AADT data, geometric/operational data, zero-crash intersections, and crash data were added as different layers in Google Earth. Note that a total of 35 intersections (having zero crashes) were collected for each calibrated intersection facility in the HSM to satisfy HSM's requirement on having at least 30 sites for SPF model fitting.

Figs. 2(a) and 2(b) show sample urban three-legged stop-controlled and urban three-legged signalized intersections, respectively. Note that SPFs for other intersection facilities in the HSM (e.g., rural four-leg signalized intersections and rural four-leg unsignalized intersections) could not be calibrated due to the very limited sample sizes. Moreover, four-leg signalized intersections in urban areas were aggregated with their three-leg counterparts to enrich the sample size. Table 1 shows descriptive statistics for the variables used in intersection SPFs development.

3.2. Segment Data Preparation

In this paper, the variable length segmentation method was employed for segmenting the roadway stretches. In this method, roadway sections were divided into homogeneous roadway segments of variable lengths. Homogeneous roadway segments imply that these segments have consistent roadway/geometric and traffic characteristics (e.g., number of lanes and AADT).

For segments data preparation, Google Earth Pro (2018) was used to geo-locate and cluster segment crashes onto the roadway network. Three data sources of the crash, AADT (kmz file format), and roadway network (readily available in Google Earth) were used for preparing and analyzing the data. The AADT shape file, which lists the AADT values for each state route at specific traffic counters, consists of street names, beginning milepost, end milepost, route ID, and AADT. Google Earth provides street names for the roadway network and the crash data source includes crash location, collision date and time, facility information, driver information, environmental, and vehicle information.

Then, the data for AADT and crashes were added as two different layers into Google Earth. Afterwards, AADT was assigned to the roadway network, dividing the roadway into different segments based on a change in AADT (and implicitly a change in the number of lanes as well). In other words, a new segment was assigned whenever there was a change either in the AADT or number of roadway lanes. Segments were then divided and separated into urban and rural segments. Afterwards, a unique ID was then created to identify each roadway segment. The unique ID generated consisted of 20 digits which included route type, followed by route name and then beginning milepost (represented by 4 digits) and end milepost (represented by 4 digits). The idea behind creating new segment IDs is to differentiate those segments having the same roadway ID.

Once unique segments IDs were created, the crash file included the latitude/longitude for each crash, where each crash was geocoded in Google Earth. Crashes were then assigned to their respective segments (for segmenting each of the U2U, U4DU, and R2U facilities). Note that a total of 35 segments (having zero crashes) were collected for each calibrated segment facility in the HSM to satisfy HSM's requirement on having at least 30 sites for SPF model fitting. Figs. 2(c), 2(d), 2(e), and 2(f) show urban four-lane divided arterial, urban four-lane undivided arterial, urban two-lane undivided arterial, and rural two-lane undivided highway segment, respectively.

Table 2 provides descriptive statistics for the variables used in segment SPFs development. It should be noted that the research team could identify the existence of left-turn lanes at road segments (and not necessarily at intersections) since two main cases arose. These two cases involved the existence of left-turn lane (at a median opening) and two-way left-turn lane median, outside the 250-feet intersection influence area, that were used exclusively for U-turns. Note that each of the intersection and segment datasets was split into 70% (training data) and 30% (validation data) as part of the SPF development.

4. Methodology

4.1. Conway–Maxwell–Poisson (COM-Poisson) Model

Crash data are mostly characterized by over-dispersion, i.e., variance is greater than the mean. However, crash dataset might experience under-dispersion, which can be generated due to the modeling

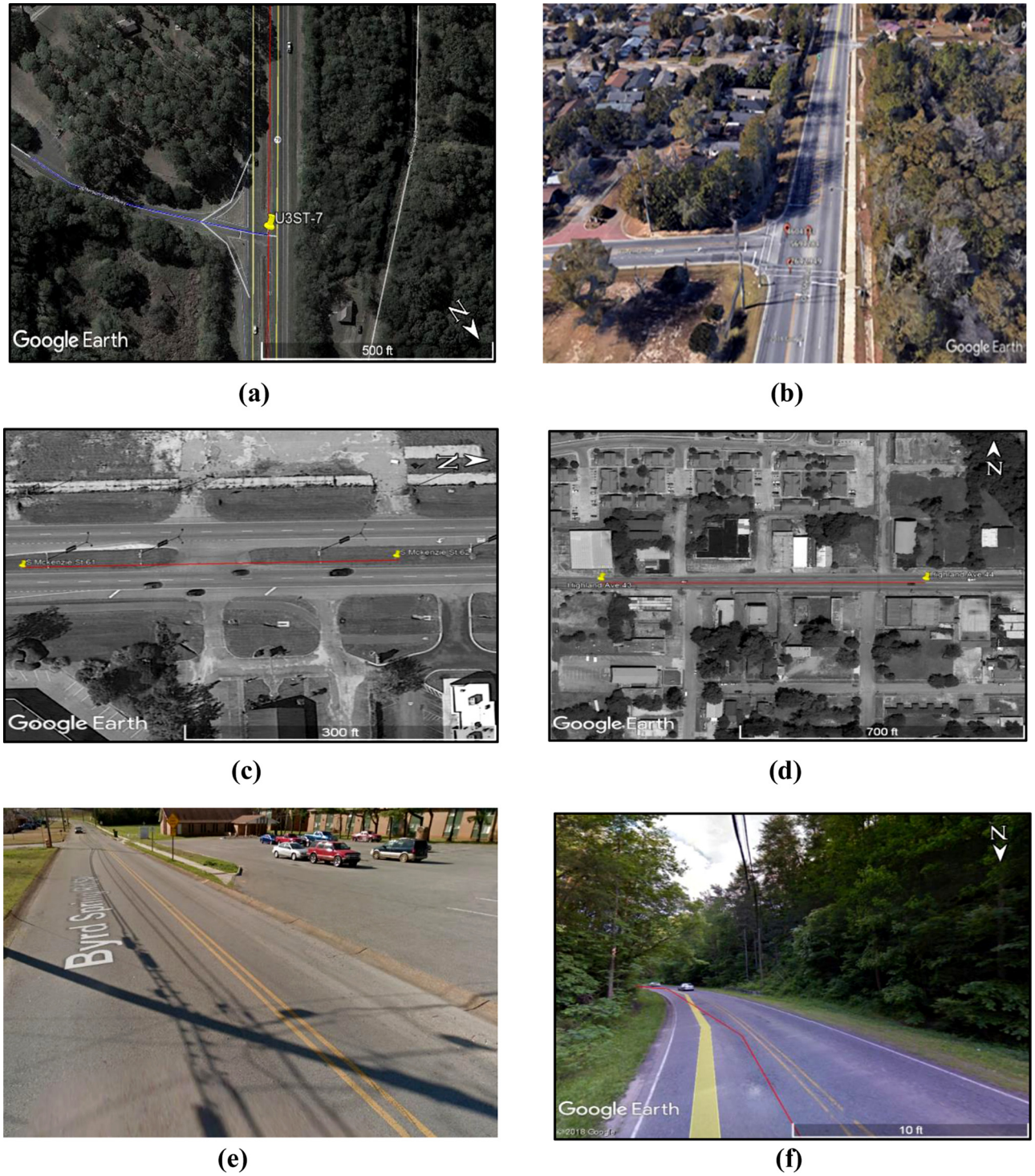

Fig. 2. (*a*) Urban Three-Legged Stop-Controlled Intersection, (*b*) Urban Three-legged Signalized Intersection, (*c*) Urban Four-Lane Divided Arterial Segment with Open Median, (*d*) Urban Four-Lane Undivided Arterial Segment, *(e)* Urban Two-Lane Undivided Arterial Segment, and (*f*) Rural Two-Lane Undivided Highway Segment (Google Earth Pro, 2018)

process (Sellers et al., 2012), data generating (Lord et al., 2008; Oh et al., 2006), and low sample mean datasets (Lord and Guikema, 2012). Although traditional Poisson and NB have been widely used for count crash data modeling for many years, they have limitations to model such under-dispersed data (Geedipally and Lord, 2011).

The COM-Poisson distribution (Conway and Maxwell, 1962) is a discrete probability distribution which can capture a wide range of dispersion types (Sellers and Shmueli, 2010). The probability mass function of COM-Poisson is defined in Equation (1):

$$P(Y=y)=\frac{\lambda^{y}}{(y!)^{\nu}\,Z(\lambda,\nu)},\ y=0,1,2,\ldots,n,\ \ \lambda>0,\ \nu\geq 0 \tag{1}$$

where: Y is a discrete random variable; $Z(\lambda,\nu)=\sum_{n=0}^{\infty}\frac{\lambda^{n}}{(n!)^{\nu}}$ is a normalizing constant; ν is dispersion parameter ($\nu>1$ is considered under-dispersed and $\nu<1$ is over-dispersed); n is total number of variables; and λ is the centering parameter.

Guikema and Goffelt (2008) proposed a new formulation by substituting $\mu=\lambda^{1/\nu}$. This is shown in Equations (2) and (3).

Table 1
: Variables' Descriptive Statistics for Intersections

Variable of Interest	USG Statistics			UST Statistics		
Continuous Variables						
Major road AADT	Max.: 56,330 Min.: 1,640 Mean: 18,015.21 St. dev.: 10,909.98			Max.: 44,200 Min.: 100 Mean: 8,852.64 St. dev.: 9,231.33		
Minor road AADT	Max.: 31,400 Min.: 501 Mean: 5,972.09 St. dev.: 4,846.88			—		
Speed limit (mph)	Max.: 55 Min.: 20 Mean: 35.02 St. dev.: 7.68			Max.: 65 Min.: 5 Mean: 30.56 St. dev.: 8.23		
Variable of Interest	USG Variable Levels	USG Statistics[a] No. of Crash.	No. of Sites	UST Variable Levels	UST Statistics[b] No. of Crash.	No. of Sites
Indicator (Categorical) Variables						
Roadway curvature and grade	Straight level	107	102	Straight level	257	257
	Straight downgrade/upgrade and curved	22	64	Straight downgrade/upgrade and curved	95	143
Presence of turn lane on major road	Turn lane exists	109	139	Turn lane exists	130	149
	No turn lane	20	27	No turn lane	222	251
Presence of turn lane on minor road	Turn lane exists	96	126	Turn lane exists	24	34
	No turn lane	33	40	No turn lane	328	366
Number of left-turn lanes on major road (both approaches)	Two, three, or four left-turn lanes	82	99	One, two, or three left-turn lanes	121	138
	No or one left-turn lane	47	67	No left-turn lanes	231	262
Number of left-turn lanes on minor road (both approaches)	Two, three, or four left-turn lanes	48	67	One left-turn lane	15	23
	No left-turn or one left-turn lane	81	99	No left-turn lanes	337	377
Number of right-turn lanes on major road (both approaches)	One, two, or three right-turn lanes	41	70	One, two right-turn lanes	36	41
	No right-turn lanes	88	96	No right-turn lanes	316	359
Number of right-turn lanes on minor road (both approaches)	One, two, or three right-turn lanes	60	90	One or two right-turn lanes	22	30
	No right-turn lanes	69	76	No right-turn lanes	330	370
Type of median on major road	Undivided	97	124	Undivided	250	281
	Divided (i.e., closed and open)	32	42	Divided (i.e., closed and open)	102	119
Presence of median on minor road	Median exists	15	27	Median exists	102	119
	No median	114	139	No median	250	281
Number of through lanes on major road (both directions)	Four or more through lanes	91	115	Four or more through lanes	101	119
	One, two, or three through lanes	38	51	One, two, or three through lanes	251	281
Number of through lanes on minor road	Three or more through lanes	34	54	Two, three, or four through lanes	331	378
	One or two through lanes	95	112	One through lane	21	22
Presence of parking on major road	Parking exists	19	25	Parking exists	65	80
	No parking	110	141	No parking	287	320
Presence of parking on minor road	Parking exists	35	43	Parking exists	189	205
	No parking	94	123	No parking	163	195
Presence of median on major road	Median exists	42	69	Median exists	102	119
	No median	87	97	No median	250	281
Presence of bicycle lane/sign on major road	Bicycle lane/sign exists	19	16	Bicycle lane/sign exists	48	60
	No bicycle lane/sign	110	150	No bicycle lane/sign	304	340
Presence of bus stop/sign on major road	Bus stop/sign exists	49	44	Bus stop/sign exists	51	57
	No bus stop/sign	80	122	No bus stop/sign	301	343
Skewed intersection	Skewed	91	115	Skewed	70	85
	Non-skewed	38	51	Non-skewed	282	315
Sight distance issue on major road	Sight dist. issue exists	17	17	Sight dist. issue exists	80	79
	No sight dist. issue	112	149	No sight dist. issue	272	321
Intersection size	Four-legged	110	126	Four-legged	165	188
	Three-legged	19	40	Three-legged	187	212

[a] *USG: Total Crashes = 192 (Training = 129, Validation = 63), Total Sites = 242 (Training = 166, Validation = 76)*
[b] *UST: Total Crashes = 506 (Training = 352, Validation = 154), Total Sites = 572 (Training = 400, Validation = 172)*

$$P(Y = y) = \frac{1}{S(\mu, \nu)}\left(\frac{\mu^y}{y!}\right)^\nu \quad (2)$$

$$S(\mu, \nu) = \sum_{n=0}^{\infty}\left(\frac{\mu^n}{n!}\right)^\nu \quad (3)$$

Considering a generalized linear model (GLM) approach, the modeling framework of COM-Poisson regression, developed by Guikema and Goffelt (2008), is considered flexible and can model both over-dispersed and under-dispersed datasets (Sellers and Shmueli, 2010). The link functions of the COM-Poisson are shown in Equations (4) and (5).

$$\ln(\mu) = \beta_0 + \sum_{i=1}^{p}\beta_i x_i \quad (4)$$

Table 2
Variables' Descriptive Statistics for Segments

Variable of Interest	Statistics Levels	U2U Statistics		U4DU Statistics		R2U Statistics	
Continuous Variables							
Major road AADT	Max.	30,570		43,420		16,880	
	Min.	201		1,450		190	
	Mean	589.08		17,848.56		3,323.46	
	St. dev.	4,346.68		8,865.11		3,724.71	
Segment length (miles)	Max.	8.78		5.13		7.06	
	Min.	0.10		0.10		0.10	
	Mean	0.73		0.49		1.48	
	St. dev.	1.28		0.75		1.56	
Variable of Interest	Variable Levels	U2U Statistics[a]		U4DU Statistics[b]		R2U Statistics[c]	
		No. of Crash.	No. of Sites	No. of Crash.	No. of Sites	No. of Crash.	No. of Sites
Indicator (Categorical) Variables							
Speed limit (mph)	Speed limit ≤ 25	49	62	—	—	—	—
	25 < Speed limit ≤ 35	39	45	—	—	—	—
	Speed limit > 35	34	38	—	—	—	—
	Speed limit ≤ 35	—	—	17	44	14	20
	35 < Speed limit ≤ 45	—	—	37	69	33	47
	Speed limit > 45	—	—	12	19	23	27
Region	East/West central	40	46	23	42	15	23
	Southeast/Southwest	59	72	29	67	28	41
	North	23	27	14	23	27	30
Roadway curvature and grade	Straight level	95	107	57	79	41	44
	Straight downgrade/upgrade and curved	27	38	9	53	29	50
Presence of turn lane	Turn lane exists	12	11	43	82	2	2
	No turn lane	110	134	23	50	68	92
Number of left-turn lanes	One or more left-turn lanes	9	9	33	60	2	2
	No left-turn lanes	113	136	33	72	68	92
Number of right-turn lanes	One or more right-turn lanes	8	7	8	20	2	2
	No right-turn lanes	114	138	58	112	68	92
Presence of parking On Major road	Parking exists	24	29	4	6	1	1
	No parking	98	116	62	126	69	93
Type of median on major road	Divided (two-way left-turn lane)	—	—	33	53	—	—
	Divided (i.e., closed and open)	—	—	14	40	—	—
	Undivided	—	—	19	39	—	—
Presence of bicycle lane/sign on major road	Bicycle lane/sign exists	7	8	4	6	0	1
	No bicycle lane/sign	115	137	62	126	70	93
Presence of bus stop/sign on major road	Bus stop/sign exists	11	15	8	21	6	7
	No bus stop/sign	111	130	58	111	64	87

[a] *U2U: Total Crashes = 184 (Training = 122, Validation = 62), Total Sites = 205 (Training = 145, Validation = 60)*
[b] *U4DU: Total Crashes = 123 (Training = 66, Validation = 57), Total Sites = 164 (Training = 132, Validation = 32)*
[c] *R2U: Total Crashes = 111 (Training = 70, Validation = 41), Total Sites = 128 (Training = 94, Validation = 34)*

$$\ln(\nu) = \beta_0 + \sum_{i=1}^{q} \beta\gamma_i z_i \tag{5}$$

where: β and γ are coefficients; and x_i and z_i are covariates. It is assumed to have p covariates used in the centering link function and q covariates used in the shape link function. Note that i in Equations (4) and (5) refers to the covariate (variable) ID.

Lord et al. (2010) modified the formulation of Guikema and Goffelt (2008) by removing the second link (Equation 5) and estimating a single shape parameter (ν). The analysis here follows the method by Lord et al. (2010) using the functional form previously shown in Equation (4). Note that the COM-Poisson model overcomes data overfitting as it captures more general patterns in all of the data and not necessarily parts of the data.

4.2. Multiplicative Adaptive Regression Splines (MARS) Technique

MARS is a non-parametric regression method and is considered a piecewise linear regression approach (an extension of linear models) that deals with non-linearities and interactions between variables. MARS requires no prior assumptions about the relationship between the response variable and predictor variables, i.e., there is no specific distribution between the response variable and the predictors. According to Put et al. (2004), MARS model is described as follows:

$$\hat{y} = a_0 + \sum_{m=1}^{M} a_m B_m(x) \tag{6}$$

where: $\hat{y}$ is the predicted response; a_0 is the coefficient of the constant basis function; $B_m(x)$ the mth basis function, which may be a single space function or a product (interaction) of two (or more) spline functions; a_m is the coefficient of the mth basis function; and M is the number of basis functions that are included in the model.

In general, three steps are involved to develop MARS model according to Put et al. (2004). The first step is a constructive phase, where the global model (shown in Equation (6)) is defined. According to Abdel-Aty and Haleem (2011), in this step, the MARS model basis functions consist of either single spline function or product of two (or more) spline functions (i.e., interactions). Those basis functions are added in a "two-at-a-time" forward stepwise procedure, which selects the best pairs of spline functions in order to improve the model. Each pair consists of one left-sided and one right-sided truncated function defined by a given knot location, as shown in Equations (7), (8), respectively.

$$[-(x-t)_+^q] = \begin{Bmatrix} (t-x)^q;\ x < t \\ 0;\ otherwise \end{Bmatrix} \tag{7}$$

$$[+(x-t)_+^q] = \begin{Bmatrix} (x-t)^q;\ x > t \\ 0;\ otherwise \end{Bmatrix} \tag{8}$$

where knot t is the quantile of the marginal distribution of a continuous

predictor.

The second step is the pruning phase, in which the least important basis functions of the model are eliminated in a "one-at-a-time" backward elimination procedure. This step is based on the generalized cross-validation (GCV) criterion, where GCV overcomes data overfitting.

Finally, in the third step, the best model is selected based on evaluation of the predictive properties of different models. The mean absolute deviance (MAD) and mean square prediction error (MSPE) criteria were used to evaluate prediction performance of different MARS and COM-Poisson models. MAD and MSPE are calculated using Equations (9) and (10), respectively. Optimal models are obtained by having smaller MAD and MSPE estimates.

$$MAD = \frac{1}{n}\sum_{i=1}^{n} |\hat{y}_i - y_i| \quad (9)$$

$$MSPE = \frac{1}{n}\sum_{i=1}^{n} (\hat{y}_i - y_i)^2 \quad (10)$$

where: n is sample size in validation dataset; $\hat{y}_i$ is observed crash frequency at site i; and y_i is predicted crash frequency for site i.

5. Results and Discussions

5.1. COM-Poisson Model Results

Using PROC COUNTREG in SAS statistical software (SAS OnDemand for Academics 2017), SPFs (based on the "COM-Poisson" model) were developed for both intersections and segments in the HSM. The final fitted models for signalized and unsignalized (or stop-controlled) intersections and segments are shown in Tables 3 and 4, respectively. While fitting the models, the three-model building techniques were examined: stepwise, forward addition, and backward elimination. Variables that were significant at an 80% confidence level (i.e., P-value $\leq$ 0.2) were retained in the models; otherwise, excluded. Other researchers have also used 80% confidence in model fitting (see for examples, Shirazi et al., 2016; Siddiqui and Abdel-Aty, 2012; Abdel-Aty et al., 2013). Note that a multicollinearity test was performed for all five fitted SPFs and no multicollinearity issue was found to exist, where all model parameters had variance inflation factors (VIFs) less than 10 (Maxwell et al., 2018).

5.1.1. Urban Signalized (USG) Intersection SPF Results

As shown in Table 3, the "-ln(NU)" coefficient (which is an indirect measure of the dispersion parameter) is negative and statistically significant, which verifies that the data are under-dispersed. It should be noted that the parameter "NU" indicates the amount of dispersion. Therefore, the Poisson distribution is recovered when NU = 1, while over-dispersion corresponds to NU < 1 and under-dispersion corresponds to NU > 1. Thus, when the value of NU is greater than 1, "-ln (NU)" is negative, which shows that the data are under-dispersed.

From Table 3, the number of bicycle-vehicle crashes was negatively correlated with the presence of curves on USG intersections, as well as the downgrade and upgrade stretches. This result shows that both bicyclists and vehicle drivers were more attentive on curved roads compared to straight roads. The crash frequency had a positive correlation with the absence of right-turn lanes on the minor approach. This shows that the existence of right-turn lanes could improve bicyclist safety at

Table 3
: Bicycle-Vehicle Crash Prediction Models (SPFs) for Urban Signalized (USG) and Urban Stop-Controlled (UST) Intersections in the HSM

Parameter	USG Model			UST Model		
	Estimate	St. Err.	P-Value	Estimate	St. Err.	P-Value
Intercept	−1.941	0.493	< .0001	−1.013	0.100	< .0001
Straight downgrade/upgrade & curved	−0.343	0.057	< .0001	−0.268	0.033	< .0001
Reference: Straight level	—	—	—	—	—	—
Presence of turn lane on major app.	N/S[a]	N/S	N/S	0.059	0.039	0.130
Reference: No turn lanes	—	—	—	—	—	—
No right-turn lanes on minor app.	0.088	0.058	0.129	N/S	N/S	N/S
Reference: One/two/three right turns	—	—	—	—	—	—
Presence of median on major app.	−0.150	0.061	0.015	N/S	N/S	N/S
Reference: No median	—	—	—	—	—	—
Divided median on major approach	N/S	N/S	N/S	−0.058	0.039	0.140
Reference: Undivided median	—	—	—	—	—	—
Bus stop/sign on major approach	0.147	0.061	0.016	N/S	N/S	N/S
Reference: No bus stop/sign	—	—	—	—	—	—
Presence of sight distance issue on major approach	N/S	N/S	N/S	0.190	0.042	< .0001
Reference: No sight distance issue	—	—	—	—	—	—
Three-legged intersection	−0.129	0.067	0.053	N/S	N/S	N/S
Reference: Four-legged intersection	—	—	—	—	—	—
Presence of parking on major app.	N/S	N/S	N/S	−0.092	0.029	0.001
Reference: No parking	—	—	—	—	—	—
Natural log. of AADT (major road)	0.118	0.046	0.011	−0.034	0.011	0.002
Natural log. of AADT (minor road)	−0.055	0.027	0.041	—	—	—
Dispersion parameter: -ln(NU)[b]	−1.984	0.117	< .0001	−2.751	0.101	< .0001
Goodness-of-Fit Statistics						
Training Dataset[c]	No. of Crashes = 129 (*Mean = 0.78, Var. = 0.36*) No. of Sites = 166			No. of Cashes = 352 (*Mean = 0.88, Var. = 0.12*) No. of Sites = 400		
Generalized (Pseudo) R-Square	0.08			0.10		
Akaike Information Criterion (AIC)	209.18			216.47		
Model Prediction Performance Assessment						
Validation Dataset[c]	No. of Crashes = 63 No. of Sites = 76			No. of Crashes = 154 No. of Sites = 172		
MAD (Mean Absolute Deviance)	0.54			0.43		
MSPE (Mean Square Pred. Error)	0.40			0.27		

[a] N/S means not significant
[b] Negative value of "-ln(NU)" validates existence of under-dispersed bicycle-vehicle crash dataset
[c] Total number of sites (having zero crashes) in the training and validation datasets is 35

Table 4
: Bicycle-Vehicle Crash Prediction Models (SPFs) for Roadway Segments in the HSM

Parameter	U2U Model		U4DU Model		R2U Model	
	Estimate[a]	P-Value	Estimate[a]	P-Value	Estimate[a]	P-Value
Intercept	0.637 (0.198)	0.001	0.488 (0.180)	0.006	−0.069 (0.067)	0.303
Natural log. of AADT	−0.059 (0.025)	0.019	N/S[b]	N/S	N/S	N/S
North region in Alabama	N/S	N/S	N/S	N/S	0.172 (0.075)	0.023
Reference: Southeast/Southwest	—	—	—	—	—	—
25 < Speed limit ≤ 35	0.094 (0.063)	0.135 0.023	—	—	—	—
Speed limit > 35			—	—	—	—
Reference: Sp. limit ≤ 25	0.159 (0.070) —	—	—	—	—	—
Speed limit ≤ 35	—	—	−0.200 (0.101)	0.047 0.011	N/S 0.131 (0.075) —	N/S 0.081 —
Speed limit > 45	—	—				
Reference: 35 < Sp. limit ≤ 45	—	—	0.357 (0.140) —	—		
Straight downgrade/upgrade & curved	−0.113 (0.055) —	0.038 —	−0.621 (0.124) —	< .0001 —	— —	— —
Reference: Straight level						
Straight level	—	—	—	—	0.247 (0.067) —	0.000 —
Reference: Straight downgrade/ upgrade & curved	—	—	—	—		
Two left-turn lanes	N/S	N/S	−0.242 (0.171) —	0.150 —	N/S —	N/S —
Reference: No left-turn or one left-turn lane	—	—				
Divided median (i.e., closed and open median types)	— —	— —	−0.503 (0.184) —	0.006 —	— —	— —
Reference: Undivided and two-way left-turn lanes						
Dispersion parameter: -ln(NU)[c]	−2.092 (0.122)	< .0001	−1.458 (0.179)	< .0001	−1.967 (0.162)	< .0001
Goodness-of-Fit Statistics						
Training Dataset[d]	No. of Crashes = 122 (*Mean = 0.84, Var. = 0.18*) No. of Sites = 145		No. of Crashes = 66 (Mean = 0.50, Var. = 0.36) No. of Sites = 132		No. of Crashes = 70 (Mean = 0.74, Var. = 0.26) No. of Sites = 94	
Generalized (Pseudo) R-Square	0.02		0.06		0.04	
Akaike Information Criterion (AIC)	164.25		187.37		131.55	
Model Prediction Performance Assessment						
Validation Dataset[d]	No. of Crashes = 62 No. of Sites = 60		No. of Crashes = 57 No. of Sites = 32		No. of Crashes = 41 No. of Sites = 34	
MAD (Mean Abs. Deviance)	0.46		0.57		0.33	
MSPE (Mean Sq. Pred. Error)	0.40		0.70		0.23	

[a] Standard error in parentheses
[b] N/S means not significant
[c] Negative value of "-ln(NU)" validates existence of under-dispersed bicycle-vehicle crash dataset
[d] Total number of sites (having *zero crashes) in the training and validation datasets is 35*

signalized intersections (also found in (Haleem and Abdel-Aty, 2010; Harwood et al., 2003)). There is a reduction in the number of bicycle-vehicle crashes with the existence of the median on the major approach of a signalized intersection. This finding conforms with that from Miranda-Moreno et al. (2011) and Sam et al. (2018).

The presence of bus stops near USG intersections was found to increase bicycle-vehicle crash frequency. The presence of bus stops effectively reduces the width of the approach road and could create sight distance issues. Furthermore, the presence of bus stops can result in higher conflicts because stationary buses may become obstacles to the moving traffic (Chin and Quddus, 2003). Three-legged signalized intersections experienced lesser bicycle-vehicle crash frequency, as opposed to the four-legged ones (also found in (Pathivada and Perumal, 2019)).

The number of bicycle-vehicle crashes was positively correlated with the traffic volume on the major approach (i.e., the annual average daily traffic or AADT). This implies that the number of bicycle crashes could increase in heavy traffic flow on the major approach at signalized intersections (also deduced in (Alluri et al., 2017; Harnen et al., 2003; Miranda-Moreno et al., 2011)). Conversely, there is a reduction in the number of bicycle-vehicle crashes with increasing vehicular traffic on the minor approaches of USG intersections (also found in (Gates et al., 2018)).

As seen in Table 3, the MAD and MSPE values from the validation dataset are relatively low, indicating a reliable and well-fitted SPF. The generalized (or Pseudo R-square) estimate is 0.08, which is an acceptable value for count regression models. Fig. 3 shows the plot of the mean of observed crashes and predicted crashes (produced from the COM-Poisson regression model) against the major and minor road AADT. The two scatter plots show that the predicted bicycle-vehicle crash counts have an upward and a downward trend against major and minor road AADT, respectively. This was the same conclusion found in the COM-Poisson model in Table 3.

5.1.2. Urban Unsignalized/Stop-Controlled (UST) Intersection SPF Results

Similar to the USG model, the "-ln (NU)" coefficient is negative and statistically significant, which validates the use of the COM-Poisson regression model to accommodate crash data under-dispersion. Similar

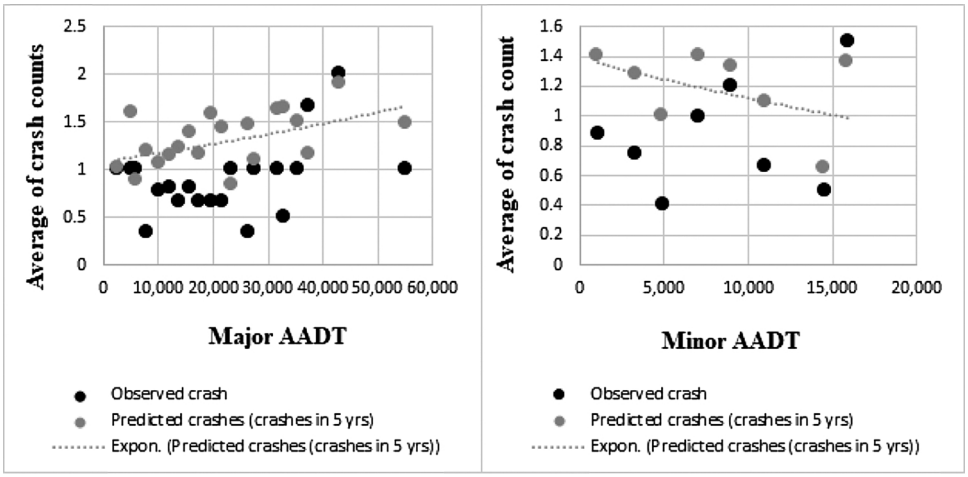

Fig. 3. Scatter Plot of Mean Observed and Predicted Crashes vs. Major and Minor AADT for Urban Signalized (USG) Intersections

to the USG intersections model results, there was a reduction in the number of bicycle-vehicle crashes on stop-controlled intersections along curves and downgrades/upgrades. The presence of a turn lane on the major approach of an UST intersection was associated with higher bicycle crash frequency. This might be due to driver inattentiveness and failure to yield for bicyclists (with the presence of turn lane on the major approach of an unsignalized intersection).

As previously concluded in the USG intersections model, the presence of divided median (i.e., closed, open, or two-way left-turn lane) was associated with a lesser number of bicycle crashes, as opposed to undivided medians. The physical separation is a key in improving bicyclist safety (Dixon et al., 2013). Sight distance had a positive correlation with the number of bicycle-vehicle crashes on the major road. The presence of a sight obstacle causes poor visibility; thus, the driver cannot detect the bicyclist with peripheral vision.

The presence of parking on the major approach was a significant factor in reducing the number of bicycle crashes at UST intersections. At stop-controlled intersections, the presence of vehicle parking and heavy traffic on a wide main road could force the drivers to slow down and be more attentive to bicyclists. Interestingly, the number of bicycle-vehicle crashes was negatively correlated with AADT on the major approach, possibly since drivers/bicyclists were more cautious in congested conditions near stop-controlled intersections.

The MAD and MSPE values from the validation dataset are relatively low, indicating a reliable and well-fitted SPF. The generalized (or Pseudo R-square) estimate is 0.10, which is an acceptable value for count regression models. Fig. 4(a) shows that there is an increase in the predicted bicycle-vehicle crash counts with increasing the major road AADT. This was the same conclusion found in the COM-Poisson model in Table 3.

5.1.3. Segment SPF Results

The significant variables from the three developed SPFs for U2U, U4DU, and R2U segments are shown in Table 4. Similar to the USG and UST intersection models, the "-ln(NU)" coefficient is again negative and statistically significant, which validates the use of the COM-Poisson regression model to accommodate crash under-dispersion.

For U2U segments, the results showed that bicycle-vehicle crashes were reduced with increasing traffic flow on major roads. Several studies concluded that the bicycle-vehicle crashes decreased with traffic volume increase (see (Siddiqui and Abdel-Aty, 2012; Weir et al., 2009). U2U segments with relatively higher speed limits (i.e., 25 < speed limit ≤ 35 mph & speed limit > 35 mph) experienced more vehicle-bicycle crashes than lower speed limit segments (i.e., speed limit ≤ 25 mph). The number of bicycle-vehicle crashes was negatively correlated with the presence of curves, which reveals that bicyclists were more cautious when riding on curved stretches, as opposed to straight/level roads.

For U4DU segments, the number of vehicle-bicycle crashes were found to decline for segments with speed limit range of less than 35 mph compared to those segments with the range between 40 and 45 mph, while there was an increase in crash counts for segments with higher speed limits (more than 45 mph), which is an intuitive finding. Bicycle-vehicle crash frequency was negatively correlated with the presence of two left-turn lanes compared to the absence or presence of

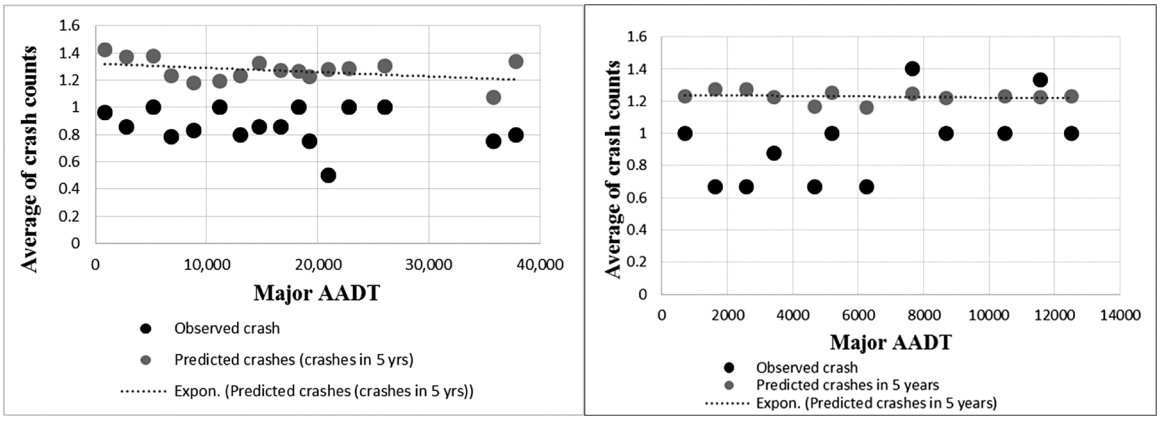

Fig. 4. Scatter Plot of Mean Observed and Predicted Crashes vs. Major AADT for: (*a*) Urban Unsignalized/Stop-Controlled (UST) Intersections, and (*b*) Urban Two-Lane Undivided (U2U) Roadway Segments

only one left-turn lane. This suggests that a greater number of left-turn lanes can effectively reduce bicycle-vehicle crashes, which might be due to the installment of speed limit signs before the horizontal curves (Koloushani et al., 2014; Tabibi et al., 2016). Divided urban four-lane segments experienced a lesser number of bicycle-vehicle crashes compared to undivided segments and those segments with two-way left-turn lanes (also deduced by Phillips et al., 2005).

For R2U segments, northern Alabama regions (e.g., Huntsville, Madison, Colbert, Cullman, Morgan, and Jackson) experienced a higher number of vehicle-bicycle crashes compared to those R2U segments located in the southeast and southwest regions of Alabama (e.g., Mobile, Montgomery, Gulf Shores, and Houston). This result can be justified by the relatively larger area and more population of northern cities compared to their counterparts in the southeast and southwest. There existed more number of bicycle-vehicle crashes on straight and level R2U segments, as opposed to segments with upgrades/downgrades & curves. This result can be attributed to the fact that drivers were possibly less cautious on straight sections and were speeding (Wu et al., 2014). As previously found in U4DU segments, there existed an increase in bicycle-vehicle crashes for R2U segments with higher speed limits (more than 45 mph) compared to those R2U segments with a speed limit range between 40 and 45 mph.

Fig. 4(b) shows that there is a decline in the predicted bicycle-vehicle crash counts with increasing the major road AADT. This was the same conclusion found in the COM-Poisson model in Table 4 (for U2U segments).

5.2. MARS Model Results

The MARS technique was used to develop SPFs for all three segments (U2U, U4DU, and R2U), plus the two intersection facilities (USG and UST). Two sample MARS models (one for U2U segments and one for UST intersections) are shown in Table 5. The R package (R Software, 2019) was used to fit the MARS models using the library "polspline".

As shown in Table 5, both MARS models revealed interaction terms. For R2U segments, one interaction term was "Northern region × Ln_AADT_{major}" and for USG intersections, the interaction was "Curve/down/upgrade × presence of bus stop/sign on major approach". To interpret the interaction term in the USG model, the equation

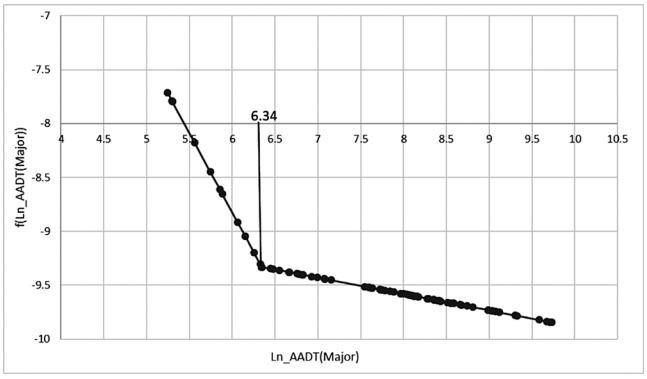

Fig. 5. Plot of Basis Function for Ln_AADT_{major} on R2U Segments

representing the interaction term is:

–1.743× presence of curve or down/upgrade straight – 0.025× presence of bus stop/sign + 0.994× curve/down/upgrade × presence of bus stop/sign

In the above equation, in the presence of curves in the vicinity of USG intersection's approach (i.e., presence of curve = 1), the equation becomes:

–1.743 + (–0.025 + 0.994) × presence of bus stop/sign

The previous equation can be re-written as "*–1.743 + 0.969 × presence of bus stop/sign*". Thus, the final coefficient for "*presence of bus stop/sign*" is "*0.969*". This implies that, with presence of curve on the signalized intersection's approaches, bicycle-vehicle crashes increase in the presence of bus stops/signs on the major approach, as opposed to absence of bus stops/signs. This was the same finding from the COM-Poisson model in Table 3.

From Table 5, the continuous variable (Ln_AADT_{major}) for R2U segment showed non-linear performance, as shown in its truncated basis function at "6.34". To understand the non-linear performance of "Ln_AADT_{major}" for R2U segments, the plot of its basis function is shown in Fig. 5. The basis function "$f(Ln\_AADT_{major})$" can be written as:

$$f(Ln\_AADT_{major}) = -1.47 \times Ln\_AADT_{major} + 1.318 \times (Ln\_AADT_{major} - 6.34)_+$$

From the above equation, the truncated term "(Ln_AADT_{major} –

Table 5
: Sample Segment and Intersection SPFs Using MARS

Basis function	R2U Total sites = 94 Total crashes = 70		USG Total sites = 166 Total crashes = 129	
	Estimate[a]	P-Value	Estimate[a]	P-Value
Intercept	8.314 (3.792)	0.028	1.477 (0.218)	0.000
Ln_AADT_{major}	−1.470 (0.609)	0.015	N/S[b]	N/S
Ln_AADT_{minor}	—	—	−0.010 (0.004)	0.009
Straight level	0.886 (0.277)	0.001	N/S	N/S
Northern region	−7.760 (2.833)	0.006	N/S	N/S
Speed limit > 45 mph	0.169 (0.099)	0.089	—	—
(Ln_AADT_{major} – 6.34)	1.318 (0.672)	0.049	N/S	N/S
North. region & Ln_AADT_{major}[c]	1.349 (0.449)	0.002	N/S	N/S
North. region & (Ln_AADT_{major} – 6.34)[c]	−1.313 (0.493)	0.008	N/S	N/S
North. region & straight level[c]	−0.347 (0.198)	0.080		
Curve and down/upgrade straight	N/S	N/S	−1.743 (0.224)	0.000
Presence of bus stop on major approach	N/S	N/S	−0.025 (0.089)	0.000
Curve/down/upgrade & presence of bus stop/sign on major approach[c]	N/S	N/S	0.994 (0.181)	0.000
(Ln_AADT_{minor} – 37)	—	—	0.015 (0.004)	0.001
Three-legged signalized intersection	—	—	−0.181 (0.081)	0.026
Right-turn lane presence on major approach	N/S	N/S	−0.144 (0.069)	0.036
Goodness-of-Fit Statistics				
Generalized R-Square	0.35		0.53	

[a] Standard error in parentheses
[b] N/S means not significant
[c] Interaction term

Table 6
: MARS and COM-Poisson Model Comparison

Facility Name	Goodness-of-Fit Criterion	MARS	COM-Poisson
Segment SPFs			
U2U	MAD (*model validation*)	0.09	0.46
(*Urban two-lane undivided*)	MSPE (*model validation*)	0.01	0.40
	Generalized R-Square (*model fitting*)	0.50	0.02
U4DU	MAD (*model validation*)	0.53	0.57
(*Urban four-lane divided and undivided*)	MSPE (*model validation*)	0.70	0.70
	Generalized R-Square (*model fitting*)	0.29	0.06
R2U	MAD (*model validation*)	0.27	0.33
(*Rural two-lane undivided*)	MSPE (*model validation*)	0.20	0.23
	Generalized R-Square (*model fitting*)	0.35	0.04
Intersection SPFs			
USG	MAD (*model validation*)	0.28	0.54
(*Urban 3- and 4-legged signalized*)	MSPE (*model validation*)	0.18	0.40
	Generalized R-Square (*model fitting*)	0.53	0.08
UST	MAD (*model validation*)	0.15	0.43
(*Urban 3- and 4-legged stop-controlled*)	MSPE (*model validation*)	0.08	0.27
	Generalized R-Square (*model fitting*)	0.36	0.10

$6.34)_+$" equals "$(\text{Ln\_AADT}_{major} - 6.34)$" whenever $\text{Ln\_AADT}_{major} > 6.34$, and zero otherwise. Fig. 5 demonstrates the non-linear performance of "Ln_AADT_{major}" against bicycle-vehicle crashes, since there exists a knot (break) at "6.34".

6. MARS and COM-Poisson Model Comparison

Table 6 shows the comparison between the fitted MARS and COM-Poisson models for the two intersection facilities (USG and UST) and the three segment facilities (U2U, U4DU, and R2U) using different goodness-of-fit criteria (i.e., MAD and MSPE to assess the prediction accuracy and generalized R-square to assess the goodness of fit of the developed model). It should be noted that lower values of MAD and MSPE and higher value of generalized R-square represent better model performance. As an example, the MARS model developed for USG intersections yielded an MAD value of 0.28 and MSPE value of 0.18, which outperformed the corresponding COM-Poisson USG intersection model (with MAD and MSPE values of 0.54 and 0.40, respectively). Moreover, the generalized R-square value from the MARS model (0.53) outperformed the corresponding COM-Poisson model (having generalized R-square value of 0.08).

Similarly, for the segment facilities, the MARS models outperformed the corresponding COM-Poisson models. For instance, the validation results of U2U segment facilities showed lower values of MAD (0.09) and MSPE (0.01) from the MARS model's prediction compared to the corresponding values from the COM-Poisson model's prediction (MAD = 0.46, MSPE = 0.40). Additionally, the generalized R-square estimate from the MARS model was higher at 0.50, as opposed to a generalized R-square value of 0.02 from the COM-Poisson model.

In general, it can be observed that MARS outperformed the corresponding COM-Poisson models for all five facility types. This can be detected in terms of the prediction performance (i.e., lower MAD and MSPE estimates), as well as model fitting (i.e., higher generalized R-square values). This suggests that MARS is a promising technique for predicting bicycle-vehicle crashes on both segments and intersections (both signalized and unsignalized). Furthermore, the fitted MARS model is interpretable (i.e., does not suffer from black-box limitation).

7. Conclusions and Recommendations

This study developed SPFs and investigated the factors affecting bicycle-vehicle crashes in Alabama using COM-Poisson and MARS techniques on various facilities in the HSM. These were U2U, U4DU, and R2U segments, and USG and UST intersections. The study applied COM-Poisson and MARS to extensive crash data collected and reviewed in detail at segments and intersections in Alabama. A total of 1,311 bicycle-vehicle crashes were analyzed. To the authors' knowledge, no previous studies have explored the application of MARS and COM-Poisson techniques in predicting bicycle-vehicle crashes on both segments and intersections, which was one of the study motivations. In addition, this study developed bicycle-vehicle crash SPFs for rural facilities (i.e., R2U) since very limited studies have attempted to investigate bicycle safety in rural areas. The study also compared the prediction performance of both MARS and COM-Poisson models (in terms of MAD and MSPE) and model fitting goodness-of-fit (i.e., the generalized R-square).

The fitted COM-Poisson models showed several significant variables affecting bicycle-vehicle safety at segments and intersections. Based on the SPFs results for USG intersections, the absence of right-turn lanes on minor roads, the presence of bus stops, and higher major road AADT were significantly associated with increased number of bicycle-vehicle crashes. However, the existence of curved and down/upgrade roads and the presence of divided medians on major approaches were found to reduce bicycle-vehicle crash frequency. From the UST intersections' model, curved and down/upgrade stretches, the presence of divided medians on major approaches, and the presence of parking spaces on major approaches were found to reduce bicycle-vehicle crash frequency. In general, results from the developed SPFs for roadway segments showed that bicycle-vehicle crash frequencies were reduced along curved and downgrade/upgrade stretches on all five facilities (USG, UST, U2U, U4DU, and R2U). Furthermore, crash occurrence was found to decline on those segments with relatively lower speed limits (e.g., less than or equal to 25 mph).

It was concluded that the MARS models outperformed the corresponding COM-Poisson models (based on MAD, MSPE, and generalized R-square values) for all five studied facilities. For example, the validation results of U2U segment facilities showed lower values of MAD (0.09) and MSPE (0.01) from the MARS model's prediction compared to the corresponding values from the COM-Poisson model's prediction (MAD = 0.46, MSPE = 0.40). Additionally, the generalized R-square estimate from the MARS model was higher at 0.50, as opposed to a generalized R-square value of 0.02 from the COM-Poisson model.

This study showed that MARS technique is a superior approach for predicting bicycle-vehicle crashes on intersections and segments in Alabama. Still, the COM-Poisson model is a promising approach for model fitting and successfully identifying the significant predictors of bicycle-vehicle crashes (as it is a simpler model to develop by transportation practitioners and does not require extensive statistical expertise). However, if the main objective of a bicycle safety study is improving bicycle crash predictions, then MARS is recommended for effectively predicting bicycle-vehicle crashes on both segments and intersections. MARS also has the capability of identifying interaction terms that are easily interpretable.

Bicycle-vehicle crashes near USG intersections can be reduced by repositioning bus stops to be further away from the intersection approaches. Sight distance improvements are suggested on UST intersections' corners through cutting down trees and eliminating skewed intersections. Moreover, installing divided medians on urban multilane major roadways can reduce bicycle-vehicle crashes.

8. Limitations and Future Studies

Similar to previous bicycle safety studies, this study faced the difficulty of obtaining bicycle traffic volume. However, the study attempted to address this by collecting the "presence/absence of bicycle lanes" at the used sites to act as a proxy measure for bicycle traffic, which was not done in most previous studies. Another limitation was

regarding the AADT data which was not available for the minor approaches of the three-leg and four-leg unsignalized intersections since most of these approaches were located on non-state-maintained roads.

Further research could compare the findings of this study with other bicycle safety studies in other states to see how bicycle-vehicle crash predictions using the MARS technique would concur or differ. The comparison can also pinpoint any differences in the significant predictors of bicycle-vehicle crashes at both segments and intersections in the analyzed states. Another research venue is to compare the MARS technique with other modeling approaches, e.g., the random-parameter negative binomial model to see how close or far the predictions are.

CRediT authorship contribution statement

Niloufar Shirani-bidabadi: Data curation, Formal analysis, Investigation, Methodology, Writing - original draft. **Naveen Mallipaddi:** Data curation, Methodology, Writing - original draft. **Kirolos Haleem:** Conceptualization, Funding acquisition, Investigation, Project administration, Supervision, Validation, Writing - review & editing. **Michael Anderson:** Writing - review & editing.

Declaration of Conflict of Interest

The authors declare that they have no conflict of interest.

Acknowledgment

The authors would like to acknowledge the Alabama Department of Transportation (ALDOT) for the grant provided to conduct this research. The opinions, findings, and conclusions in this paper are those of the authors and not necessarily those of the State of Alabama Department of Transportation. The authors would also like to acknowledge the effort made by Ms. Brittany Wood in the data collection process.

References

Abdel-Aty, M., Haleem, K., 2011. Analyzing Angle Crashes at Unsignalized Intersections Using Machine Learning Techniques. Accident Analysis & Prevention 43, 461–470.
Abdel-Aty, M., Lee, J., Siddiqui, C., Choi, K., 2013. Geographical Unit Based Analysis in the Context of Transportation Safety Planning. Transportation Research Part A: Policy and Practice 49, 62–75.
Alluri, P., Raihan, M., Saha, D., Wu, W., Huq, A., Nafis, S., Gan, A., 2017. Statewide Analysis of Bicycle Crashes. Report BDV29-977-23. Florida Department of Transportation, Tallahassee, FL.
American Association of State Highways and Transportation Officials, 2010. Highway Safety Manual. "AASHTO". First Edition. Transportation Research Board of the National Academies, Washington, D.C.
American Community Survey Reports, 2018. American Community Survey Reports. Accessed July 12. https://www.census.gov/prod/2014pubs/acs-25.pdf.
Amoh-Gyimah, R., Saberia, M., Sarvi, M., 2016. Macroscopic Modeling of Pedestrian and Bicycle Crashes: A Cross-Comparison of Estimation Methods. Accident Analysis & Prevention 93, 147–159.
Bargegol, I., Gilani, V., Ghasedi, M., Ghorbanzadeh, M., 2016. Delay Modeling of Unsignalized Roundabouts Using Neural Network and Regression. Computational Research Progress in Applied Science & Engineering 2 (1), 28–34.
Bauer, K., Harwood, D., 2000. Statistical Models of at-Grade Intersections Crashes - Addendum. Report FHWA-RD-99-094. Federal Highway Administration (FHWA), McLean, VA.
Cai, Q., Abdel-Aty, M., Castro, S., 2020. Explore Effects of Bicycle Facilities and Exposure on Bicycle Safety at Intersections. International Journal of Sustainable Transportation 1–12.
Centre for Disease Control and Prevention, 2019. Motor Vehicle Safety. "CDC". Accessed May 12. https://www.cdc.gov/motorvehiclesafety/bicycle/index.html.
Chen, P., 2015. Built Environment Factors in Explaining the Automobile-Involved Bicycle Crash Frequencies: A Spatial Statistic Approach. Safety Science 79, 336–343.
Chen, C., Wang, H., Roll, J., Nordback, K., Wang, Y., 2020. Using Bicycle App Data to Develop Safety Performance Functions (SPFs) for Bicyclists at Intersections: A Generic Framework. Transportation Research Part A: Policy and Practice 132, 1034–1052.
Chimba, D., Emaasit, D., Cherry, C., Pannell, Z., 2014. In: Patterning Demographic and Socio-Economic Characteristics Affecting Pedestrian and Bicycle Crash Frequency. Proceedings of the 93th Annual Meeting of the Transportation Research Board, Washington, D.C.
Chin, H., Quddus, M., 2003. Applying the Random Effect Negative Binomial Model to Examine Traffic Accident Occurrence at Signalized Intersections. Accident Analysis & Prevention 35 (2), 253–259.
Conway, R., Maxwell, W., 1962. A Queuing Model with State Dependent Service Rates. Journal of Industrial Engineering 12, 132–136.
Dixon, K., Yi, X., Brown, L., Layton, R., 2013. Developing an Oregon Access Management Best Practices Manual. Oregon Department of Transportation, Salem, OR Federal Highway Administration (Washington, D.C.), & Oregon Transportation Research and Education Consortium (OTREC), Portland, OR.
Dolatsara, H., 2014. Development of Safety Performance Functions for Non-Motorized Traffic Safety. M.S. Thesis. Western Michigan University, Kalamazoo, MI.
Donnell, E., Gayah, V., Jovanis, P., 2014. Safety Performance Functions. Report FHWA-PA-2014-007-PSU WO 1. Pennsylvania Department of Transportation, Harrisburg, PA.
Gates, T., Savolainen, P., Stapleton, S., Kirsch, T., Miraskar, S., 2016. Development of Safety Performance Functions and Other Decision Support Tools to Assess Pedestrian and Bicycle Safety. Final Report. Transportation Research Center for Livable Communities (TRCLC), Kalamazoo, MI.
Gaweesh, S., Ahmed, M., Piccorelli, A., 2019. Developing Crash Prediction Models Using Parametric and Nonparametric Approaches for Rural Mountainous Freeways: A Case Study on Wyoming Interstate 80. Accident Analysis & Prevention 123, 176–189.
Geedipally, S., Lord, D., 2011. Examination of Crash Variances Estimated by Poisson–Gamma and Conway–Maxwell–Poisson Models. Transportation Research Record No. 2241. Transportation Research Board of the National Academies, Washington, D.C, pp. 59–67.
Google Earth Pro, 2018. Google Earth Pro Application. V. 7.3. Accessed February 14. https://www.google.com/earth/.
Google Maps, 2017. Google maps. Accessed June 19. http://www.maps.google.com/.
Guikema, S., Goffelt, J., 2008. A Flexible Count Data Regression Model for Risk Analysis. Risk Analysis: An International Journal 28 (1), 213–223.
Haleem, K., Abdel-Aty, M., 2010. Examining Traffic Crash Injury Severity at Unsignalized Intersections. Journal of Safety Research 41 (4), 347–357.
Hamann, C.J., Peek-Asa, C., 2013. On-Road Bicycle Facilities and Bicycle Crashes in Iowa, 2007–2010. Accident Analysis & Prevention 56, 103–109.
Harnen, S., Wong, S., Umar, R., Hashim, W., 2003. Motorcycle Crash Prediction Model for Non-Signalized Intersections. IATSS Research 27, 58–65.
Harwood, D., Bauer, K., Potts, I., Torbic, D., Richard, K., Rabbani, E., Hauer, E., Elefteriadou, L., Griffith, M., 2003. Safety Effectiveness of Intersection Left-and Right-Turn Lanes. Transportation Research Record No. 1840. Transportation Research Board of the National Academies, Washington, D.C, pp. 131–139.
Koloushani, M.R., Fatemi, A., Tabibi, M., 2014. In: Application of Global Positioning System Data Collected by Mobile Mapping System for Automatic Control of Safety Standards in Horizontal Curves. Proceedings of the 93th Annual Meeting of the Transportation Research Board, Washington, D.C.
Kononov, J., Bailey, B., Allery, B., 2008. In: Explanatory Examination of the Functional Form of Safety Performance Functions of Urban Freeways. Proceedings of the 87th Annual Meeting of the Transportation Research Board, Washington, D.C..
Kwigizile, V., Oh, J., Van Houten, R., Kwayu, K., 2016. Development of Decision Support Tools to Assess Pedestrian and Bicycle Safety: Development of Safety Performance Functions. Final Report. Transportation Research Center for Livable Communities (TRCLC), Kalamazoo, MI.
Li, X., Lord, D., Zhang, Y., Xie, Y., 2008. Predicting Motor Vehicle Crashes Using Support Vector Machine Models. Accident Analysis & Prevention 40, 1611–1618.
Lord, D., Geedipally, S., Guikema, S., 2010. Extension of the Application of Conway-Maxwell-PoissonModels: Analyzing Traffic Crash Data Exhibiting Underdispersion. Risk Analysis: An International Journal 30 (8), 1268–1276.
Lord, D., Guikema, S., 2012. The Conway–Maxwell–Poisson Model for Analyzing Crash Data. Applied Stochastic Models in Business and Industry 28 (2), 122–127.
Lord, D., Guikema, S., Geedipally, S., 2008. Application of the Conway–Maxwell–Poisson Generalized Linear Model for Analyzing Motor Vehicle Crashes. Accident Analysis & Prevention 40 (3), 1123–1134.
Lusk, A., Asgarzadeh, M., Farvid, M., 2015. Database Improvements for Motor Vehicle/Bicycle Crash Analysis. Injury Prevention 21, 221–230.
Maxwell, O., Mayowa, B., Chinedu, I., Peace, A., 2018. Modelling Count Data; A Generalized Linear Model Framework. American Journal of Mathematics and Statistics 8 (6), 179–183.
Mehta, G., Lou, Y., 2013. Safety Performance Function Calibration and Development for the State of Alabama: Two-Lane Two-Way Rural Roads and Four-Lane Divided Highways. Transportation Research Record No. 2398. Transportation Research Board of the National Academies, Washington, D.C, pp. 75–82.
Miranda-Moreno, L., Morency, P., El-Geneidy, A., 2011. The Link between Built Environment, Pedestrian Activity and Pedestrian–Vehicle Collision Occurrence at Signalized Intersections. Accident Analysis & Prevention 43 (5), 1624–1634.
Mukoko, K., Pulugurtha, S., 2020. Examining the Influence of Network, Land Use, and Demographic Characteristics to Estimate the Number of Bicycle-Vehicle Crashes on Urban Roads. IATSS Research 44, 8–16.
National Highway Traffic Safety Administration, 2019. "NHTSA". Accessed May 12. https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812382.
Nordback, K., Marshall, W., Janson, B., 2014. Bicyclist Safety Performance Functions for a U.S. City. Accident Analysis & Prevention 65, 114–122.
Oh, J., Washington, S., Nam, D., 2006. Accident Prediction Model for Railway-Highway Interfaces. Accident Analysis & Prevention 38 (2), 346–356.
Park, J., Abdel-Aty, M., Wang, L., Lee, G., Hong, J., 2019. Influence of Multiple Freeway Design Features on Freight Traffic Safety. Journal of Advanced Transportation, Article ID, 5739496 8 pages.
Pathivada, B., Perumal, V., 2019. Analyzing Dilemma Driver Behavior at Signalized

Intersection under Mixed Traffic Conditions. Transportation Research Part F: Traffic Psychology and Behaviour 60, 111–120.
Phillips, S., Carter, D., Hummer, J., Foyle, R., 2005. In: Safety Comparison of Four-Lane Median Divided and Five-Lane with TWLTL Segments. the 3rd International Symposium on Highway Geometric Designs, Chicago, IL. Paper presented at.
Pulugurtha, S., Thakur, V., 2015. Evaluating the Effectiveness of on-Street Bicycle Lane and Assessing Risk to Bicyclists in Charlotte, North Carolina. Accident Analysis & Prevention 76, 34–41.
Put, R., Xu, Q., Massart, D., Heyden, Y., 2004. Multivariate Adaptive Regression Splines (MARS) in Chromatographic Quantitative Structure-Rentention Relationship Studies. Journal of Chormatography 1055, 11–19.
Software, R., 2019. R. Software. Accessed April 10. http://www.r-project.org/.
Rista, E., Barrette, T., Savolainen, P., 2018. In: Safety Performance of Median U-Turns on Urban Boulevards. Proceedings of the 97th Annual Meeting of the Transportation Research Board, Washington, D.C.
Sam, E., Daniels, S., Brijs, K., Brijs, T., Wets, G., 2018. Modelling Public Bus/Minibus Transport Accident Severity in Ghana. Accident Analysis & Prevention 119, 114–121.
SAS OnDemand for Academics, 2017. SAS OnDemand for Academics. Accessed December 28. https://www.sas.com/en_us/software/on-demand-for-academics.html.
Sellers, K., Borle, S., Shmueli, G., 2012. The COM-Poisson Model for Count Data: A Survey of Methods and Applications. Applied Stochastic Models in Business and Industry 28 (2), 104–116.
Sellers, K., Shmueli, G., 2010. A Flexible Regression Model for Count Data. The Annals of Applied Statistic 4 (2), 943–961.
Shirazi, M., Lord, D., Geedipally, S., 2016. Sample-Size Guidleines for Recalibrating Crash Prediction Models: Recommendations for the Highway Safety Manual. Accident Analysis & Prevention 93, 160–168.
Siddiqui, C., Abdel-Aty, M., 2012. Nature of Modeling Boundary Pedestrian Crashes at Zones. Transportation Research Record No. 2299. Transportation Research Board of the National Academies, Washington, D.C, pp. 31–40.
Srinivasan, R., Carter, D., 2010. Development of Safety Performance Functions for North Carolina. Report FHWA/NC/2010-09. North Carolina Department of Transportation, Murphy, NC.
Tabibi, M., Koloushani, M., Fatemi, A., 2016. Introducing Automatic Control Method of Speed Limit Signs Safety Standards before Horizontal Curves via Mobile Mapping System. Amirkabir Journal of Civil Engineering 48 (3), 301–314.
Thomas, L., Lan, B., Sanders, R., Frackelton, A., Gardner, S., Hintze, M., 2017. In: In Pursuit of Safety: Systemic Bicycle Crash Analysis in Seattle, WA. Proceedings of the 96th Annual Meeting of the Transportation Research Board, Washington, D.C.
Ulak, M.B., Ozguven, E.E., Karabag, H.H., Ghorbanzadeh, M., Moses, R., Dulebenets, M., 2020. Development of Safety Performance Functions for Restricted Crossing U-Turn Intersections. Journal of Transportation Engineering, Part A: Systems 146 (6), 04020038.
Weir, M., Weintraub, J., Humphreys, E., Seto, E., Bhatia, R., 2009. Analyzing an Area-Level Model of Vehicle-Pedestrian Injury Collisions with Implications of Land Use and Transportation Planning. Accident Analysis & Prevention 41 (1), 137–145.
Wu, Q., Chen, F., Zhang, G., Liu, X., Wang, H., Bogus, S., 2014. Mixed Logit Model-Based Driver Injury Severity Investigations in Single-and Multi-Vehicle Crashes on Rural Two-Lane Highways. Accident Analysis & Prevention 72, 105–115.
Xie, Y., Chen, C., 2016. Calibration of Safety Performance Functions for Massachusetts Urban and Suburban Intersections. Massachusetts Department of Transportation, Boston, MA.
Yao, S., Loo, B., 2016. Safety in Numbers for Cyclists beyond National-Level and City-Level Data: A Study on the Non-Linearity of Risk within the City of Hong Kong. Injury Prevention 22 (6), 379–385.
Yue, L., Abdel-Aty, M., Lee, J., Farid, A., 2019. Effects of Signalization at Rural Intersections Considering the Elderly Driving Population. Transportation Research Record No. 2673. Transportation Research Board of the National Academies, Washington, D.C, pp. 743–757.
Zangenehpour, S., Miranda-Moreno, L., Saunier, N., 2015. Automated Classification Based on Video Data at Intersections with Heavy Pedestrian and Bicycle Traffic: Methodology and Application. Transportation Research Part C: Emerging Technologies 56, 161–176.