

## Research Article

# Predicting Wet-Road Crashes Using the Finite-Mixture Zero-Truncated Negative Binomial Model

Ying Chen<sup>1</sup> and Zhongxiang Huang<sup>2</sup>

<sup>1</sup>*School of Traffic and Transportation Engineering, School of Architecture, Changsha University of Science & Technology, Changsha 410114, China*

<sup>2</sup>*School of Traffic and Transportation Engineering, Changsha University of Science & Technology, Changsha 410114, China*

Correspondence should be addressed to Ying Chen; [cycsust@163.com](mailto:cycsust@163.com)

Received 28 July 2020; Revised 23 September 2020; Accepted 28 September 2020; Published 21 October 2020

Academic Editor: Yajie Zou

Copyright © 2020 Ying Chen and Zhongxiang Huang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Inclement weather affects traffic safety in various ways. Crashes on rainy days not only cause fatalities and injuries but also significantly increase travel time. Accurately predicting crash risk under inclement weather conditions is helpful and informative to both roadway agencies and roadway users. Safety researchers have proposed various analytic methods to predict crashes. However, most of them require complete roadway inventory, traffic, and crash data. Data incompleteness is a challenge in many developing countries. It is common that safety researchers only have access to data on sites where a crash has occurred (i.e., zero-truncated data). The conventional crash models are not applicable to zero-truncated safety data. This paper proposes a finite-mixture zero-truncated negative binomial (FMZTNB) model structure. The model is applied to three-year wet-road crash data on 395 divided roadway segments (total 586 km), and the parameters are estimated using the Markov chain Monte Carlo (MCMC) method. Comparison indicates that the proposed FMZTNB model has better fitting performance and is more accurate in predicting the number of wet-road crashes. The model is capable of capturing the heterogeneity within the sample crash data. In addition, lane width showed mixed effects in different components on wet-road crashes, which are not observed in conventional modeling approaches. Practitioners are encouraged to consider the finite-mixture zero-truncated modeling approach when complete safety dataset is not available.

## 1. Introduction

According to the World Health Organization, more than 1.3 million roadway users died each year as a result of traffic crashes, and the cost of traffic crashes accounted for about 3% of the gross domestic product in most countries [1]. Road traffic injuries and deaths are a global problem, and traffic crashes are a leading cause for nonnatural death. Safety researchers and practitioners have made continuous efforts to reduce the number and severity of crashes.

A traffic crash is usually caused by one or several factors, including humans (i.e., vehicle driver, motorist, bicyclist, and pedestrians), vehicles, roadway facilities, and environment. Weather affects traffic safety, demand, selection of transportation mode, driving capabilities, vehicle

performance (i.e., stability, maneuverability, and traction), and roadway infrastructure (i.e., pavement friction) through visibility impairments, precipitation, and temperature. Inclement weather not only increases crash risk but also significantly affects users' travel time. According to crash statistics, more than 20 percent of crashes and more than 15 percent of traffic fatalities are weather-related [2]. It is necessary to accurately predict the occurrence of crashes under inclement weather conditions.

Statistical modeling approach has been extensively used in recent two to three decades to quantitatively predict number of crashes. Specifically, safety researchers have proposed various models for developing crash counts, e.g., Poisson, negative binomial (NB), Sichel [3, 4], Conway–Maxwell–Poisson [5, 6], zero-inflated Poisson [7],

Poisson–Tweedie [8], Tobit [9–12], machine learning techniques [13, 14], etc. For a detailed review of the crash regression techniques, readers can refer to the article by Lord and Mannering [15]. In all these models, crash counts are treated as response variables, and more importantly, all of these models require complete roadway inventory, traffic, and crash data. Safety data (e.g., roadway inventory, traffic, operation, and crash information) play a critical role in crash prediction model development, hotspot identification, and safety effectiveness evaluation. Inaccurate or incomplete crash records with the conventional crash prediction models may lead to various misleading results. These errors not only result in inefficient use of limited resources for safety improvements but also cause additional loss of lives. However, data incompleteness is a challenge in many developing countries. For example, only information on segments or intersections where a crash has occurred is collected. This type of data is known as zero-truncated data. Previous studies have shown that the conventional count models are not adequate to model zero-truncated crash data [16, 17]. How to develop reliable crash prediction models using zero-truncated data is an important topic for safety analysts.

This study is an extension of a recent study on modeling zero-truncated crash data [16]. The primary objective is to develop safety performance functions for wet-road crashes when zero's are truncated in the safety data considering the heterogeneity. Particularly, this paper proposes a finite-mixture zero-truncated negative binomial (FMZTNB) model structure and examines if the FMZTNB model provides better modeling results than the commonly used models.

The rest of this paper is organized as follows: Section 2 reviews the literature pertaining to the influence of weather on safety. Section 3 describes the details of the zero-truncated models. Section 4 briefly documents the zero-truncated data. Section 5 presents the modeling results, and Section 6 summarizes the study.

## 2. Literature Review

Because of the great influence of weather on roadway safety, transportation researchers have made continuous efforts in understanding the relationship between different weather conditions and traffic crashes.

Shankar et al. [18] conducted one of the earliest studies on the effect of weather conditions on roadway crashes. The researchers developed a negative binomial crash model with roadway geometrical and environmental factors. The modeling results suggest that both maximum rainfall and number of rainy days play significant and positive role in number of total crashes.

Maze et al. [19] studied how inclement weather affects traffic demand, traffic safety, and traffic flow relationships. The researchers pointed out that certain types of severe weather conditions (e.g., winter storms) bring a higher risk of being involved in a crash by 13 to 25 times. Weather conditions also impact the crash severity, but it varies depending on specific weather condition and crash location.

Qiu and Nixon [20] conducted a systemic review on the effect of adverse weather on the occurrence of roadway crashes. The researchers reviewed 112 studies conducted between 1967 and 2005 that had examined the association between weather and traffic crashes. Crash rates from each study were combined through a meta-analysis method. The researchers conclude that the crash rate usually increases during precipitation. Snow has a greater effect than rain does on crash occurrence. Specifically, snow can increase the crash rate by 84% and the injury rate by 75%.

Jung et al. [21] analyzed the influence of four weather factors (i.e., rainfall intensity, water film depth, temperature, and wind speed and direction) on the injury severity of rainy day multivehicle crashes. The study found that wind speed is associated with the outcome of crashes.

Recently, Das et al. [22] developed safety performance functions for two types of roadways (i.e., rural two-lane highway and rural multilane highway) in two states (i.e., Ohio and Washington). The researchers included speed measures and weather conditions in the models. Modeling results revealed that precipitation is negatively associated with number of crashes. This result is inconsistent with most previous studies, and the researchers noted that the vehicle speeds might reduce during the wet-weather conditions, hence resulting in fewer crashes.

To summarize, extensive studies have been conducted to analyze the relationship between weather and safety. Overall, crash rates increase significantly during inclement weather conditions. In the previous studies, almost all of them include weather data as factors in the regression models, and none of them have focused on developing a safety performance function for wet-weather crashes specifically. In addition, previous studies have used the common count models (e.g., negative binomial), which require complete safety data. Zero-truncated data are common in developing countries, and zero-truncated models have been proposed by researchers to analyze crash data in recent years [16, 17]. To the best of the authors' knowledge, no efforts have been made to analyze zero-truncated wet-road crashes. This study aims to fill this gap.

## 3. Methodology

This section discusses three crash modeling approaches: (1) the commonly used negative binomial model; (2) zero-truncated NB model; and (3) finite-mixture zero-truncated NB model.

**3.1. Conventional NB Model.** As has been mentioned in Section 2, various statistical methods have been developed by safety researchers to predict number of crashes. The NB model is still the most commonly used approach and is recommended by the first edition of *Highway Safety Manual (HSM)* [23]. This section briefly introduces the structure of the NB model.

The commonly used NB model assumes that the number of crashes occurred at a given site (a segment or an

intersection) during a certain period follows Poisson distribution as follows:

$$y_{i,t} | \lambda_{i,t} \sim \text{Poisson}(\lambda_{i,t}). \quad (1)$$

The probability mass function (PMF) of crash count is shown as follows:

$$P(y_{i,t} | \lambda_{i,t}) = \frac{\lambda_{i,t}^{y_{i,t}} \times e^{-\lambda_{i,t}}}{y_{i,t}!}, \quad y_{i,t} = 0, 1, 2, \dots, \quad (2)$$

where  $y$  denotes the crash count. The subscripts  $i$  and  $t$  represent site index and study period, respectively.  $\lambda_{i,t}$  is the Poisson rate for the site during the period. For the ease of readers, the subscripts  $i$  and  $t$  are omitted in the rest of this paper.

Furthermore, assume that the Poisson rate  $\lambda$  follows gamma distribution:

$$\lambda | \mu, \alpha \sim \text{Gamma}(\mu, \alpha), \quad (3)$$

where  $\mu$  is the mean for  $\lambda$  and  $\alpha$  is the shape parameter (positive).

Assuming that the mean  $\mu$  is associated with roadway features (e.g., traffic volume, segment length, and geometric characteristics),

$$\mu = f(\text{traffic volume, roadway features}). \quad (4)$$

Interpreting  $\lambda$  from equation (2), the PMF of the NB distribution can be obtained as

$$p(y | \mu, \alpha) = \int_0^\infty \frac{\lambda^y \times \exp(-\lambda)}{y!} \frac{1}{\mu^\alpha} \times \frac{\lambda^{\alpha-1} \times \exp(-\lambda/\mu)}{\Gamma(\alpha)} d\lambda. \quad (5)$$

The PMF of  $y$  is shown as follows:

$$p(y | \mu, \alpha) = \frac{\Gamma(y + \alpha)}{\Gamma(y + 1) \times \Gamma(\alpha)} \times \left( \frac{\mu/\alpha}{1 + (\mu/\alpha)} \right)^y \times \left( \frac{1}{1 + (\mu/\alpha)} \right)^\alpha, \quad (6)$$

where  $y$  is the response variable (i.e., crash count),  $\mu$  indicates the mean response of the observation, and  $\alpha$  is the

dispersion parameter (i.e., shape parameter in the Gamma distribution). For the detailed derivative of the NB model, readers can refer to [24]. It is important to note that, in the conventional NB model structure, the response variable  $y$  takes the values of all nonnegative counts (i.e., 0, 1, 2, 3, ...). In other words, all the observed crash counts should be included in the model development. Since the NB model has closed-form, the parameters can be easily estimated. Many software packages have been developed to estimate the unknown parameters, for example, the MASS package of R [25, 26].

**3.2. Zero-Truncated NB Model.** The NB model has been widely used in analyzing overdispersed count data; however, it requires completed observed data. When the zero's are truncated, the assumption of the NB model cannot be satisfied, and the estimated parameters are biased. Statisticians proposed truncated models [27]. In the truncated count model, the response variable,  $y$ , is also considered to follow Poisson distribution. But, it only takes positive numbers (i.e., conditional on that  $y > 0$ ) as follows:

$$P(y | \lambda, y > 0) = \frac{P(y)}{P(y > 0)} = \frac{P(y)}{1 - P(y = 0)}, \quad y = 1, 2, 3, \dots \quad (7)$$

From equation (2), it can be derived that

$$P(y = 0 | \lambda) = \frac{\lambda^0 \times e^{-\lambda}}{0!} = e^{-\lambda}. \quad (8)$$

Substituting equation (8) into (7), the zero-truncated Poisson distribution can be obtained as follows:

$$P(y | \lambda, y > 0) = \frac{P(y)}{1 - e^{-\lambda}}, \quad y = 1, 2, 3, \dots, \quad (9)$$

where  $y$  is the response variable (truncated) and  $\lambda$  is the Poisson rate. Similarly, assuming that the Poisson rate  $\lambda$  follows Gamma distribution, the zero-truncated NB model can be obtained as follows:

$$P(y = 0 | \mu, \alpha) = \frac{\Gamma(y + \alpha)}{\Gamma(\alpha) \times \Gamma(y + 1)} \times \left( \frac{\mu/\alpha}{1 + (\mu/\alpha)} \right)^y \times \left( \frac{1}{1 + (\mu/\alpha)} \right)^\alpha = \left( \frac{1}{1 + (\mu/\alpha)} \right)^\alpha, \quad (10)$$

$$P(y | \mu, \alpha, y > 0) = \frac{\Gamma(y + \alpha)}{\Gamma(\alpha) \times \Gamma(y + 1)} \times \left( \frac{\mu/\alpha}{1 + (\mu/\alpha)} \right)^y \times \frac{(1/(1 + (\mu/\alpha)))^\alpha}{(1 - (1/(1 + (\mu/\alpha))))^\alpha}, \quad (11)$$

where  $\mu$  is the mean response of the observation and  $\alpha$  is the dispersion parameter.

Compared to the conventional NB model, the zero-truncated NB model can be viewed as a conditional NB distribution that the response variable takes nonzero values. The conditional distribution (i.e., positive NB) brings complexity in estimating parameters. A few

software packages are available for estimating the ZTNB model, for example, VGAM with R [28].

**3.3. Finite-Mixture Zero-Truncated NB Model.** In both the conventional NB and zero-truncated NB models, the distribution of the response variable has only one component,

i.e., there is only one Poisson mean. The finite-mixture models assume that the response variables arise from two or more unobserved components with unknown proportions. This provides significant modeling flexibility than the conventional single component models [29]. As has been mentioned, statisticians have proposed the  $K$ -component finite mixture of negative binomial regression models (i.e., FMNB- $K$ ) as follows [29, 30]:

$$p(y) = \sum_{k=1}^K w_k \left[ \frac{\Gamma(y + \alpha_k)}{\Gamma(y + 1)\Gamma(\alpha_k)} \left( \frac{\mu_k}{\mu_k + \alpha_k} \right)^y \left( \frac{\alpha_k}{\mu_k + \alpha_k} \right)^{\alpha_k} \right], \quad (12)$$

$$p(y) = \sum_{k=1}^K w_k \left[ \frac{\Gamma(y + \alpha_k)}{\Gamma(\alpha_k) \times \Gamma(y + 1)} \times \left( \frac{\mu_k/\alpha_k}{1 + \mu_k/\alpha_k} \right)^y \times \frac{(1/(1 + (\mu_k/\alpha_k)))^{\alpha_k}}{(1 - (1/(1 + \mu/\alpha)))^\alpha} \right], \quad (13)$$

where  $y$  is the zero-truncated response variable (i.e., crash counts;  $y = 1, 2, 3, 4, \dots$ );  $w_k$  is the weight factor of component  $k$  which sum to 1 ( $\sum_{k=1}^K w_k = 1$ );  $\mu_k$  is the Poisson mean of component  $k$ ; and  $\alpha_k$  is the dispersion parameter of component  $k$ .

In both the FMNB- $K$  and FMZTNB- $K$  models, a function is used to link the Poisson mean and roadway features; therefore,

$$\mu = \sum_{k=1}^K f_k(\text{traffic volume, roadway features}). \quad (14)$$

It can be seen that when  $K=1$ , the FMNB- $K$  and FMZTNB- $K$  models reduce to NB and ZTNB models, respectively. The FMNB models allow for additional heterogeneity within components not captured by the independent variables.

It is important to note that, as the number of components  $K$  increases, the FMNB model becomes more flexible. However, it also brings complexity in the parameter estimation. Previous studies have indicated that a two-component finite mixture of NB regression models (FMNB-2) was quite enough to characterize crash data [31–33]. Thus, this study considers the two-component finite mixture of zero-truncated NB model (FMZTNB-2) in the analyses.

In terms of parameter estimation, the commonly used maximum likelihood estimation (MLE) algorithm will not generate reliable results due to the complicated likelihood function in the FMZTNB-2 model. An alternative is the Gibbs sampling technique, also known as the Markov chain Monte Carlo (MCMC) method, which has been frequently used in estimating parameters of finite-mixture models [29, 34]. Package “rjags” is used to draw the samplings [35], and the FMNZTB-2 MCMC model is developed using JAGS (Just Another Gibbs Sampler) [36]. The truncation is represented using function  $T(\cdot)$  in the JAGS.

where  $y$  is the response variable ( $y = 0, 1, 2, 3, 4, \dots$ );  $w_k$  is the weight factor of component  $k$  which sum to 1 ( $\sum_{k=1}^K w_k = 1$ );  $\mu_k$  is the Poisson mean of component  $k$ ; and  $\alpha_k$  is the dispersion parameter of component  $k$ .

Analogous to equation (12), the  $K$ -component finite mixture of zero-truncated NB model (FMZTNB- $K$ ) can be constructed as

## 4. Data

This study collected data on 395 rural multilane-divided roadway segments, including traffic volume, lane width, average shoulder width, and median width. Three years of wet-road crash data were collected. A wet-road crash is defined as that the weather condition was rain, snow, or hail, or the surface condition was wet, snowy, ice, or standing water at the time of the crash occurred. In terms of independent variables, this paper mainly considered data availability and potential effects on the occurrence of crashes during rainy weather conditions from published literature [19–21]. Finally, the following six variables were selected from the dataset: segment length, traffic volume, lane width, average outside shoulder width, average inside shoulder width, and median width. Descriptive statistics of the roadway and crash data are illustrated in Table 1.

It is worth mentioning that the minimum crash count of the sample segments is 1 (see the last row in Table 1), rather than 0. This is because when collecting the roadway data, only information on segments where at least one crash had occurred is available to the authors. In other words, the safety data is zero-truncated.

## 5. Modeling Results

Previous studies have revealed that the commonly used NB model is not applicable for modeling zero-truncated crash data [16, 17]. The parameters can be heavily biased, and the results are not reliable. Thus, the conventional NB model is not used to the data collected in this study. This section presents the results of the ZTNB model and the FMZTNB-2 model, separately.

**5.1. Modeling Result of ZTNB.** The authors developed the ZTNB model with the data described in Section 4 with the following functional form.

TABLE 1: Descriptive statistics of data (sample size: 395).

Variable	Mean	Minimum	Maximum	Standard deviation
Segment length (km)	1.483	0.172	3.080	0.969
Traffic volume (veh/day)	11569.41	3192	26935	5505.763
Lane width (m)	3.826	2.9	5.0	0.513
Average outside shoulder width (m)	2.515	0.61	3.05	0.758
Average inside shoulder width (m)	1.536	0	3.05	0.705
Median width (m)	6.352	3.05	14.64	3.107
Wet-road crash count	2.42	1	15	2.439

$$\mu = \exp(\beta_0 + \beta_1 \log(\text{ADT}) + \beta_2 \text{LW} + \beta_3 \text{OSH} + \beta_4 \text{ISH} + \beta_5 \times \text{MW}), \quad (15)$$

where  $\mu$  is the mean of the observed crash data; ADT is traffic volume; LW is lane width (m); OSH is average outside shoulder width; ISH is average inside shoulder width (m); MW is median width (m);  $\beta_0, \beta_1, \dots, \beta_5$  are unknown parameters to be estimated. It is important to note that the length of a segment is considered as an offset variable, meaning that the number of crashes is proportional to the segment length. This assumption is consistent with the *HSM*.

Although studies have pointed out that varying dispersion parameter (i.e.,  $\alpha$  in equations (10) and (11)) benefits the crash prediction models [37–39], this study assumed that it is fixed among all the sites to make the computation easier and consistent with the FMZTNB-2 model in Section 5.2 [37, 40–42].

The modeling results of the ZTNB model is shown in Table 2. As can be seen, the parameters for traffic volume, average outside shoulder width, and average inside shoulder width are all statistically significant at the level of 90 percent or higher. Specifically, as the traffic volume increase, the predicted number of wet-road crashes also increases. The parameters for the other three roadway features are all negative, indicating that, with the increase of shoulder width or median width, the predicted number of wet-road crashes will decrease. For example, with one meter increase in average outside shoulder width, the predicted number of wet-roadway crashes will decrease by 14.6 percent (i.e.,  $1 - e^{-0.158}$ ). This is expected, as outside shoulders become wider, it provides additional recovery spaces for vehicles which slide away from the traveling lane due to the reduced skid number during rainy days. The results are in line with several previous studies [21, 43]. On the other hand, the parameter for lane width is  $-0.1$ , and the result is not statistically significant. The dispersion parameter,  $\alpha$ , is estimated as 1.615, which is also insignificant.

This study used four types of goodness-of-measure (GOF) to evaluate the model performance: Akaike

TABLE 2: Estimating results of the ZTNB model.

Variable	Estimate	Std. err.	P value	Significant level
Intercept, $\beta_0$	-3.638	1.698	—	Not significant
Log (ADT), $\beta_1$	<b>0.770</b>	0.169	<0.001	99.9%
Lane width, $\beta_2$	-0.100	0.159	0.530	Not significant
Ave. out. SHD, $\beta_3$	<b>-0.158</b>	0.082	0.054	90.0%
Ave. in. SHD, $\beta_4$	<b>-0.310</b>	0.080	<0.001	99.9%
Median width, $\beta_5$	<b>-0.373</b>	0.034	<0.001	99.9%
Disp. par., $\alpha$	1.615	5.465	0.409	Not significant
AIC	1142.29	—	—	—
BIC	1170.14	—	—	—
MAE	0.64	—	—	—
RMSE	2.52	—	—	—

Note. Std. err. = standard error. Disp. par. = dispersion parameter. Bold and underline values indicate statistically significant at 90.0% or higher. — means not applicable.

information criterion (AIC), Bayesian information criterion (BIC), mean absolute error (MAE), and root mean square error (RMSE). The AIC, BIC, MAE, and RMSE for the ZTNB model are 1142.29, 1170.14, 0.64, and 2.52, respectively (see the last four rows in Table 2).

**5.2. Modeling Result of FMZTNB-2.** As has been mentioned in Section 3, this study utilized MCMC approach to estimate the parameters of the FMZTNB-2 model. Noninformative priors were used for hyperparameters. This study performed 1,000,000 MCMC iterations with two different chains, and the first 20,000 samples of each chain were discarded as burn-in samples from the MCMC outputs. Gelman–Rubin (G–R) convergence statistics and visual history plots were used to verify the MCMC process [44, 45]. The functional forms linking the Poisson mean and the roadway features are similar to those of the ZTNB model, except that there are two forms in the components, as shown in the following equations.

$$\mu_{c1} = \exp(\beta_{0,c1} + \beta_{1,c1} \log(\text{ADT}) + \beta_{2,c1} \text{LW} + \beta_{3,c1} \text{OSH} + \beta_{4,c1} \text{ISH} + \beta_{5,c1} \text{MW}), \quad (16)$$

$$\mu_{c2} = \exp(\beta_{0,c2} + \beta_{1,c2} \log(\text{ADT}) + \beta_{2,c2} \text{LW} + \beta_{3,c2} \text{OSH} + \beta_{4,c2} \text{ISH} + \beta_{5,c2} \text{MW}), \quad (17)$$

$$\mu = wt \times \mu_{c1} + (1 - wt) \times \mu_{c2}, \quad (18)$$



TABLE 3: Estimating results of the FMZTNB-2 model.

Variable	Estimate	Component 1		Estimate	Component 2	
		Std. err.	P value		Std. err.	P value
Intercept	-4.047	2.819	0.151	<b>-4.485</b>	<b>1.190</b>	<0.001
Log (ADT)	<b>0.718</b>	<b>0.018</b>	<0.001	<b>0.993</b>	<b>0.195</b>	<0.001
Lane width	<b>-0.080</b>	<b>0.010</b>	<0.001	<b>0.203</b>	<b>0.012</b>	<0.001
Ave. Out. SHD	<b>-0.094</b>	<b>0.022</b>	<0.001	<b>-0.183</b>	<b>0.014</b>	<0.001
Ave. In. SHD	-0.241	0.177	0.174	<b>-0.879</b>	<b>0.166</b>	<0.001
Median width	<b>-0.416</b>	<b>0.010</b>	<0.001	-0.494	0.251	0.049
Disp. par.	<b>1.499</b>	<b>0.142</b>	<0.001	4.619	1.856	0.013
Weight	<b>0.712</b>	<b>0.082</b>	<0.001	<b>0.288</b>	-	-
AIC	1020.54	-	-	-	-	-
BIC	1088.32	-	-	-	-	-
MAE	0.22	-	-	-	-	-
RMSE	2.18	-	-	-	-	-

Note. Std. err. = standard error. Disp. Par. = dispersion parameter. Bold and italic values indicate statistically significant at level of 90.0 percent or higher. — means not applicable.

TABLE 4: Prediction comparison between ZTNB and FMZTNB models (three example sites).

Site number (level)	ZTNB			FMZTNB-2		
	Prediction	Std. err.	90% PI	Prediction	Std. err.	90% PI
138 (low)	0.0040	0.0636	[0.0036–0.0046]	0.0005	0.0007	[0.0005–0.0005]
90 (moderate)	0.0654	0.2608	[0.0392–0.109]	0.0627	0.0449	[0.0574–0.0684]
65 (high)	0.7350	1.0341	[0.0968–5.5788]	0.7625	0.5903	[0.2398–2.425]

Note. Std. err. = standard error. PI = prediction interval.

where  $\mu$  is the mean of the observed crash data;  $\mu_{c1}$  and  $\mu_{c2}$  are the mean of observations in the two components, respectively; ADT is traffic volume; LW is lane width (m); OSH is average outside shoulder width; ISH is average inside shoulder width (m); MW is median width (m); and  $\beta'$ s are parameters to be estimated.

The modeling results of the FMZTNB-2 model are documented in Table 3. First, the estimated weight factor for component 1 is 0.712, with a standard error of 0.082. This result is statistically significant, indicating that the sample data include two components. Component 1 accounts for about 71.2 percent, and component 2 accounts for about 28.8 percent (i.e.,  $1 - 0.712$ ). Second, most of the parameters in both components are statistically significant (except average inside shoulder width in component 1 and median width in component 2). Overall, the signs of the parameters in the FMZTNB-2 model are the same as the corresponding parameters in the ZTNB model. For example, the parameter for average outside shoulder width is  $-0.158$  in the ZTNB model. They are  $-0.094$  and  $-0.183$  in the two components, respectively, of the FMZTNB-2 model. All of them indicate that wider outside shoulders are associated with fewer wet-road crashes. The parameter for average inside shoulder in component 1 and that for median width in component 2 is not significant. Another interesting finding is that the estimated parameters for lane width in the two components have different signs, i.e., it is negative ( $-0.080$ ) in component 1 and positive ( $0.203$ ) in component 2. In other words, the lane width is negatively associated with wet-road crashes in the first group of

segments (i.e., component 1); however, it is positively associated with wet-road crashes in the second group of segments (i.e., component 2). This is in line with a few studies which have reported controversy effects of lane width on safety [46–48].

Finally, the AIC, BIC, MAE, and RMSE for the FMZTNB-2 model are 1020.54, 1088.32, 0.22 and 2.14, respectively (see the last four rows in Table 3). In addition to model goodness-of-fit, this paper also analyzed the prediction performance of the two models using three sites. The three sites represent relative low, moderate, and high crash levels, respectively. The crash mean prediction, standard deviation, as well as 90 percentile confidence intervals of the three sites by the two models are tabulated in Table 4. The results indicate that, for the three sites, the predicted crash mean (i.e., number of wet-weather crashes) between the two models are fairly close (except for the first site, which has a very small crash mean). For site 90, the predicted number of crashes of the ZTNB and FMZTNB-2 models are 0.0645 and 0.0627, respectively. Their standard deviation values are 0.2606 and 0.0449, respectively. The crash predictions with FMZTNB-2 model have significantly lower standard deviation values and narrower intervals, indicating that the model has higher prediction accuracy.

The FMZTNB-2 model shows superiority in modeling the wet-weather crash data. First, the FMZTNB-2 model fits the dataset better than the ZTNB model in terms of GOF measures (e.g., AIC, BIC, MAE, and RMSE). Second, the predictions using FMZTNB-2 model have lower standard deviations and narrower prediction intervals,

indicating that the predictions are more accurate. Finally, a few interesting relationships between variables and crashes are observed from the FMZTNB-2 model. For example, the parameters of lane width are opposite in the two components, indicating that this factor have mixed effects at different locations. These results indicate that the FMZTNB-2 model captures the heterogeneity of the crash data better than the ZTNB model.

## 6. Conclusions

Inclement weather increases both crash risk and travel time. Efforts have been made in the past decades to predict the occurrence of traffic crashes. However, very few of the previous studies have focused on predicting wet-road crashes. Most of the commonly used crash prediction models require complete roadway inventory, traffic, and crash data. Data missing is relative common in developing countries. How to analyze zero-truncated crash data and predict the number of wet-road crashes is the primary objective of this study. To better capture the heterogeneity of wet-road crash data, this study developed the two-component finite-mixture zero-truncated negative binomial model. The model is applied to three-year wet-road crash on 395 rural-divided roadways. The model results are compared with those based on zero-truncated negative binomial model. Comparison indicates that the proposed FMZTNB-2 model fits the wet-road crash data better than the ZTNB model. It is worth mentioning that, the wet-weather crash data were not modeled with the conventional NB model since previous studies have demonstrated that the application of NB model on truncated data is not recommended. There are trade-offs of using ZTNB or FMZTNB models in crash analyses. With zero-truncated data, the sample size is smaller than that of full data. The reduced sample size might increase uncertainty of parameter estimates.

There are some limitations with this study. First, only a number of roadway characteristics (i.e., segment length, lane width, shoulder width, and median width) and traffic data are available to the authors. There are other factors affecting the occurrence of wet-roadway crashes (e.g., precipitation, number of rainy days per year, and surface skid number). Unfortunately, they are not accessible to the authors. Second, previous studies have shown that the varying forms of dispersion parameter and weight factor for the components in the finite-mixture models improve both crash prediction and hotspot identification [33, 37, 49–51]. In this study, fixed dispersion parameter and weight factor were used to simplify the parameter estimation process. In the future, it is necessary to collect more data, especially those closed related to wet-road crashes, and to examine if varying forms of dispersion parameter and weight factor will further improve the model performance. Finally, the finite-mixture model provides better results than the previously proposed zero-truncated model (e.g., goodness-of-fit and prediction). However, parameter estimates with the FMZTNB-2 model require MCMC, and it increases the computational

time, which may be challenging for practitioners. The parameter estimating method in the FMZTNB-2 model needs to be further simplified in the future.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This research was sponsored jointly by the National Natural Science Foundation of China (project no. 51978082); the Outstanding Youth Foundation of Hunan Education Department (project no. 19B022); and the Young Teacher Development Foundation of Changsha University of Science & Technology (project no. 2019QJCZ056).

## References

- [1] World Health Organization, *Global Status Report on Road Safety 2018: Summary*, World Health Organization, Geneva, Switzerland, 2018.
- [2] FHWA, *How Do Weather Events Impact Roads?*, FHWA, Washington, DC, USA, 2020, [https://ops.fhwa.dot.gov/weather/q1\\_roadimpact.htm](https://ops.fhwa.dot.gov/weather/q1_roadimpact.htm).
- [3] L. Wu, Y. Zou, and D. Lord, "Comparison of sichel and negative binomial models in hot spot identification," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2460, no. 1, pp. 107–116, 2014.
- [4] Y. Zou, D. Lord, Y. Zhang, and Y. Peng, "Comparison of sichel and negative binomial models in estimating empirical bayes estimates," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2392, no. 1, pp. 11–21, 2013.
- [5] D. Lord, S. R. Geedipally, and S. D. Guikema, "Extension of the application of conway-maxwell-Poisson models: analyzing traffic crash data exhibiting underdispersion," *Risk Analysis*, vol. 30, no. 8, pp. 1268–1276, 2010.
- [6] D. Lord, S. D. Guikema, and S. R. Geedipally, "Application of the conway-maxwell-Poisson generalized linear model for analyzing motor vehicle crashes," *Accident Analysis & Prevention*, vol. 40, no. 3, pp. 1123–1134, 2008.
- [7] D. Lord, S. P. Washington, and J. N. Ivan, "Poisson, poisson-gamma and zero-inflated regression models of motor vehicle crashes: balancing statistical fit and theory," *Accident Analysis and Prevention*, vol. 37, no. 1, pp. 35–46, 2005.
- [8] D. Saha, P. Alluri, E. Dumbaugh, and A. Gan, "Application of the Poisson-tweedie distribution in analyzing crash frequency data," *Accident Analysis & Prevention*, vol. 137, Article ID 105456, 2020.
- [9] Y. Guo, Z. Li, P. Liu, and Y. Wu, "Modeling correlation and heterogeneity in crash rates by collision types using full bayesian random parameters multivariate tobit model," *Accident Analysis & Prevention*, vol. 128, pp. 164–174, 2019.
- [10] P. C. Anastasopoulos, F. L. Mannering, V. N. Shankar, and J. E. Haddock, "A study of factors affecting highway accident

- rates using the random-parameters Tobit model," *Accident Analysis & Prevention*, vol. 45, pp. 628–633, 2012.
- [11] Y. Guo, T. Sayed, and M. Essa, "Real-time conflict-based bayesian Tobit models for safety evaluation of signalized intersections," *Accident Analysis & Prevention*, vol. 144, Article ID 105660, 2020.
  - [12] Y. Guo, Z. Li, and T. Sayed, "Analysis of crash rates at freeway diverge areas using bayesian tobit modeling framework," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2673, no. 4, pp. 652–662, 2019.
  - [13] S. Das, L. Minjares-Kyle, L. Wu, and R. H. Henk, "Understanding crash potential associated with teen driving: survey analysis using multivariate graphical method," *Journal of Safety Research*, vol. 70, pp. 213–222, 2019.
  - [14] X. Yang, Y. Zou, J. Tang, J. Liang, and M. Ijaz, "Evaluation of short-term freeway speed prediction based on periodic analysis using statistical models and machine learning models," *Journal of Advanced Transportation*, vol. 2020, Article ID 9628957, 16 pages, 2020.
  - [15] D. Lord and F. Mannering, "The statistical analysis of crash-frequency data: a review and assessment of methodological alternatives," *Transportation Research Part A: Policy and Practice*, vol. 44, no. 5, pp. 291–305, 2010.
  - [16] Y. Chen and Z. Huang, "Modeling intersection traffic crashes using a zero-truncated negative binomial model," *China Journal of Highway and Transport*, vol. 36, no. 15, pp. 138–145, 2020.
  - [17] D.-G. Kim and Y. Lee, "Modelling crash frequencies at signalized intersections with a truncated count data model," *International Journal of Urban Sciences*, vol. 17, no. 1, pp. 85–94, 2013.
  - [18] V. Shankar, F. Mannering, and W. Barfield, "Effect of roadway geometrics and environmental factors on rural freeway accident frequencies," *Accident Analysis & Prevention*, vol. 27, no. 3, pp. 371–389, 1995.
  - [19] T. H. Maze, M. Agarwal, and G. Burchett, "Whether weather matters to traffic demand, traffic safety, and traffic operations and flow," *Transportation Research Record: Journal of Transportation Research Board*, vol. 1948, no. 1, pp. 170–176, 2006.
  - [20] L. Qiu and W. A. Nixon, "Effects of adverse weather on traffic crashes: systematic review and meta-analysis," *Transportation Research Record*, vol. 2055, no. 1, pp. 139–146, 2008.
  - [21] S. Jung, X. Qin, and D. A. Noyce, "Injury severity of multi-vehicle crash in rainy weather," *Journal of Transportation Engineering*, vol. 138, no. 1, pp. 50–59, 2012.
  - [22] S. Das, S. R. Geedipally, and K. Fitzpatrick, "Inclusion of speed and weather measures in safety performance functions for rural roadways," *IATSS Research*, 2020.
  - [23] AASHTO, *Highway Safety Manual*, American Association of State Highway and Transportation Officials, Washington, DC, USA, 2010.
  - [24] J. M. Hilbe, *Negative Binomial Regression*, Cambridge University Press, Cambridge, UK, 2nd edition, 2011.
  - [25] B. Ripley, B. Venables, D. M. Bates, K. Hornik, A. Gebhardt, and D. Firth, "Package "mass"," 2014, <http://cran.r-project.org/web/packages/MASS/MASS.pdf>.
  - [26] R Core Team, "R: a language and environment for statistical computing," 2019, <http://www.R-project.org/>.
  - [27] A. C. Cohen, "Estimating the parameter in a conditional Poisson distribution," *Biometrics*, vol. 16, no. 2, pp. 203–211, 1960.
  - [28] T. Yee, *Vgam: Vector Generalized Linear and Additive Models*, Springer, Berlin, Germany, 2017.
  - [29] B.-J. Park, D. Lord, and L. Wu, "Finite mixture modeling approach for developing crash Modification factors in highway safety analysis," *Accident Analysis & Prevention*, vol. 97, pp. 274–287, 2016.
  - [30] X. Zhong, Y. Zou, S. Yuan, Z. Dong, and M. Ijaz, "Finite mixture survival model for examining the variability of urban arterial travel time for buses, passenger cars and taxis," *IET Intelligent Transport Systems*, vol. 45, 2020, In press.
  - [31] B.-J. Park and D. Lord, "Application of finite mixture models for vehicle crash data analysis," *Accident Analysis and Prevention*, vol. 41, no. 4, pp. 683–691, 2009.
  - [32] B.-J. Park, D. Lord, and J. D. Hart, "Bias properties of bayesian statistics in finite mixture of negative binomial regression models in crash data analysis," *Accident Analysis and Prevention*, vol. 42, no. 2, pp. 741–749, 2010.
  - [33] Y. J. Zou, J. E. Ash, B. J. Park, D. Lord, and L. T. Wu, "Empirical bayes estimates of finite mixture of negative binomial regression models and its application to highway safety," *Journal of Applied Statistics*, vol. 45, no. 9, pp. 1652–1669, 2018.
  - [34] B.-J. Park, D. Lord, and C. Lee, "Finite mixture modeling for vehicle crash data with application to hotspot identification," *Accident Analysis & Prevention*, vol. 71, pp. 319–326, 2014.
  - [35] M. Plummer, A. Stukalov, and M. Denwood, "Bayesian graphical models using MCMC," *R Package Version*, vol. 4, no. 6, 2016.
  - [36] A. Inglis, A. Ahmed, B. Wundervald, and E. Prado, "Jags: just another gibbs sampler," 2018, <http://mcmc-jags.sourceforge.net/2004>.
  - [37] Y. Meng, L. Wu, C. Ma, X. Guo, and X. Wang, "A comparative analysis of intersection hotspot identification: fixed vs. varying dispersion parameters in negative binomial models," *Journal of Transportation Safety & Security*, pp. 1–18, 2020.
  - [38] E. Hauer, "Statistical road safety modeling," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1897, no. 1, pp. 81–87, 2004.
  - [39] E. Hauer, "Overdispersion in modelling accidents on road sections and in empirical bayes estimation," *Accident Analysis and Prevention*, vol. 33, no. 6, pp. 799–808, 2001.
  - [40] D. Lord and L. F. Miranda-Moreno, "Effects of low sample mean values and small sample size on the estimation of the fixed dispersion parameter of Poisson-gamma models for modeling motor vehicle crashes: a bayesian perspective," *Safety Science*, vol. 46, no. 5, pp. 751–770, 2008.
  - [41] D. Lord and P. Y.-J. Park, "Investigating the effects of the fixed and varying dispersion parameters of Poisson-gamma models on Empirical Bayes estimates," *Accident Analysis & Prevention*, vol. 40, no. 4, pp. 1441–1457, 2008.
  - [42] S.-P. Miaou and D. Lord, "Modeling traffic crash-flow relationships for intersections: dispersion parameter, functional form, and bayes versus empirical bayes methods," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1840, no. 1, pp. 31–40, 2003.
  - [43] K. Wang, S. Zhao, and E. Jackson, "Multivariate Poisson lognormal modeling of weather-related crashes on freeways," *Transportation Research Record*, vol. 2672, no. 38, pp. 184–198, 2018.
  - [44] P. D. Hoff, *A First Course in Bayesian Statistical Methods*, Springer, New York, NY, USA, 2009.
  - [45] S. P. Brooks and A. Gelman, "General methods for monitoring convergence of iterative simulations," *Journal of Computational and Graphical Statistics*, vol. 7, no. 4, pp. 434–455, 1998.



- [46] D. M. Belmont, "Effect of shoulder width on accidents on two-lane tangents," 2014, <http://www.ktc.uky.edu/files/2012/09/1980-The-Effect-of-Lane-and-Shoulder-Widths-on-Accident-Reductions-on-Rural-Two-Lane-Roads-Report-No.-561.pdf>.
- [47] E. Hauer, "Lane width and safety," 2014, <http://www.roadsafetyresearch.com/>.
- [48] I. B. Potts, D. W. Harwood, and K. R. Richard, "Relationship of lane width to safety on urban and suburban arterials," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2023, no. 1, pp. 63–82, 2007.
- [49] Y. Zou, K. Henrickson, L. Wu, Y. Wang, and Z. Zhang, "Application of the empirical bayes method with the finite mixture model for identifying accident-prone spots," *Mathematical Problems in Engineering*, vol. 2015, p. 10, Article ID 958206, 2015.
- [50] Y. Zou, Y. Zhang, and D. Lord, "Analyzing different functional forms of the varying weight parameter for finite mixture of negative binomial regression models," *Analytic Methods in Accident Research*, vol. 1, pp. 39–52, 2014.
- [51] Y. Zou, Y. Zhang, and D. Lord, "Application of finite mixture of negative binomial regression models with varying weight parameters for vehicle crash data analysis," *Accident Analysis & Prevention*, vol. 50, pp. 1042–1051, 2013.