Taylor & Francis
Taylor & Francis Group

Transportmetrica A: Transport Science

ISSN: 2324-9935 (Print) 2324-9943 (Online) Journal homepage: https://www.tandfonline.com/loi/ttra21

Accounting for heterogeneity in traffic crash prediction: exploring the usage of a dynamic state-space approach

Chunjiao Dong, Qian Yang, Dongchu Cui & Kun Xie

To cite this article: Chunjiao Dong, Qian Yang, Dongchu Cui & Kun Xie (2019): Accounting for heterogeneity in traffic crash prediction: exploring the usage of a dynamic state-space approach, Transportmetrica A: Transport Science, DOI: 10.1080/23249935.2019.1594446

To link to this article: https://doi.org/10.1080/23249935.2019.1594446

Accepted author version posted online: 14 Mar 2019.
Published online: 27 Mar 2019.

TRANSPORTMETRICA A: TRANSPORT SCIENCE
https://doi.org/10.1080/23249935.2019.1594446

Accounting for heterogeneity in traffic crash prediction: exploring the usage of a dynamic state-space approach

Chunjiao Dong[a], Qian Yang[b], Dongchu Cui[c] and Kun Xie[d]

[a]Key Laboratory of Transport Industry of Big Data Application Technologies for Comprehensive Transport, Ministry of Transport, Beijing Jiaotong University, Beijing, People's Republic of China; [b]School of Civil Engineering, Shijiazhuang Tiedao University, Shijiazhuang, People's Republic of China; [c]School of Economics and Management, Yanshan University, Qinhuangdao City, People's Republic of China; [d]National Demonstration Center for Experimental Traffic and Transportation Education, School of Traffic and Transportation, Beijing Jiaotong University, Beijing, People's Republic of China

ABSTRACT
A dynamic state-space model is proposed to predict the crash counts. The outcomes of a multivariate regression model that identifies dynamics relationship between the examined factors and the traffic crashes have been incorporated in the proposed sate-space modes as an initial value to describe the state transition process. The KEF and VBAKF algorithms have been developed to estimate the proposed models and the developed models are referred to as SSMKEF and SSM-VBAKF models, respectively. The findings suggest that the proposed state-space model has better prediction accuracy and robustness with the VBAKF algorithm as the estimation method and the prediction accuracy that measured by RMSD can be improved by 23.81% compared to the KEF algorithm. The findings suggest that the proposed SSM-VBAKF and SSM-KEF models can better address the heterogeneity issues and a significant number of zeros in correlated crash data, and provide sufficient fit to the multivariate correlated crash data.

ARTICLE HISTORY
Received 21 June 2018
Accepted 8 March 2019

KEYWORDS
Highway safety; dynamic state-space model; traffic crashes; SSM-KEF model; SSM-VBAKF model

1. Introduction

Although motor vehicle travel provides an unparalleled degree of mobility and is a major mean of transportation in the United States, according to the report of NHTSA (NHTSA 2017a), traffic crashes were the leading cause of death. Except a slight increase in 2012, there has been a general downward trend of traffic fatalities over the past decade. However, in the most recent two years, there are two large percentage increases in traffic fatalities (NHTSA 2017b). One is in 2016, there were 1976 more traffic fatalities than in 2015 – a 5.6-percent increase, which is lower than the 8.4-percent increase from 2014 to 2015. In total, 37,461 people were killed in traffic crashes on roadways during 2016 (NHTSA 2017b). To reduce deaths, injuries, and relevant medical costs from traffic crashes, with the data that obtained from police reports, local weather stations, and state highway-asset-management

CONTACT Chunjiao Dong cjdong@bjtu.edu.cn Key Laboratory of Transport Industry of Big Data Application Technologies for Comprehensive Transport, Ministry of Transport, Beijing Jiaotong University, No. 3, Shangyuancun, Haidian District, Beijing 100044, People's Republic of China

databases, many methods have been developed to investigate the relationship between the influence factors and traffic crash outcomes and intend to provide effective countermeasures. Lord and Mannering (2010) provided a comprehensive review about the methods and approaches that previously applied to traffic crash analyses along with their strengths and weaknesses. Basically, the applied methodologies can be classified into two categories. One is focusing on the development of statistical models and another is based on the machine learning approaches.

To account for the integer nature of the crash data, a variety of regression methods that are based on Poisson distribution as well as some extensions of the Poisson model have been applied for traffic crash analyses over the years. For example, as an extension of the Poisson model, the negative binomial (or Poisson-gamma) model was proposed to overcome potential over-dispersion issues in the crash data that Poisson models cannot handle (Daniels et al. 2010; El-Basyouny and Sayed 2006; Kim and Washington 2006; Lord and Bonneson 2005; Lord 2006; Malyshkina and Mannering 2010; Miaou and Lord 2003). To address the under-dispersed issue, Poisson-lognormal model have been proposed as an alternative of the most commonly used negative binomial/Poisson-gamma model for traffic crash analyses (Miaou, Song, and Mallick 2003; Lord and Miranda-Moreno 2008; Aguero-Valverde and Jovanis 2008). Although the negative binomial (or Poisson-gamma) and Poisson-lognormal offers more flexibility than the Poisson models, the estimations process can be very complex and the results can be adversely affected by low sample-mean values and small sample sizes (Miaou, Song, and Mallick 2003). To handle the crash data that characterized by a significant amount of zeros, zero-inflated models, such as zero-inflated Poisson and zero-inflated negative binomial have been developed to account for more zeros than a Poisson or negative binomial/Poisson-gamma model can deal with (Carson and Mannering 2001; Lee and Mannering 2002; Kumara and Chin 2003; Shankar et al. 2003; Washington, Karlaftis, and Mannering 2010). The extensions and generalization of the Poisson models also include Conway-Maxwell-Poisson models (Lord and Bonneson 2007) and Gamma models (Oh, Washington, and Nam 2006). To account for the unobserved heterogeneity from one roadway site to another, the random-parameter feature has been incorporated in these generalized Poisson models and the estimated parameters are allowed to vary across each individual observation in the dataset (Milton, Shankar, and Mannering 2008). Furthermore, bivariate/multivariate model formulations have been proposed to jointly model more than one crash type simultaneously, since the counts of specific crash types are not independent (Dong et al. 2015, 2016, 2017, 2018). The applied bivariate/multivariate models including the multivariate Poisson model (Ma and Kockelman 2006), the bivariate negative binomial model (Dong et al. 2015), the multivariate Poisson-lognormal model (Park and Lord 2007; El-Basyouny and Sayed 2009; Dong et al. 2014a), the multivariate zero-inflated Poisson model (Dong et al. 2014b), and the multivariate random-parameters zero-inflated negative binomial regression model (Dong et al. 2014c) can better address the correlation issues in crash data and provide new insights.

Although the Poisson model has served as a starting point for traffic crash analysis and its extension and generalization formulations have been proposed to account for a variety of methodological issues associated with traffic crash data for several decades, researchers have often identified the crash data exhibiting some characteristics that make the application of the proposed statistical models problematic. Specifically, the statistical models can be adversely affected by low sample-means and can produce biased estimates in small samples (Lord and Mannering 2010). As an alternative, the machine learning approach based models, including Artificial Neural Network (ANN) and Support Vector Machine (SVM) models have been applied to crash data and used as data analytic methods because of their ability to deal with massive amounts of multi-dimensional data. In addition, because of the modeling flexibility, accurate predictive ability, and good generalization ability, the machine learning-based models have been considered as accurate and generic mathematical models in the field of traffic safety.

Because the commonly used statistical models assume the pre-defined underlying relationship between dependent and independent variables and the violation of the assumption would lead to erroneous estimation, ANN and Bayesian neural network (BNN) models have been employed to address the traffic safety issues for many years (Chang 2005; Abdelwahab and Abdel-Aty 2002; Xie, Lord, and Zhang 2007; Kunt, Aghayan, and Noii 2011; Jadaan, Al-Fayyad, and Gammoh 2014; Akin and Akbas 2010). Although both ANN and BNN models have similar multilevel network structures, they are different in predicting the traffic crashes. For ANN, the weights are assumed to fix. However, the weights of BNN are assuming to follow a probability distribution and the predictions will be integrated over all the probability weights. Basically, the ANN can be characterized by three features: network architecture, model of a neuron, and learning algorithms. Though the ANN and BNN models show better linear/non-linear approximation properties than conventional statistical approaches, these models often cannot be generalized to other data sets (Lord and Mannering 2010).

The SVM models have recently been introduced for traffic safety analyses (Zhang and Xie 2007; Li et al. 2008), which are a new class of models that are based on statistical learning theory and structural risk minimization (Kecman 2005). These models are supposed to approximate any multivariate function to any desired accuracy with a set of related supervised learning methods. It has been found that the SVM models showed better or comparable results to the outcomes predicted by ANN/BNN and other statistical models (Li et al. 2008; Yu and Abdel-Aty 2014; Chen et al. 2016; Dong, Huang, and Zheng 2015; Ren and Zhou 2011; Yu and Abdel-Aty 2013; Kecman 2005). However, like ANN and BNN, the SVM models often cannot be generalized to other data sets and they all tend to behave as black-boxes, which cannot provide the interpretable parameters as statistical models do. Other than the ANN/BNN and SVM models, other machine learning methods have been introduced in traffic safety analyses. Abdel-Aty and Haleem (2011) proposed multivariate adaptive regression splines (MARS) to predict vehicle angle crashes using the data that obtained from Florida. The results showed that MARS outperformed the NB models. The proposed MARS models showed promising results after screening the covariates using the random forest. The findings suggested that MARS is an efficient technique for predicting crashes at unsignalized intersections.

Although more and more influence factors have been incorporated and the proposed models became more and more advanced, there are still some factors are not available and the models result in bias estimations and erroneous predictions. The variations in the effects of variables across the sampled observations that are unknown to the researchers are referred to as unobserved heterogeneity (Mannering, Shankar, and Bhat 2016). The unobserved heterogeneity among sampled observations could arise from spatial correlations, temporal correlations, or a combination of two. For the spatial correlations, data collected from the same geographic entity might share unobserved effects. For the temporal correlations, data collected from the same geographic entity over successive time periods could share unobserved effects. To address the issues of unobserved heterogeneity and overcome the limitations of the statistical models and the machine learning approaches, we proposed a dynamic state-space model for crash data analyses. The state-space model framework has been successfully applied in the field of transportation to solve a broad range of problems (Shumway and Stoffer 2000; Durbin 2000; Stathopoulos and Karlaftis 2003; Dong et al. 2014d), although there aren't lots of applications in the area of traffic safety.

The proposed dynamic state-space models can better account for the characteristics of crash data in terms of non-negative integers with a small range and perform a comprehensive analysis that aims to predict the crash counts. In the proposed state-space models, the analyzed roadway entity (a roadway segment in this study) has been considered as a system and the traffic volume has been considered as the control input variable. Two estimation methods, including KEF and VBAKF algorithms have been developed to estimate the proposed state-space models. The KEF estimation algorithm assumes the measurement noise follows a Gaussian distribution and the VBAKF algorithm assumes the measurement noise is unknown. In addition, the outcome of a multivariate regression model that considers the traffic factors, geometric design features, and environmental characteristics as the independent variables has been incorporated in the proposed state-space models as an initial value to describe the state transition process. With the proposed state-space model, we intend to provide an alternative framework for analyzing crash data that are measured or observed through a stochastic process.

2. Methodology

The traffic flow has been considered as the control input in the proposed state-space model. The study assumed that the effects of influence factors on traffic crashes were captured by the state variables with a logistic output function. Other unobserved factors were considered as process and sensor noise. The proposed state-space model for traffic crash prediction assumes that the state of roadway entity at a time t evolved from the prior state at time $t-1$ according to the equation

$$x_{t+1} = A_t x_t + B_t q_t + w_t \tag{1}$$

where x_t is the state vector at time t, q_t is the vector containing any control inputs, A_t is the state transition matrix, B_t is the control input matrix that applies the effect of each control input variable in the vector q_t on the state vector, and vector w_t denotes process noise that is zero-mean, white Gaussian, stochastic process with covariance matrixes $\mathbf{Q}$.

For k types of crashes, the state vector that includes k state variables can be written as

$$\mathbf{X}_t = \begin{bmatrix} \ln \lambda_{1t} \\ \ln \lambda_{2t} \\ \vdots \\ \ln \lambda_{kt} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} \beta_{1it} u_{it} + \varepsilon_1 \\ \sum_{i=1}^{n} \beta_{2it} u_{it} + \varepsilon_2 \\ \vdots \\ \sum_{i=1}^{n} \beta_{kit} u_{it} + \varepsilon_k \end{bmatrix} \tag{2}$$

where λ_{it} is the expected number of crashes for crash type i at time t, which is a function of explanatory variables u_{tk} (e.g. roadway geometric design features, pavement conditions, and environmental characteristics), $\boldsymbol{\beta}$ is a vector of estimable parameters, and EXP(ε_i) is a gamma-distributed error term with mean 1 and variance α.

To evaluate how the changes of geometric design features, pavement conditions, and environmental characteristics affect traffic safety, the regression models have been embedded in the proposed model. Assuming that $y_t = \ln n_t$, and n is the number of observed traffic crashes, the measurement equation of the roadway entity can be written as

$$y_t = H_t x_t + v_t \tag{3}$$

where H_t is the transformation matrix that maps the state vector parameters into the measurement domain, and v_t is the vector containing the measurement noise terms for each observation in the measurement vector, which is assumed to be zero-mean Gaussian white noise with covariance **R**.

The state vector x_t cannot be observed and the Kalman filter provides an algorithm to determine an estimate of x_t. As an estimation theory for state-space models, Kalman filter provides a recursive solution through a linear optimal filtering to estimate state variables. For the nonlinear system, a linearization process will be performed to approximate the nonlinear system with a linear time varying (LTV) system at each step, which would result in an extended Kalman filter (EKF) (Sepasi, Ghorbani, and Liaw 2014). In this study, the EKF algorithms are employed to estimate the parameters and perform the traffic crash predication. The employed EKF has several merits. It doesn't require integrations backward in time. In addition, it requires no derivations of a tangent linear operator or adjoint equations. Furthermore, the computational requirements are affordable and comparable to other commonly used methods (Evensen 2003). Compared to the Kalman filter, the employed EKF was designed to address the huge computational requirements that are associated with the error covariance matrix. With the EKF algorithm as the estimation method, the developed state-space models are referred to as SSM-EKF models. At each step, using the first order of a Taylor-series, matrices of transition and measurement function are linearized close to the operation point. SSM-EKF model starts filtering with the best available information on the initial state and error covariance. The proposed SSM-EKF that involves two stages: prediction and measurement update is shown below.

Method: SSM-EKF

- Initialize x_0 and P_0
- For $t = 1, 2, \ldots$

-**Prediction:**

Project the state ahead $\hat{X}_t^- = A\hat{X}_{t-1} + Bu_t$

Project the error covariance ahead $P_t^- = AP_{t-1}A^T + Q$

-**Update:**

Compute the Kalman Gain $K_t = P_t^- H^T (HP_t^- H^T + R)^{-1}$

Update estimate with observations $y_t \hat{X}_t = \hat{X}_t^- + K_t(y_t - H\hat{X}_t^-)$

Update the error covariance $P_t = (I - K_t H)P_k^-$

In the proposed SSM-EKF model, the measurement noise covariance matrix **R** are assumed to known. However, in most situations, the noise covariance **R** is unknown. In case of unknown covariance matrix, the variational Bayesian adaptive Kalman filter (Mbalawata et al. 2015) is employed and the Markovian dynamic model prior for the unknown measurement noise covariance is defined as

$$\mathbf{R}_t \sim p(\mathbf{R}_t|\mathbf{R}_{t-1}) \tag{4}$$

The extract Bayesian filter for Equation (4) is computationally intractable. However, the distribution can be jointly represented by the filtering distribution of the state and covariance matrix and approximated with the free-form variational Bayesian approximation as follows:

$$p(x_t, R_t|y_{1:t-1}) \approx \mathrm{N}(x_t|\hat{\mathbf{X}}_t, \mathbf{P}_t)\mathrm{IW}(R_t|\upsilon_t, V_t) \tag{5}$$

where $\hat{\mathbf{X}}_t$ and $\mathbf{P}_t$ are given by the standard Kalman filter, and v_k and $\mathbf{V}_k$ are the parameters of the inverse Wishart (IW) distribution. The proposed covariance can be computed as the mean of the inverse Wishart distribution:

$$\mathbf{R}_t = \frac{1}{\upsilon_k - d - 1}\mathbf{V}_k \tag{6}$$

Särkkä and Nummenmaa (2009) suggested that Equation (4) is hard to construct explicitly and Särkkä and Hartikainen (2013) proposed a heuristic dynamic model for the covariance

$$\begin{aligned} \upsilon_t^- &= \rho(\upsilon_{t-1} - d - 1) + d + 1 \\ \mathbf{R}_t^- &= \mathbf{B}\mathbf{R}_{t-1}\mathbf{B}^T \end{aligned} \tag{7}$$

where υ_t^- and R_t^- are prior parameters, ρ is a real number and $0 < \rho \leq 1$, which controls the forgetting of the previous estimates of the measurement covariance matrix by decreasing the degrees of freedom exponentially, and $\mathbf{B}$ is a matrix, $0 < |\mathbf{B}| \leq 1$, which can be used to model the deterministic dynamics of the covariance matrix.

Based on the VNAKF algorithm, the SSM-VBAKF models can be developed, which is slightly different compared to the SSM-KEF models. For the prediction, Equation (8) is computed after projecting the error covariance P_t^-. For the update process, set $\upsilon_t = \upsilon_t^- + 1$ and the update of the measurement noise covariance, as shown in Equation (8), is added at the end.

$$\begin{aligned} \mathbf{R}_t^{j+1} = {} & \left(\frac{\upsilon_{t-1} - d - 1}{\upsilon_t - d - 1}\right)\mathbf{R}_t^- + \left(\frac{1}{\upsilon_t - d - 1}\right)\mathbf{H}_t\mathbf{P}_t^{j+1}\mathbf{H}_t^T \\ & + \left(\frac{1}{\upsilon_t - d - 1}\right)(\mathbf{y}_t - \mathbf{H}_t\hat{\mathbf{X}}_t^{j+1})(\mathbf{y}_t - \mathbf{H}_t\hat{\mathbf{X}}_t^{j+1})^T \end{aligned} \tag{8}$$

The iteration will be continuing until the convergence achieved (i.e. N times for $j = 1, 2, \ldots, N$). In SSM-VBAKF, the number of iterations N depends on the problems. However, the previous research (Mbalawata et al. 2015) found that the algorithm required only a few iterations to converge (such as $N = 5$). In addition, a criterion can be set up to determine a suitable time to stop by monitoring the changes in the estimates at each iteration.

3. Data description

The data are obtained from the Tennessee Roadway Information Management System (TRIMS) and the Pavement Management System (PMS) that are maintained in the Tennessee Department of Transportation (TDOT). After the initial screening, 1587 roadway segments that are in Knox County are chosen for the analyses and a total of 15,179 traffic crashes were reported on these selected roadway segments from 2010 to 2014. In TRIMS, the crash data have been classified into five categories according to the injury severities, which are fatal, incapacitating injury, non-incapacitating injury, possible-injury, and PDO crashes. Because the number of fatal crashes is much less compared to the number of other crash categories, the fatal crashes and incapacitating injury crashes have been combined and referred to as major injury crashes. Correspondingly, the possible-injury and PDO crashes have been combined and referred to as no-injury crashes. The non-incapacitating injury crashes are referred to as minor injury crashes. A few of pervious literature (Chang and Chien 2013; Wu et al. 2014; Pahukula, Hernandez, and Unnikrishnan 2015) has used a similar classification method for injury outcomes. The final dataset includes 405 (2.67%) major injury crashes, 3806 (25.07%) minor injury crashes, and 10,969 (72.26%) PDO crashes. Individual roadway segment experienced from 0 to 22 crashes per year with a mean of 1.91 and a standard deviation of 2.65. As expected, a significant amount of zeros is observed. The obtained data have been classified into three groups. The crash data from 2010 to 2012 have been used to develop a regression model and the estimated parameters β can be used to initialize the state transition matrix $\mathbf{A}$. The crash data of 2013 have been used as the observation y to calibrate the prediction accuracy of the proposed state-space model and the crash data of 2014 have been used as the verification set.

For the selected 1587 roadway segments, the traffic factors, geometric design features, pavement factors, and environmental characteristics are linked to the number of crashes through the common variable *id_number*. In other words, the dataset contains detailed information on roadway segments. The considered traffic factors include the thousand passenger car annual average daily traffic (AADT), thousand truck AADT, and posted speed limits. The thousand passenger car AADT from 2010 to 2014 varies from 0.72 to 30.24 with a mean of 7.36 and a standard deviation of 6.78 and the thousand truck AADT from 2010 to 2014 varies from 0.12 to 2.66 with a mean of 0.76 and a standard deviation of 0.67. The variable of the posted speed limit has been considered as the categorical variable with two classes. For 65.97% (5235) roadway segment, the posted speed limit is less than 55 mph and the posted speed limit of 34.03% (2700) roadway segment is no less than 55 mph.

Important measurements of geometric design features considered in this study include segment length, degree of horizontal curvature, median widths, outsider shoulder widths, number of through lanes, lane widths, median types, and shoulder type. Among them, the segment length, degree of horizontal curvature, median widths, and outsider shoulder widths are considered as the continuous variables and the others are considered as the categorical variables. Other than the traffic factors and geometric design features, the impacts of pavement surface characteristics are considered to better address traffic safety issues for roadway design and maintenance. The considered pavement surface characteristics include the international roughness index (IRI) and rut depth (RD). The analyzed IRI varies from 25.54 to 182.75 with a mean of 65.45 and a standard deviation of 27.86, which is calculated using a quarter-car vehicle math model and the response is accumulated to yield a roughness index with units of slope (in/mi). Another pavement condition indicator is the RD, which is measured at roadway speeds with a laser/inertial profilograph. The analyzed RD varies from 0.07 to 0.57 with a mean of 0.15 and a standard deviation of 0.07.

The environmental factors, including terrain types, lighting condition, and land use type are considered. Two terrain types are examined, which include rolling terrace (62.26%)

Table 1. Summary statistics of analyzed continuous variables.

Variable	Mean	Std. Dev.	Min.	Max.
Independent variable				
The number of major injury crashes per year per roadway segment	0.05	0.30	0	4
The number of minor injury crashes per year per roadway segment	0.48	1.17	0	10
The number of no injury crashes per year per roadway segment	1.38	2.36	0	19
Traffic factors				
Thousand passenger car AADT	7.36	6.78	0.72	30.24
Thousand truck AADT	0.76	0.67	0.12	2.66
Geometric design features				
Segment length (miles)	0.82	1.13	0.17	12.43
Degree of horizontal curvature	1.51	3.35	0.00	14.24
Median widths	1.12	2.24	0.00	12.49
Outside shoulder widths	3.35	1.96	3.37	8.25
Pavement factors				
International roughness index	65.45	27.86	25.54	182.75
Rut depth (in.)	0.15	0.07	0.07	0.57

Table 2. Summary statistics of analyzed categorical variables.

Variable	Category	Frequency	Percent
Traffic factor			
Posted speed limits	$<$ 55 mph	5235	65.97
	$\geq$ 55 mph	2700	34.03
Geometric design features			
Number of through lanes	6	1480	18.65
	4	4105	51.73
	2	2350	29.62
Lane widths (ft)	12	2170	27.35
	11	4185	52.74
	10	1580	19.91
Median type	1 for non-traversable median	2990	28.80
	0 for traversable median	4945	20.98
Shoulder type	2 for pavement	1725	50.22
	1 for gravel	4140	37.68
	0 for dirt	2070	62.32
Environmental factors			
Terrain type	1 for mountainous	2995	37.74
	0 for rolling	4940	62.26
Land use type	2 for residential	4105	51.73
	1 for commercial	1905	24.01
	0 for rural	1925	24.26
Indicator for lighting	1 for lighting exists on the roadway segments	3505	44.17
	0 for others	4430	55.83

and mountainous terrace (37.74%). Lighting condition was considered as a category variable, which indicates whether lighting devises are provided at the roadway segments. Three types of land use are considered, including commercial (24.01%), rural (24.26%), and residential (51.73%). These variables are considered because they might have potential significant effects on traffic safety. The descriptive statistics of continuous variables and categorical variables are shown in Tables 1 and 2, respectively.

4. Modeling results

The MATLAB was employed for model development. The data from 2010 to 2012 were used as the input to develop a MVNB regression models and the estimate parameters, as shown in Table 3, were used as the initial parameters for the state-space models. More background information about MVNB regression models can be found in Done et al. (2014b and 2015). For the proposed state-space models, two types of estimation algorithms were employed, which are KEF and VBKEF algorithms and the developed models are referred to as SSM-KEF model and SSM-VBAKF model. The MVBN models that are based on four-year data were also employed as the comparison method to verify the effectiveness of the proposed state-space models.

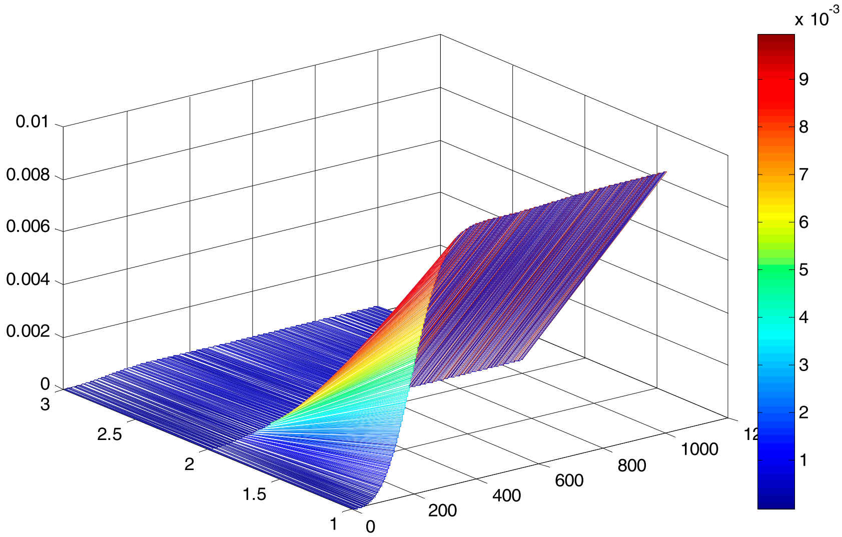
Table 3 provides the estimation results of the MVNB regression models that are based on three-year data. The results show that the variables of thousand passenger car AADT, thousand truck AADT, degree of horizontal curvature, rut depth, and mountainous terrain type have significant positive effects on major injury crashes, and the variables of median widths, posted speed limits less than 55 mph, non-traversable median, residential land use type, and lighting exists on the roadway segments have significant negative effects on major injury crashes. In other words, for the continuous variables, increasing the values of thousand passenger car AADT, thousand truck AADT, degree of horizontal curvature, and rut depth will reduce the likelihood of major injury crashes, and decreasing the values of median widths will increase the likelihood of major injury crashes. For the categorical variables, compared to the rolling terrain type, the mountainous terrain type is associated with a higher risk of major injury crashes. The posted speed limits less than 55 mph, non-traversable median, residential land use, and lighting exists on the roadway segments are associated with lower risk of major injury crashes, compared to the conditions of posted speed limits greater than (or equal to) 55 mph, traversable median, commercial and rural land use, and no lighting existing, respectively.

All the significant variables for the major injury crashes have significant effects for the minor injury and non-injury crashes and the direction of the significant variables are consistent. The comparison of the estimated parameters and *t*-statistics in Table 3 suggests not only that the significance level of the examined factors are quite consistent, but also that the estimated coefficients have the expected and consistent algebraic sign. However, for the minor injury crashes, one more variable – commercial land use has been found to have a significant positive effect. Other than the variable of commercial land use, three more variables, including segment length, outside shoulder widths, and number of through lanes have been found to have significant effects on non-injury crashes. Among them, segment length and number of through lanes are positively associated with non-injury crashes and outside shoulder widths is negatively associated with non-injury crashes.

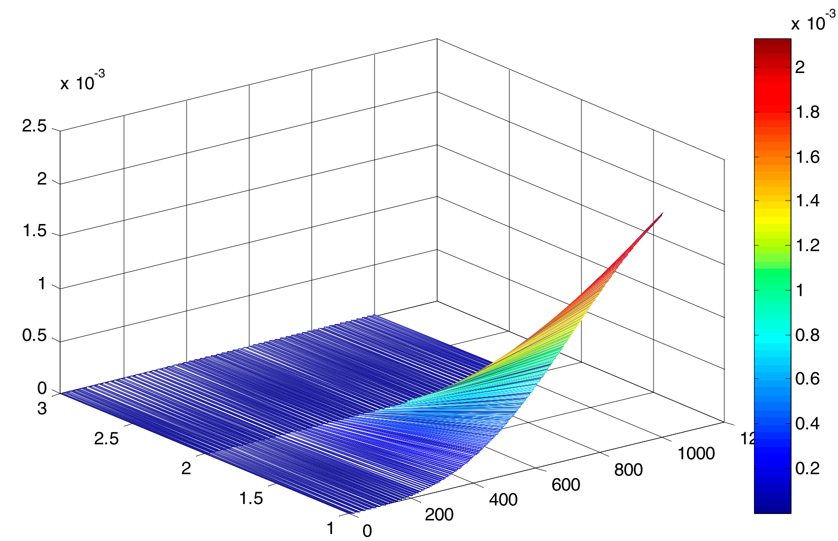
The obtained coefficients are used as the initial values for the proposed state-space models. The proposed models estimate a process by using a form of feedback control. In other words, the filter estimates the process state at some time and then uses the measurements as the feedback. As such, the *a priori* estimates are obtained for the next step through the predictor equations projecting forward the current state and error covariance estimates. An improved *a posteriori* estimate is obtained through the measurement update equations by incorporating a new observation into the *a prior* estimate. The convergence process of Kalman gains of SSM-KEF and SSM-VBAKF models are shown in Figure 1. It can be seen that the SSM-VBAKF model has better performances in terms of narrow search space and more efficient convergence process. Compared to the SSM-VBAKF model, the proposed SSM-KEF model can achieve the same accuracy with comparable search space. Because the

Table 3. The estimated parameters from MVNB model.

	Major injury crashes				Minor injury crashes				PDO crashes			
Variables	Estimates	Std. Err.	*t* Value	Pr > |*t*|	Estimates	Std. Err.	*t* Value	Pr > |*t*|	Estimates	Std. Err.	*t* Value	Pr > |*t*|
Intercept	2.44	0.58	4.23	< .0001	0.97	0.25	3.95	0.0001	2.58	0.75	3.45	0.0006
Thousand passenger car AADT	0.70	0.15	4.57	< .0001	1.26	0.18	7.15	< .0001	2.58	0.62	4.17	< .0001
Thousand truck AADT	0.68	0.29	2.39	0.0173	0.44	0.13	3.53	0.0004	0.71	0.23	3.02	0.0027
Segment length (miles)	–	–	–	–	5.92	2.03	2.91	0.0037	0.12	0.05	2.58	0.0100
Degree of horizontal curvature	7.09	3.26	2.18	0.0299	0.81	0.06	14.16	< .0001	5.07	0.37	13.66	< .0001
Median widths	−0.76	0.14	−5.31	< .0001	−4.40	1.39	−3.17	0.0016	−8.64	3.64	−2.37	0.0179
Outside shoulder widths	–	–	–	–	−1.21	0.13	−9.59	< .0001	−8.79	0.95	−9.23	< .0001
Rut depth (in.)	0.52	0.19	2.75	0.0061	1.81	0.11	16.45	< .0001	9.52	0.70	13.53	< .0001
Posted speed limits < 55 mph	−4.24	0.66	−6.41	< .0001	−0.88	0.24	−3.67	0.0003	−0.41	0.08	−5.12	< .0001
Number of through lanes (3 lanes)	7.07	1.09	6.50	< .0001	2.42	0.95	2.54	0.0113	4.37	1.70	2.57	0.0104
Non-traversable median	−8.30	1.29	−6.41	< .0001	−0.35	0.07	−5.33	< .0001	−1.24	0.48	−2.56	0.0108
Mountainous terrain type	0.44	0.11	4.12	< .0001	0.66	0.06	10.34	< .0001	8.68	2.52	3.45	0.0006
Residential land use	−2.82	1.13	−2.48	0.0133	−0.99	0.16	−6.21	< .0001	−0.76	0.24	−3.17	0.0016
Commercial land use	–	–	–	–	0.28	0.06	4.89	< .0001	2.62	1.09	2.39	0.0169
Lighting exists on the roadway segments	−0.67	0.13	−5.12	< .0001	−4.36	0.87	−5.01	< .0001	−1.39	0.31	−4.56	< .0001



(a)



(b)

Figure 1. The comparison of convergence process of Kalman gains. (a) The convergence process of Kalman gains of SSM-EKF model, (b) The convergence process of Kalman gains of SSM-VBAKF model.

SSM-VBAKF model allows the usage of unknown measurement noise covariance matrix, it provides more flexible and convincing results.

The developed SSM-KEF and SSM-VBAKF models have been used to predict the crash counts for the analyzed 1538 roadway segments. The MVNB models are used as a comparison method. Two commonly used evaluation measurements, including Mean Absolute

Table 4. Comparison of prediction results.

	Major injury crashes	Minor injury crashes	No-injury crashes	Total
Observation				
Observed mean	0.06	0.45	1.32	1.82
Observed Std. Dev	0.33	1.15	2.28	2.56
SSM-KEF model				
Predicted mean	0.20	0.50	1.33	2.03
Predicted Std. Dev.	0.54	1.50	2.69	3.07
MAE	0.15	0.14	0.26	0.49
RMSD	0.37	0.48	0.63	0.87
SSM-VBAKF model				
Predicted mean	0.12	0.46	1.17	1.75
Predicted Std. Dev.	0.32	1.12	2.44	2.67
MAE	0.11	0.19	0.25	0.44
RMSD	0.25	0.35	0.56	0.68
MVNB model				
Predicted mean	0.21	0.78	1.70	2.70
Predicted Std. Dev.	0.64	1.62	3.04	3.48
MAE	0.17	0.35	0.50	0.97
RMSD	0.44	0.71	0.97	1.46

Error (MAE) and Root mean squared error (RMSE) have been employed to assess the model performances in terms of prediction accuracy and robustness. The prediction results are shown in Table 4. Results in Table 4 show that the predicted means from the proposed SSM-VBAKF model (0.12, 0.46, 1.17, and 1.75 for a major injury, minor injury, no-injury, and all crashes, respectively) are very close to the observed means (0.06, 0.45, 1.32, and 1.82). The SSM-KEF model results in a mean of 0.20, 0.50, 1.33, and 2.03 for major, minor, and no-injury crashes, which is comparable to the results of SSM-VBAKF model. Compared to the SSM-VBAKF and SSM-KEF models, the MVNB model results in a mean of 0.21, 0.78, 1.70, and 2.70, which indicates the worse performances. We believe that a very importation feature of the proposed model is that it can provide a good prediction of the chance that the roadway segment is in the crash-free state or some crash-prone propensity states.

For all the observed samples, the proposed SSM-VBAKF model results in a MAE of 0.11, 0.19, 0.25, and 0.44, and a RMSD of 0.25, 0.35, 0.56, and 0.68 for a major injury, minor injury, no-injury, and all crashes, respectively. The SSM-KEF model results in a MAE of 0.15, 0.14, 0.26, and 0.49, and a RMSD of 0.37, 0.48, 0.63, and 0.87 for major injury, minor injury, no-injury, and all crashes, respectively. The MVNB model results in a MAE of 0.17, 0.35, 0.50, and 0.97, and a RMSD of 0.44, 0.71, 0.97, and 1.46 for a major injury, minor injury, no-injury, and all crashes, respectively. The results suggest that the proposed SSM-VBAKF model predicts better than the proposed SSM-KEF model and MVNB model. In summary, compared to the SSM-KEF model and MVNB model, the proposed SSM-VBAKF model results in the most accurate predictions that are measured MAE and RMSD. We hypothesize that the proposed state-space models better addresses the issue of temporal correlation and allows for excess zero counts in correlated data.

The findings indicate that the predictions from the proposed SSM-VBAKF models have significant improvements over the comparable models, in both accuracy and robustness. Compared to the SSM-KEF model, the RMSD improvement of the proposed SSM-VBAKF model ranges from 12.34% to 33.39% with a mean of 23.81%. Compared to the MVNB model, the RMSD improvements of the proposed SSM-VBAKF model ranges from 42.80%

to 53.26% with a mean of 47.74%. Clearly, the improvement is significant for the proposed models over MVNB models. Therefore, the proposed model seems to be a better alternative for crash count predictions. The best-performing result of the proposed SSM-VBAKF model for major injury crashes has a RMSD of 0.25, an 33.39% and 44.51% improvement from the SSM-KEF model and MVNB model, respectively. The proposed SSM-VBAKF model performs worse for no-injury crashes, with a 0.56 RMSD for all observed samples. However, it is still better than the results of the SSM-KEF model and MVNB model and the prediction accuracy has been improved by 12.34% and 42.80%, respectively.

The proposed models have better performances in terms of small error variances than the comparison models, since the proposed models are based on a dynamic state-space model, which contains a dynamic equation governing the state dynamics and the observations have been used to calibrate the prediction accuracy. The overall performances of the proposed SSM-VBAKF model for all crashes show an 22.09% improvement over the SSM-KEF model and an 53.26% improvement over the MVNB model. It is clear that the predictions obtained from the proposed models are superior to those obtained from the comparison models. The greatest difference is demonstrated for the no-injury crashes where the proposed SSM-VBAKF model yields a RMSD of 0.56 compared to a 0.97 RMSD value from the MVNB models. The improvements of the proposed SSM-KEF model over the MVNB model are also significant, which are 44.51%, 50.38%, 42.80%, and 53.26% for a major injury, minor injury, no-injury, and all crashes, respectively. The predicted results are compared to the observed values and the results are shown in Figure 2.

Figure 2(a) displays the empirical frequency distribution of major injury crashes and the predicted distribution from SSM-VBAKF, SSM-KEF, and MVNB models. As expected, a significant amount of zeros (representing crash-free states) is present. Such a pattern is typical of roadway segment based crash counts. It is evident that the proposed models provide a sufficient fit to the multivariate correlated crash data. The results in Figure 2 (b) to (d) are consistent with Figure 2 (a), which shows the proposed models can better account for the over-dispersion issues and a significant amount of zeros in correlated crash data. The comparison results suggest that the SSM-VBAKF models provide a better goodness of fit compared to the SSM-EKF models and MVNB models. Compared to the proposed SSM-VBAKF model, the proposed SSM-EKF model provides comparative prediction results and the proposed SSM-VBAKF model and SSM-KEF model give a better goodness of fit to the data than MVNB models. The findings reveal that the proposed models have superior predictions while the MVNB model has relatively large overestimations. These examples demonstrate the superior prediction accuracy and robustness of the proposed model for crash count predictions across injury severities.

The prediction results are further compared in Figure 3. From the plots of predicted values versus measurement values, we see that the predicted values from the proposed SSM-VBAKF are quite close to the measurement values for all crash types. The results are further confirming the findings that the proposed SSM-VBAKF models are the superior alternative models for crash count predictions.

As the graphical analysis of the crash count patterns shows, the proposed models, particularly for the zero crash counts, capture the general pattern of the crash count distribution and reflect in a manner of crash count distribution across injury severities. In addition, the proposed models enable the predicted crash counts to retain a smoother pattern, which is less sensitive to the high-frequency variations of crash counts, in contrast to the pattern

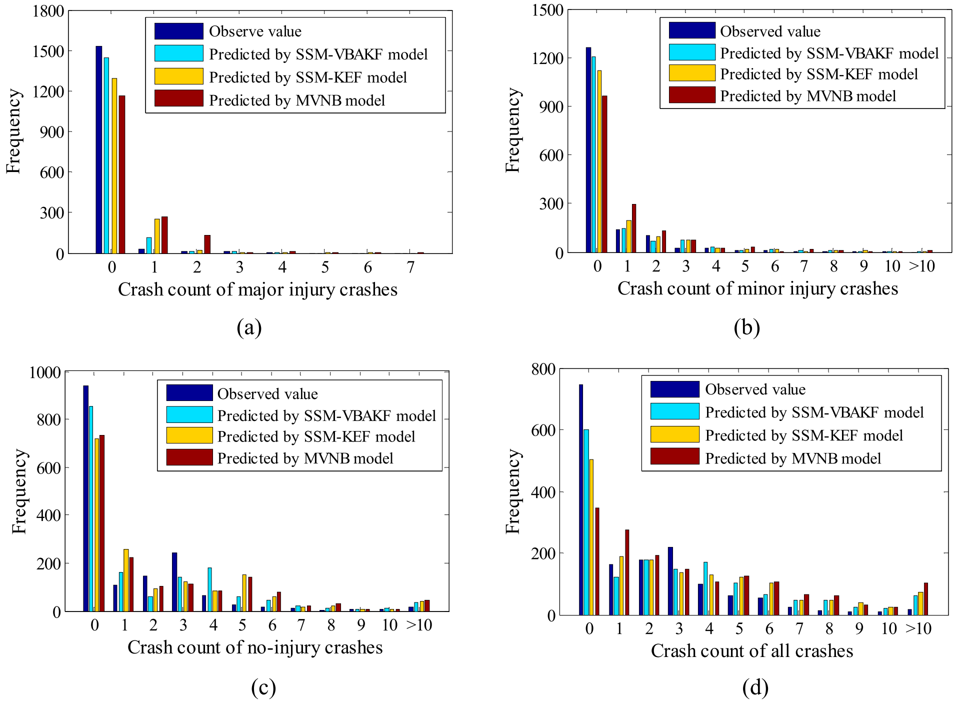


Figure 2. The frequency distribution of crash counts. (a) Distribution comparison of major injury crashes, (b) Distribution comparison of minor injury crashes, (c) Distribution comparison of no-injury crashes, (d) Distribution comparison of all crashes.

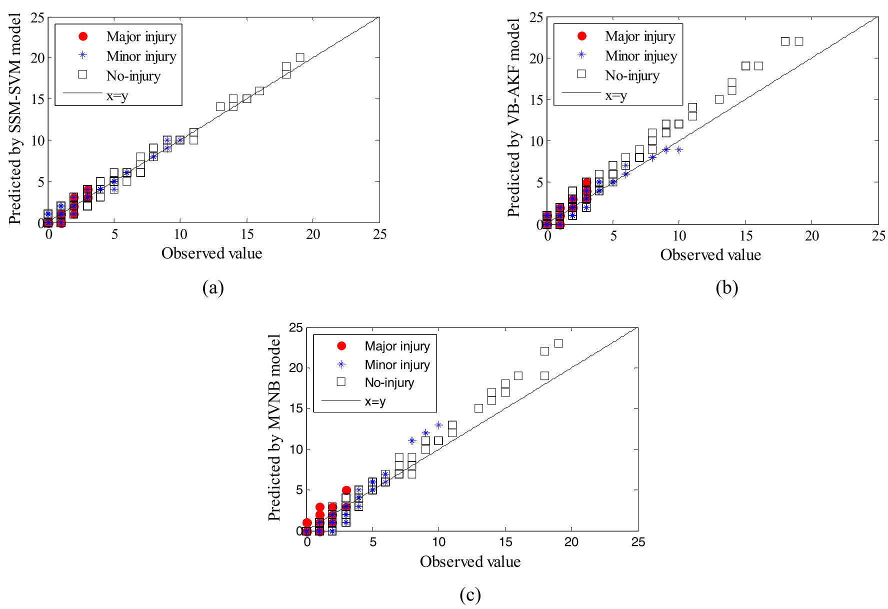


Figure 3. The comparison of model performances. (a) Predicted value of SSM-VBAKF model vs. observed value, (b) Predicted value of SSM-KEF model vs. observed value, (c) Predicted value of MVNB model vs. observed value.

obtained from the comparison models. In other words, we believe that the robustness of the proposed models is an advantage.

5. Conclusions

In this study, we presented a dynamic state-space model with two estimation methods. The proposed dynamic state-space models can better account for the characteristics of crash data in terms of non-negative integers with a small range and perform a comprehensive analysis that aims to predict the crash counts. With the SSM-KEF models, the estimation process assumes that the measurement noise follows Gaussian distribution. In the SSM-VBAKF models, the measurement noise covariance matrix is assumed to unknown. To identify dynamics relationship between the examined variables and the traffic crashes, and enhance the prediction accuracy and robustness, the estimated coefficients from the multivariate regression models have been incorporated as the initial values for the proposed state-space models. Five-year crash data have been employed to demonstrate the application of the proposed models with two different estimation algorithms. The investigation results provide sufficient evidence for the following conclusions:

(1) The proposed models have superior performances in terms of prediction power compared to the MVNB models. The overall performances of the proposed SSM-VBAKF and SSM-KEF models for all crashes show an 47.74% and 24.97% RMSD improvement, respectively, over the MVNB model.
(2) The proposed SSM-VBAKF model has better prediction accuracy and robustness compared to the proposed SSM-KEF model and the prediction accuracy that is measured by RMSD can be improved by 23.81%.
(3) The proposed models provide a sufficient fit to the multivariate correlated crash data. The results show that the proposed models can better account for the over-dispersion issues and a significant amount of zeros in correlated crash data.
(4) The proposed models can capture the general pattern of the crash count distribution and reflect in a manner of crash count distribution across injury severities.

The findings suggest that the proposed model can better address the heterogeneity issues in correlated crash data and is a superior alternative for traffic crash predictions. The proposed models can perform traffic crashes prediction for a given facility. The proposed methodology could be applied to other roadway networks if appropriate influence factors are available. The application of the proposed models will enable academic researchers, traffic engineers, and decision-makers in government agencies to gain the information on traffic crash predications. By employing the proposed models with relative cases and developing them for various time periods, the traffic and transportation engineers can obtain traffic crash predictions with their needs. Further investigation of the proposed models includes the predictions of dynamic spatial-temporal pattern in crash data.

Acknowledgements

Special thanks to TDOT for providing the TRIMS data.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This research is supported by funding provided by the Fundamental Research Funds for the Central Universities [grant number 2019RC027].

References

Abdel-Aty, M., and K. Haleem. 2011. "Analyzing Angle Crashes at Unsignalized Intersections Using Machine Learning Techniques." *Accident Analysis & Prevention* 43 (1): 461–470.

Abdelwahab, H. T., and M. A. Abdel-Aty. 2002. "Artificial Neural Networks and Logit Models for Traffic Safety Analysis of Toll Plazas." *Transportation Research Record: Journal of the Transportation Research Board* 1784 (1): 115–125.

Aguero-Valverde, J., and P. P. Jovanis. 2008. "Analysis of Road Crash Frequency with Spatial Models." *Transportation Research Record: Journal of the Transportation Research Board* 2061: 55–63.

Akin, D., and B. Akbas. 2010. "A Neural Network (NN) Model to Predict Intersection Crashes Based Upon Driver, Vehicle and Roadway Surface Characteristics." *Scientific Research and Essays* 5 (19): 2837–2847.

Carson, J., and F. Mannering. 2001. "The Effect of Ice Warning Signs on Ice-accident Frequencies and Severities." *Accident Analysis and Prevention* 33 (1): 99–109.

Chang, L. 2005. "Analysis of Freeway Accident Frequencies: Negative Binomial Regression Versus Artificial Neural Network." *Safety Science* 43 (8): 541–557.

Chang, L. Y., and J. T. Chien. 2013. "Analysis of Driver Injury Severity in Truck-involved Accidents Using a Non-parametric Classification Tree Model." *Safety Science* 51 (1): 17–22.

Chen, C., G. Zhang, Z. Qian, R. A. Tarefder, and Z. Tian. 2016. "Investigating Driver Injury Severity Patterns in Rollover Crashes Using Support Vector Machine Models." *Accident Analysis and Prevention* 90: 128–139.

Daniels, S., T. Brijs, E. Nuyts, and G. Wets. 2010. "Explaining Variation in Safety Performance of Roundabouts." *Accident Analysis and Prevention* 42 (2): 393–402.

Dong, C., D. B. Clarke, S. S. Nambisan, and B. Huang. 2016. "Analyzing Injury Crashes Using Random-parameter Bivariate Regression Models." *Transportmetrica A: Transport Science* 12 (9): 794–810.

Dong, C., D. B. Clarke, S. H. Richards, and B. Huang. 2014a. "Differences in Passenger Car and Large Truck Involved Crash Frequencies at Urban Signalized Intersections: An Exploratory Analysis." *Accident Analysis and Prevention* 62: 87–94.

Dong, C., D. B. Clarke, X. Yan, A. Khattak, and B. Huang. 2014c. "Multivariate Random-parameters Zero-inflated Negative Binomial Regression Model: An Application to Estimate Crash Frequencies at Intersections." *Accident Analysis and Prevention* 70: 320–329.

Dong, N., H. Huang, and L. Zheng. 2015. "Support Vector Machine in Crash Prediction at the Level of Traffic Analysis Zones: Assessing the Spatial Proximity Effects." *Accident Analysis and Prevention* 90: 128–139.

Dong, C., S. S. Nambisan, S. H. Richards, and Z. Ma. 2015. "Assessment of the Effects of Highway Geometric Design Features on the Frequency of Truck Involved Crashes Using Bivariate Regression." *Transportation Research Part A: Policy and Practice* 75: 30–41.

Dong, C., S. H. Richards, D. B. Clarke, X. Zhou, and Z. Ma. 2014b. "Examining Signalized Intersection Crash Frequency Using Multivariate Zero-inflated Poisson Regression." *Safety Science* 70: 63–69.

Dong, C., C. Shao, S. H. Richards, and L. D. Han. 2014d. "Flow Rate and Time Mean Speed Predictions for the Urban Freeway Network Using State Space Models." *Transportation Research Part C: Emerging Technologies* 43 (1): 20–32.

Dong, C., J. Shi, B. Huang, X. Chen, and Z. Ma. 2017. "A Study of Factors Affecting Intersection Crash Frequencies Using Random-parameter Multivariate Zero-inflated Models." *International Journal of Injury Control and Safety Promotion* 24 (2): 208–221.

Dong, C., K. Xie, J. Zeng, and X. Li. 2018. "Multivariate Dynamic Tobit Models with Lagged Observed Dependent Variables: An Effectiveness Analysis of Highway Safety Laws." *Accident Analysis and Prevention* 113: 292–302.

Durbin, J. 2000. "The Foreman Lecture: The State Space Approach to Time Series Analysis and Its Potential for Official Statistics." *Australian and New Zealand Journal of Statistics* 42 (1): 1–23.

El-Basyouny, K., and T. Sayed. 2006. "Comparison of two Negative Binomial Regression Techniques in Developing Accident Prediction Models." *Transportation Research Record: Journal of the Transportation Research Board* 1950: 9–16.

El-Basyouny, K., and T. Sayed. 2009. "Collision Prediction Models Using Multivariate Poisson-lognormal Regression." *Accident Analysis and Prevention* 41 (4): 820–828.

Evensen. 2003. "The Ensemble Kalman Filter: Theoretical Formulation and Practical Implementation." *Ocean Dynamics* 53: 343–367.

Jadaan, K. S., M. Al-Fayyad, and H. F. Gammoh. 2014. "Prediction of Road Traffic Accidents in Jordan Using Artificial Neural Network (ANN)." *Journal of Traffic and Logistics Engineering* 2 (2): 92–94.

Kecman, V. 2005. "Support Vector Machines–an Introduction." In *Support Vector Machines: Theory and Applications*, edited by L. Wang, 1–48. Berlin: Springer-Verlag.

Kim, D., and S. Washington. 2006. "The Significance of Endogeneity Problems in Crash Models: An Examination of Left-turn Lanes in Intersection Crash Models." *Accident Analysis and Prevention* 38 (6): 1094–1100.

Kumara, S. S. P., and H. C. Chin. 2003. "Modeling Accident Occurrence at Signalized tee Intersections with Special Emphasis on Excess Zeros." *Traffic Injury Prevention* 4 (1): 53–57.

Kunt, M. M., I. Aghayan, and N. Noii. 2011. "Prediction for Traffic Accident Severity: Comparing the Artificial Neural Network, Genetic Algorithm, Combined Genetic Algorithm and Pattern Search Methods." *Transport* 26 (4): 353–366.

Lee, J., and F. Mannering. 2002. "Impact of Roadside Features on the Frequency and Severity of Run-off-roadway Accidents: An Empirical Analysis." *Accident Analysis and Prevention* 34 (2): 149–161.

Li, X., D. Lord, Y. Zhang, and Y. Xie. 2008. "Predicting Motor Vehicle Crashes Using Support Vector Machine Models." *Accident Analysis and Prevention* 40 (4): 1611–1618.

Lord, D. 2006. "Modeling Motor Vehicle Crashes Using Poisson-gamma Models: Examining the Effects of Low Sample Mean Values and Small Sample Size on the Estimation of the Fixed Dispersion Parameter." *Accident Analysis and Prevention* 38 (4): 751–766.

Lord, D., and J. A. Bonneson. 2005. "Calibration of Predictive Models for Estimating Safety of Ramp Design Configurations." *Transportation Research Record: Journal of the Transportation Research Board* 1908: 88–95.

Lord, D., and J. A. Bonneson. 2007. "Development of Accident Modification Factors for Rural Frontage Road Segments in Texas." *Transportation Research Record: Journal of the Transportation Research Board* 2023: 20–27.

Lord, D., and F. Mannering. 2010. "The Statistical Analysis of Crash-frequency Data: A Review and Assessment of Methodological Alternatives." *Transportation Research Part A* 44 (5): 291–305.

Lord, D., and L. F. Miranda-Moreno. 2008. "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* 46 (5): 751–770.

Ma, J., and K. M. Kockelman. 2006. "Bayesian Multivariate Poisson Regression for Models of Injury Count by Severity." *Transportation Research Record: Journal of the Transportation Research Board* 1950: 24–34.

Malyshkina, N., and F. Mannering. 2010. "Empirical Assessment of the Impact of Highway Design Exceptions on the Frequency and Severity of Vehicle Accidents." *Accident Analysis and Prevention* 42 (1): 131–139.

Mannering, F. L., V. Shankar, and C. R. Bhat. 2016. "Unobserved Heterogeneity and the Statistical Analysis of Highway Accident Data." *Analytic Methods in Accident Research* 11: 1–16.

Mbalawata, I. S., S. Särkkä, M. Vihola, and H. Haario. 2015. "Adaptive Metropolis Algorithm Using Variational Bayesian Adaptive Kalman Filter." *Computational Statistical and Data Analysis* 83: 101–115.

Miaou, S.-P., and D. Lord. 2003. "Modeling Traffic Crash-flow Relationships for Intersections: Dispersion Parameter, Functional Form, and Bayes Versus Empirical Bayes." *Transportation Research Record: Journal of the Transportation Research Board* 1840: 31–40.

Miaou, S.-P., J. J. Song, and B. K. Mallick. 2003. "Roadway Traffic Crash Mapping: A Space-time Modeling Approach." *Journal of Transportation and Statistics* 6 (1): 33–57.

Milton, J., V. Shankar, and F. Mannering. 2008. "Highway Accident Severities and the Mixed Logit Model: An Exploratory Empirical Analysis." *Accident Analysis and Prevention* 40 (1): 260–266.

NHTSA. 2017a. *Summary of Motor Vehicle Crashes. Traffic Safety Facts, 2015 Data*. Washington, DC: National Highway Traffic Safety Administration, DOT HS 812 376.

NHTSA. 2017b. *2016 Fatal Motor Vehicle Crashes: Overview. Traffic Safety Facts, Research Note*. Washington, DC: National Highway Safety Administration, DOT HS 812 456.

Oh, J., S. P. Washington, and D. Nam. 2006. "Accident Prediction Model for Railway-highway Interfaces." *Accident Analysis and Prevention* 38 (2): 346–356.

Pahukula, J., S. Hernandez, and A. Unnikrishnan. 2015. "A Time of day Analysis of Crashes Involving Large Truck in Urban Areas." *Accident Analysis and Prevention* 75: 155–163.

Park, E.-S., and D. Lord. 2007. "Multivariate Poisson-lognormal Models for Jointly Modeling Crash Frequency by Severity." *Transportation Research Record: Journal of the Transportation Research Board* 2019: 1–6.

Ren, G., and Z. Zhou. 2011. "Traffic Safety Forecasting Method by Particle Swarm Optimization and Support Vector Machine." *Expert Systems with Applications* 38 (8): 10420–10424.

Sepasi, S., R. Ghorbani, and B. Y. Liaw. 2014. "A Novel on-Board State-of-charge Estimation Method for Aged Li-ion Batteries Based on Model Adaptive Extended Kalman Filter." *Journal of Power Sources* 245: 337–344.

Shankar, V. N., G. F. Ulfarsson, R. M. Pendyala, and M. B. Nebergall. 2003. "Modeling Crashes Involving Pedestrians and Motorized Traffic." *Safety Science* 41 (7): 627–640.

Shumway, R. H., and D. S. Stoffer. 2000. *Time Series Analysis and Its Applications*. New York, NY: Springer.

Stathopoulos, A., and M. G. Karlaftis. 2003. "A Multivariate State Space Approach for Urban Traffic Flow Modeling and Prediction." *Transportation Research Part C: Emerging Technologies* 11 (2): 121–135.

Washington, S. P., M. G. Karlaftis, and F. L. Mannering. 2010. *Statistical and Econometric Methods for Transportation Data Analysis*. 2nd ed. Boca Raton, FL: Chapman Hall/CRC.

Wu, Q., F. Chen, G. Zhang, X. C. Liu, H. Wang, and S. M. Bogus. 2014. "Mixed Logit Model-based Driver Injury Severity Investigations in Single- and Multi-vehicle Crashes on Rural Two-lane Highways." *Accident Analysis and Prevention* 72: 105–115.

Xie, Y., D. Lord, and Y. Zhang. 2007. "Predicting Motor Vehicle Collisions Using Bayesian Neural Networks: An Empirical Analysis." *Accident Analysis and Prevention* 39 (5): 922–933.

Yu, R., and M. Abdel-Aty. 2013. "Utilizing Support Vector Machine in Real-time Crash Risk Evaluation." *Accident Analysis and Prevention* 51: 252–259.

Yu, R., and M. Abdel-Aty. 2014. "Analyzing Crash Injury Severity for a Mountainous Freeway Incorporating Real-time Traffic and Weather Data." *Safety Science* 63: 50–56.

Zhang, Y., and Y. Xie. 2007. "Forecasting of Short-term Freeway Volume with v-Support Vector Machines." *Transportation Research Record: Journal of the Transportation Research Board* 2024: 92–99.