Predicting motor vehicle collisions using Bayesian neural network models: An empirical analysis

Yuanchang Xie [1], Dominique Lord *, Yunlong Zhang [2]

Zachry Department of Civil Engineering, Texas A&M University, 3136 TAMU, College Station, TX 77843-3136, United States

Abstract

Statistical models have frequently been used in highway safety studies. They can be utilized for various purposes, including establishing relationships between variables, screening covariates and predicting values. Generalized linear models (GLM) and hierarchical Bayes models (HBM) have been the most common types of model favored by transportation safety analysts. Over the last few years, researchers have proposed the back-propagation neural network (BPNN) model for modeling the phenomenon under study. Compared to GLMs and HBMs, BPNNs have received much less attention in highway safety modeling. The reasons are attributed to the complexity for estimating this kind of model as well as the problem related to "over-fitting" the data. To circumvent the latter problem, some statisticians have proposed the use of Bayesian neural network (BNN) models. These models have been shown to perform better than BPNN models while at the same time reducing the difficulty associated with over-fitting the data.

The objective of this study is to evaluate the application of BNN models for predicting motor vehicle crashes. To accomplish this objective, a series of models was estimated using data collected on rural frontage roads in Texas. Three types of models were compared: BPNN, BNN and the negative binomial (NB) regression models. The results of this study show that in general both types of neural network models perform better than the NB regression model in terms of data prediction. Although the BPNN model can occasionally provide better or approximately equivalent prediction performance compared to the BNN model, in most cases its prediction performance is worse than the BNN model. In addition, the data fitting performance of the BPNN model is consistently worse than the BNN model, which suggests that the BNN model has better generalization abilities than the BPNN model and can effectively alleviate the over-fitting problem without significantly compromising the nonlinear approximation ability. The results also show that BNNs could be used for other useful analyses in highway safety, including the development of accident modification factors and for improving the prediction capabilities for evaluating different highway design alternatives.
© 2007 Elsevier Ltd. All rights reserved.

Keywords: Artificial neural networks; Bayesian methods; Negative binomial; Statistical model; Frontage roads

1. Introduction

Statistical or crash prediction models have frequently been used in highway safety studies. They can be used to identify major contributing factors or establish relationships between crashes and explanatory variables, such as traffic flows, types of traffic control, and highway geometric variables, with the aim that effective countermeasures could be implemented to reduce the number and severity of motor vehicle collisions occurring on different types of highway entities. The models can also be utilized to predict crash frequencies on sites that have not been used for estimating the original models or with different traffic flow and highway geometric conditions. The predicted results could be used in costs/benefit analyses and, if the predicted values are reliably estimated, could greatly help allocate the limited funds available to improve highway safety via the proper identification of hazardous sites (Hauer, 1996; Hauer et al., 2004a; Miaou and Song, 2005).

Previous studies that documented the development and application of crash prediction models have usually focused on statistical regression techniques. Most of these techniques are based on the generalized linear modeling (GLM) framework. GLMs have been very popular because they have explicit theoretical foundations and can produce interpretable coefficients for each explanatory variable included in the model. In addition, this modeling framework can be easily estimated using

* Corresponding author. Tel.: +1 979 458 3949; fax: +1 979 845 6481.
 E-mail addresses: ycxie@tamu.edu (Y. Xie), d-lord@tamu.edu (D. Lord), yzhang@civil.tamu.edu (Y. Zhang).
 [1] Tel.: +1 979 862 3553.
 [2] Tel.: +1 979 845 9902; fax: +1 979 845 6481.

commercial statistical software programs, such as SAS (SAS, 2002) or Genstat (Payne, 2000) among others. Hierarchical Bayes models have also been proposed for modeling motor vehicle collisions (Schluter et al., 1997; Tunaru, 2002; Miaou and Lord, 2003; Qin et al., 2005; Miaou and Song, 2005). These models have been found to offer superior statistical properties compared to GLMs when crash data are subjected to low sample mean values and small sample size (Lord, 2006; Lord and Miranda-Moreno, 2006).

Over the last few years, researchers in various fields of research, including highway safety, have proposed the use of neural network models for modeling the phenomenon under study (Mussone et al., 1999; Abdelwahab and Abdel-Aty, 2002; Riviere et al., 2006). For the application of neural networks, the entire dataset is usually divided into two subsets, defined as training and testing sets, respectively. Neural networks are trained using the data included in training dataset such that the underlying relationship between crash frequency and explanatory variables can be established. The testing dataset is then used to evaluate the performance of the trained neural network models. One important criticism that has been raised about these models is that although conventional neural network models, such as back-propagation neural networks (BPNN), can fit the training data with high precision, when it comes to prediction, they may produce predicted values with unacceptable variances (MATLAB, 2006). One of the major reasons causing this phenomenon is over-fitting. Neural network models that suffer from the over-fitting problem generally have poor generalization ability, which limits their applicability for crash predictions, even though they may possess better linear and nonlinear approximation abilities than statistical regression methods.

The objective of this study is to evaluate the application of Bayesian neural network (BNN) models for predicting motor vehicle crashes. BNN models can effectively reduce the over-fitting phenomenon while still keep the strong nonlinear approximation ability of neural networks (Marzban and Witt, 2001). To accomplish the objective of this study, a series of models is estimated using data collected on rural frontage roads in Texas. Three types of models, BPNN, BNN and the negative binomial (NB) (or Poisson-gamma) models, are estimated and their performances are compared using criteria employed for assessing the fit and prediction of models (Oh et al., 2003). Furthermore, a sensitivity analysis is also performed to illustrate how to extract the underlying relationship between crash frequency and explanatory variables from the trained neural networks. The results of this study will show that in general both neural network models perform better than the NB regression model in terms of data fitting and prediction. Although the BPNN model can sometimes provide better predictions than the BNN model, in most cases its data fitting and prediction performances are worse than the BNN model, which suggests that the BNN model effectively alleviates the over-fitting problem without significantly sacrificing its nonlinear approximation ability. The results will also show that BNNs could be used for useful analyses in highway safety, including the development of accident modification factors (AMFs) and for improving the prediction capabilities for evaluating different highway design alternatives.

This paper is divided into seven sections. The second and third sections provide a description on statistical regression techniques, BPNN, and BNN models. The fourth and fifth sections cover the data collection activities, and the modeling effort carried out in this work. The sixth section presents the results of the model comparison and the sensitivity analysis. The last section summarizes the key results and conclusions of the study and provides recommendations for further studies.

2. Background

Most traffic crash prediction models (sometimes referred to as safety performance function or SPF) have been based on statistical regression techniques. Initially, linear regression models were the model of choice for modeling motor vehicle collisions (Okamoto and Koshi, 1989; Miaou and Lum, 1993). Linear models, when not corrected for unequal variance, assume that the number of crashes follows a normal distribution. It has been reported in the literature that these models are inadequate for modeling count data, since crash data exhibit non-constant variance (Miaou et al., 1996).

Due to the inadequacy of linear regression models for analyzing discrete, nonnegative, sporadic, and asymmetrically distributed random events (Miaou and Lum, 1993), GLMs have been proposed for modeling crash frequency. Among GLMs, several types of models have been used by researchers, including Poisson regression (Miaou and Lum, 1993; Miaou, 1994; Maher and Summersgill, 1996; Oh et al., 2004, 2006; Lord et al., 2005b), Poisson-gamma or NB regression (Miaou, 1994; Shankar et al., 1995; Maher and Summersgill, 1996; Milton and Mannering, 1998; Persaud et al., 2002; Hiselius, 2004; Oh et al., 2004, 2006; Lord et al., 2005b; Donnell and Mason, 2006; Lord, 2006), Gamma regression model (Oh et al., 2006), and other variations of the NB regression model (Chin and Quddus, 2003; Miaou and Lord, 2003; Lord et al., 2005a; El-Basyouny and Sayed, 2006). It is generally agreed that when the sample variance is significantly greater than sample mean, NB models should be used in lieu of Poisson regression models. On the other hand, if the sample variance is significantly smaller than the sample mean, which is defined as underdispersion, gamma models are the models of choice (Oh et al., 2006). Zero-inflated models (Poisson and NB) have been proposed for modeling crash data with an apparent excess of zero observations (Shankar et al., 1997; Lee and Mannering, 2002; Kumara and Chin, 2003), but their application has been discredited when the characteristics and the nature of the data do not warrant the application of such models (Lord et al., 2005b, 2007; Warton, 2005). More recently, Bayesian multivariate generalized linear mixed model (Song et al., 2006), and multivariate Poisson regression model (Miaou and Song, 2005; Ma and Kockelman, 2006; Park and Lord, 2006) have also been proposed for modeling motor vehicle collisions.

All regression-based models such as the ones described above share one common characteristic: they need a well-defined function relating the dependent variable (crash frequencies) to

the independent (explanatory) variables. This function is often referred to as the "rate function" or "functional form" in the traffic safety literature. The specification of the functional form can significantly affect the goodness-of-fit of GLMs (Miaou and Lord, 2003). The functional form is usually estimated via a trial and error process based on the transportation safety analyst's experience, and can seldom be completely optimized. Normally, the functional form depends on the nature of the data and its selection should be based on the combination of statistical and logical properties linking the crash data to the covariates of the model (Miaou and Lord, 2003).

Compared to statistical regression models, the application of neural network models for crash data modeling has received much less attention. The primary reason is attributed to the complexity for estimating these models. Other criticisms that have impeded on their use include the following (Vogt and Bared, 1998):

(1) over-fitting when the sample size is small;
(2) unlike regression models, neural network models essentially work as black-boxes and do not generate interpretable parameters for each explanatory variable.

For the first criticism, it has been reported that similar to neural network models many regression models also suffer from the over-fitting problem (Marzban and Witt, 2001). To respond to the criticism that neural network models work as black-boxes, Fish and Blodgett (2003) and Delen et al. (2006) proposed a sensitivity analysis approach to quantify the effect of each input variable on the network output. Despite of these disadvantages, neural network models have some significant advantages over statistical regression models. First, neural network models do not require the establishment of a functional form. Statistical regression models, on the other hand, have to specify an appropriate functional form linking the dependent variable and independent variables (*Note*: the perfect functional form is unknown). Second, research has shown that standard multilayer feed-forward neural network models can approximate any continuous function defined on a compact set with arbitrary accuracy given enough hidden neurons are used (Hornik et al., 1989), though this strong ability may sometimes lead to over-fitting.

To avoid the over-fitting problem and improve the generalization ability of neural network models, a number of approaches have been proposed in the literature. One of these approaches includes adding a weight-decay or regularization term in the estimation process (Marzban and Witt, 2001; Liang, 2005). However, Marzban and Witt (2001) discussed that this improvement of generalization of neural network models impedes on their nonlinear approximation ability. On the other hand, they noted that the Bayesian inference method can improve the neural networks generalization ability without compromising the nonlinearity properties. The development of BNN was first initiated by Mackay (1992) and further developed by Neal (1995). Based on the previous BNN models, Liang (2005) introduced an improved BNN model by incorporating a prior on both the network connections and the weights. This modification gives the network more flexibility for choosing hidden neurons and input

variables. In Liang's study, the proposed BNN model was trained using an Evolutionary Monte Carlo (EMC) algorithm and was compared to a number of popular models such as the BPNN and the Box–Jenkins model for nonlinear time series forecasting. The testing results showed that the proposed BNN model consistently outperformed other prediction methods. Although BNN models began to gain popularity in late 1990s and have been used even more since 2000, only one application of BNN models in traffic safety has been identified (Riviere et al., 2006), and so far BNN models have not been utilized for modeling crash frequency. In view of the aforementioned advantages of neural network models over statistical regression models combined with the improvement made by incorporating Bayesian inference to neural networks, it is of valuable interest to investigate whether BNN models can be used to efficiently model motor vehicle crashes and whether they perform better than statistical regression models and conventional neural network models for predicting values.

3. Methodology

The most widely used statistical model in highway safety remains the NB regression model (when the data exhibit over-dispersion), and it has been used as a benchmark by many researchers. Consequently, the NB regression model was used in this study and compared with the BNN model proposed by Liang (2005). Besides, a BPNN model was also estimated using the same data for comparison purposes as BPNNs are one of the most commonly used neural network models. In the following sections, the characteristics of the NB regression, BPNN, and BNN models are briefly described. Details on the theory behind the BNN model and its training algorithm are presented in Liang (2005).

3.1. Negative binomial regression models

Let $\{(x_1, y_1), \ldots, (x_i, y_i), \ldots, (x_n, y_n)\}$ be a set of collected accident data. x_i is a vector consisting of the accident related characteristics of site i, and y_i represents the number of crashes reported at site i. A typical NB regression model is given by the following (Miaou, 1994):

$$\text{Prob}(Y_i = y_i) = \frac{\Gamma(y_i + \phi)}{\Gamma(y_i + 1)\Gamma(\phi)}\left(\frac{\mu_i}{\mu_i + \phi}\right)^{y_i}\left(\frac{\phi}{\phi + \mu_i}\right)^{\phi} \quad (1)$$

$$\text{expectation of } Y_i \text{ is } E(Y_i) = \mu_i = g(x_i) \quad (2)$$

$$\text{variance of } Y_i \text{ is } \text{Var}(Y_i) = \mu_i + \frac{\mu_i^2}{\phi} \quad (3)$$

where the Y_i is the identically independent variable following a NB distribution, $g(x_i)$ the functional form of the NB regression model, $i = 1, 2, \ldots, n$; n the total number of observations, and ϕ is the inverse dispersion parameter (assumed to be fixed in this study, see Miaou and Lord (2003) and Mitra and Washington (2007) about this assumption).

Although there are numerous NB regression models that have been used to model crash frequency, most of them use the same

probability density function (Eq. (1)) (*Note*: other parameterization exists, see Cameron and Trivedi, 1998) and only the functional form varies from model to model. Usually, the functional forms are built empirically and the best one is selected using a trial and error process (often based on the judgment of the safety analyst and through some statistical tools, such as the goodness-of-fit, etc.). For this study, the following functional form was used:

$$g(x_i) = \beta_0 \left(\frac{365 \times F_i L_i \, \text{off}_i}{1,000,000} \right) \exp\{\beta_1 \, \text{LW}_1 + \beta_2 \, \text{RS}_i\} \qquad (4)$$

where $x_i = [F_i, L_i, \text{LW}_i, \text{RS}_i, \text{off}_i]$, F_i is the ADT for segment i (veh/day), L_i the length of segment i (mile), LW_i the lane width of segment i (ft), RS_i the right shoulder width of segment i (ft), off_i the offset of segment i, describing the number of years during which the accident data y_i was collected (in this study, the variable off_i was equal to 5), and β_0, β_1, β_2 are the regression coefficients to be estimated.

3.2. BP neural networks

Fig. 1 shows the structure of a typical BPNN model that was used for modeling crash frequency in this study. The transfer function for the hidden layer, *fh*, was chosen to be a tangent (tanh) function and a linear function was used as the transfer function for the output layer, *fo*. Again, let $\{(x_1, y_1), \ldots, (x_i, y_i), \ldots, (x_n, y_n)\}$ be a set of collected accident data. The prediction result $\hat{y}_i$ using this BPNN structure was given by the following equation:

$$\hat{y}_i = f(x_i, \psi) = b_2$$
$$+ \sum_{j=1}^{M} \left\{ w_2(j) \tanh \left(\sum_{k=1}^{P} w_1(j,k) x_{ik} + b_1(j) \right) \right\} \qquad (5)$$

where P is the number of input neurons, M the number of hidden neurons, $b_1(j)$ and b_2 the biases, $w_2(j)$ the weights connecting hidden layer and output layer, $w_1(j,k)$ the weights connecting input layer and hidden layer, x_{ik} the kth element of the ith input, $x_i = [x_{i1}, \ldots, x_{ik}, \ldots, x_{iP}]$, the ith input, Ψ a vector contains all

Fig. 1. A typical single-output BPNN with single hidden layer.

the network parameters ($b_1(j)$, b_2, $w_1(j,k)$, and $w_2(j,k)$), $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, M$, and $k = 1, 2, \ldots, P$.

To render the BPNN model comparable to the NB regression model, the input dimension to the BPNN model was set to 4, which describes the ADT, segment length, lane width, and right shoulder width of each site, respectively. The number of neurons in the output layers was set to 1 and the output is the predicted number of crashes for each site. The BPNN was built and trained using the MATLAB neural network toolbox (MATLAB, 2006). A standard method for training BPNNs is to minimize the error term shown in the following equation:

$$E = \frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i)^2 \qquad (6)$$

As discussed in the background section, a regularization or weight-decay term is often added to Eq. (6) to improve the generalization ability of BPNNs. In this study, the BPNN model was trained by minimizing Eq. (7) (MATLAB, 2006):

$$E_r = \eta \frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i)^2 + (1 - \eta) \frac{1}{n_p} \sum_{i=1}^{n_p} (\psi_i)^2 \qquad (7)$$

where n_p is the number of network parameters, including weights and bias, ψ_i the ith element in the network parameter vector, and η is the performance ratio.

3.3. Bayesian neural networks

The BNN model used in this study was initially proposed by Liang (2003, 2005). In the BNN model, Dr. Liang used a fully connected multilayer feed-forward network structure with one hidden layer. The simplified network structure is illustrated in Fig. 2. For the BNN model, the transfer functions used in the hidden layer and the output layer are the same as those used in the BPNN model.

Although the network structure of the proposed BNN model is very similar to the BPNN structure, they are different in the prediction mechanism and the training process. First, an example is given to illustrate the differences in the prediction mechanism. Assume there are n sets of accident data $\{(x_1, y_1), \ldots, (x_i, y_i), \ldots, (x_n, y_n)\}$, where the definitions of x_i and y_i are the same as what we used for the NB regression and BPNN models. Let θ denotes all the network parameters or weights, β_j, α_k, and γ_{jk} ($j = 1, \ldots, M$; $k = 0, \ldots, P$), in Fig. 2. The predicted number of accidents for site i using BNNs is given by the following equation (Neal, 1995):

$$\hat{y}_i = \int f_B(x_i, \theta) P(\theta | (x_1, y_1), \ldots, (x_n, y_n)) \, d\theta \qquad (8)$$

where $f_B(x_i, \theta)$ is defined as:

$$f_B(x_i, \theta) = \alpha_0 + \sum_{k=1}^{P} (\alpha_k x_{ik})$$
$$+ \sum_{j=1}^{M} \left\{ \beta_j \tanh \left(\sum_{k=1}^{P} \gamma_{jk} x_{ik} + \gamma_{j0} \right) \right\} \qquad (9)$$
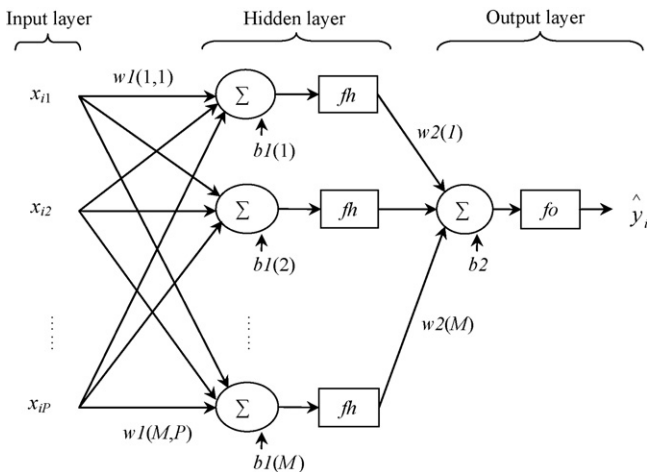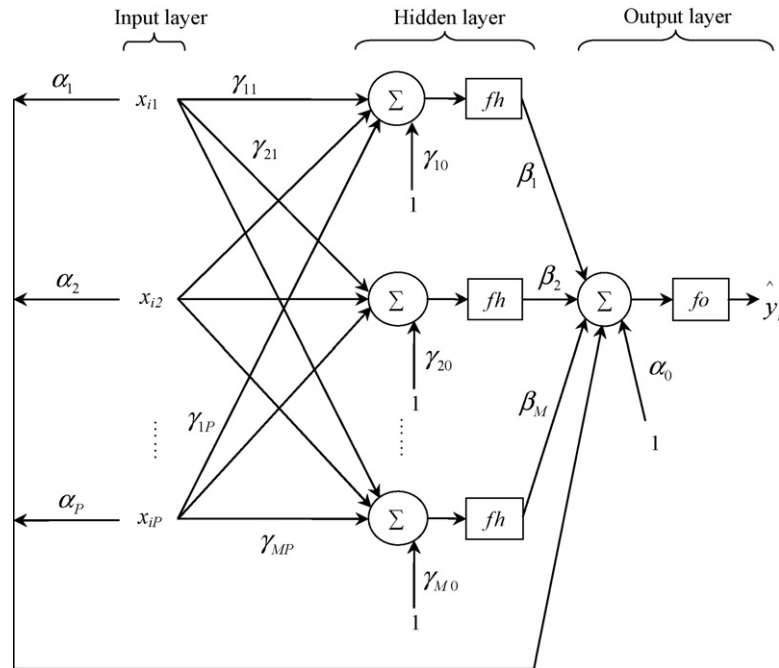
Fig. 2. A fully connected multilayer feed-forward neural network.

$P(\theta|(x_1, y_1), \ldots, (x_n, y_n))$ in Eq. (8) is the posterior distribution of θ given observed data $\{(x_1, y_1), \ldots, (x_n, y_n)\}$. One can see the main difference between BNNs and BPNNs is that for BPNNs the network parameter ψ is fixed; while for BNNs the network parameter θ follows a certain probability distribution, and the prediction process for BNNs is to evaluate the integral of $f_B(x_i, \theta)P(\theta|(x_1, y_1), \ldots, (x_n, y_n))$ over all possible values of θ as shown in Eq. (8).

The actual BNN model is more complicated than the example given above. Readers are referred to Liang (2003, 2005) for a more detailed description of the BNN model and its EMC training algorithm. For the curious readers, Appendix A provides a brief summary of the theory behind the BNN model proposed by Liang (2005).

4. Characteristics of data

In order to compare the models evaluated in this work, data collected for a research project related to estimating the safety performance of rural frontage roads in Texas was used in this

Fig. 3. Crash count distribution for the 88 rural frontage road segments in Texas.

Table 1
Descriptive statistics of the explanatory variables

	Length (mile)	ADT (vpd)	Right shoulder width (ft)	Lane width (ft)	Number of crash in 5 years
Min	0.69	110	0	9	0
Max	5.34	6400	9	13	6
Mean	2.16	939	1.38	10.67	1.39
S.D.	0.99	1186	2.12	0.84	1.28

study (Lord and Bonneson, 2006). In this dataset, there were 88 sites consisting of rural two-way frontage roads located in central Texas. During the 5-year period, 122 crashes occurred on the study sites (all severities). Fig. 3 illustrates the distribution of the crash counts for the 88 sites. This figure shows that no crash occurred at 28 sites (nearly 32%) during the 5-year period. The mean and standard deviation (S.D.) of the sample crash data are 1.39 and 1.28, respectively. This implies that a NB regression model is more suitable than a Poisson regression model. Table 1 summarizes the descriptive statistics for the explanatory variables.

5. Implementation of models

The NB model was estimated using SAS (SAS, 2002). To implement the BPNN and BNN models, the number of hidden neurons (M) needs to be decided. For the BPNN model, the performance ratio (η) also needs to be specified. A commonly used method to choose the two parameters is cross-validation, which was used in this study. The data was randomly separated into two parts, one is for training and the other is for testing. The training part consists of approximately 3/4 of the total data. Different numbers of hidden neurons and performance ratios
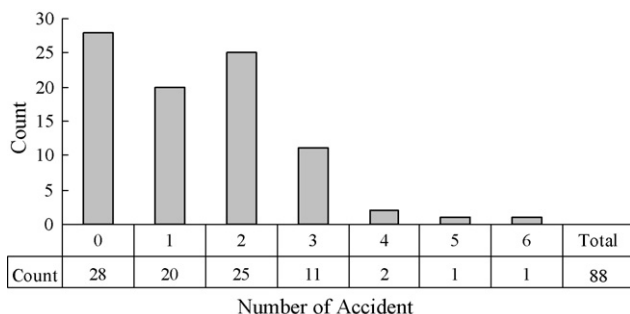
were tested, and those correspond to the lowest testing errors (defined similarly as Eq. (6)) were chosen.

For the BPNN model, $M = 3, 4, \ldots, 9, 10$ and $\eta = 0.05, 0.1, 0.15, \ldots, 0.9, 0.95$ were tested, and the best number of hidden neurons and performance ratio were chosen to be 8 and 0.75, respectively. For the BNN model, $M = 3, 4, \ldots, 9, 10$ were tried, and the best number of hidden neurons was chosen to be 5.

The BPNN was built and trained using the MATLAB neural network toolbox (MATLAB, 2006). The number of training epochs was set to be 10,000; the learning rate was 0.05; the training goal was set to be 0.0001. Default settings were used for the remaining parameters. The training of the BNN model is quite different from the BPNN model, and parameters suggested in Liang (2005) were used. It is well known that for neural network models multiple runs may produce different results. Thus for each scenario in this study, the neural network models were run 10 independent times, and their average performances were used for comparison.

6. Results of analysis

This section is divided into two parts. The first part describes the results of the analyses for the prediction performance of the three models. The second part presents the results of the sensitivity analyses for the BNN model.

6.1. Comparison of predictive performance

To evaluate the effects of sample sizes of the training datasets on model performances, three scenarios were evaluated. Table 2 shows the characteristics of the three scenarios; the data were divided into training and testing subsets. It should be pointed out that the sites were randomly selected in each dataset.

Two evaluation criteria as proposed by Oh et al. (2003) were adopted to compare both the prediction and fitting performances of the three models. These evaluation criteria are described in the following equations:

$$\text{mean absolute deviation (MAD)} = \frac{1}{n} \sum_{i=1}^{n} |\hat{y}_i - y_i| \qquad (10)$$

$$\text{mean squared prediction error (MSPE)} = \frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i)^2 \qquad (11)$$

In these equations, $\hat{y}_i$ and y_i are the predicted and observed values, respectively, and n is the size of training or testing subsets. Values closer to zero indicate better model performance for both evaluation criteria. MAD is used to estimate the prediction deviation. MSPE is employed for determining the variance of the difference between predicted and observed results.

Table 2
Sample size of training and testing datasets

Scenarios number	Training set size	Testing set size
1	60	28
2	70	18
3	80	8

Table 3 shows the modeling results for the NB regression model. This table shows that most of the estimated coefficients have large *p*-values that are insignificant; the magnitude of the coefficients varies significantly (*Note*: all the coefficients have the proper sign, e.g., larger shoulder widths are usually associated with a reduction in run-off-the-road collisions, see Hughes et al., 2004) between different training set sizes, for example, from training set size 70 to 80, the estimated coefficients for lane width and intercept have more than doubled. This instability is attributed to the problems associated with low sample mean (LSM) values and small sample size (SSS). These problems affect the dispersion parameter and, consequently, the estimation of the confidence intervals (i.e., standard errors) for each coefficient in the model (see Lord, 2006; Lord and Miranda-Moreno, 2006; Zhang et al., 2006).

The training and testing performances of the three models are summarized in Table 4. For all training set sizes, the training and testing performances of the NB models did not perform very well compared to the other two models. This seems to support that neural network models can better approximate nonlinear functions. In addition, the output of the NB regression model shows that it may also suffers from over-fitting problem especially for the training size equal to 70 and 80. This is consistent with the discussions in a previous work by Marzban and Witt (2001).

Table 4 also shows that for the training process, the MAD and MSPE are consistently the lowest for all training sizes for the BNN model. This means that the BPNN model has the best training performance. Both the BPNN and BNN models outperform the NB model when the MAD and MSPE values are evaluated for all testing data sets. For training set size equal to 60, the BNN model performs better than the BPNN model for both training and testing. While for training set size equal to 70, the BNN model slightly underperforms the BPNN model in terms of testing performance. When the training set size is equal to 80, both neural network models perform approximately the same in terms of testing MAD and MSPE values, but the BNN model has better training performance.

To further compare the BPNN and BNN models, the previous test was repeated independently for another three times based on three randomly selected training data. For simplicity, the parameter estimation results of the NB model were omitted. Only the MAD and MSPE values of the three models are shown in Table 5. It can be seen that for almost all cases the neural network models perform better than the NB model for both training and testing. The only exception is the training set size 80 for the additional data set 1, where the BPNN model underperforms the NB model in terms of testing MAD and MSPE. This again suggests that in general neural networks have better nonlinear approximation ability. The values in Table 5 also show that in all cases the BNN model produces better training and testing results than the BPNN model. Although in some cases the testing MADs of the BPNN and BNN models are the same (e.g., training set size 60 for the additional data sets 2 and 3), the corresponding MSPE and training MAD values of the BNN model are smaller than those of the BPNN model. This finding is consistent with the discussions in Marzban and Witt (2001). They argued

Table 3
Modeling results of the NB regression models

Training set size	Parameter	Estimate	Standard error	Wald 95% confidence limits		$P > \chi^2$	Log likelihood
60	Intercept	0.5781	1.7987	−2.9473	4.1034	0.7479	−54.8055
	Lane width	−0.1020	0.1732	−0.4413	0.2374	0.5560	
	Right shoulder	−0.1742	0.0719	−0.3152	−0.0332	0.0154	
	Dispersion parameter[1]	0.1367	0.1541	−0.1653	0.4387	–	
70	Intercept	0.3078	1.5454	−2.7210	3.3367	0.8421	−63.1060
	Lane width	−0.0795	0.1489	−0.3714	0.2124	0.5935	
	Right shoulder	−0.1585	0.0640	−0.2838	−0.0331	0.0133	
	Dispersion parameter[1]	0.0890	0.1259	−0.1578	0.3358	–	
80	Intercept	1.3469	1.5450	−1.6812	4.3750	0.3833	−77.7407
	Lane width	−0.1760	0.1483	−0.4668	0.1147	0.2353	
	Right shoulder	−0.1228	0.0629	−0.2461	0.0005	0.0509	
	Dispersion parameter[1]	0.1748	0.1412	−0.1019	0.4516	–	

Note—1: the dispersion parameter is shown as non-significant. This is caused by the problems associated with small sample size and low sample mean. In reality, there should be a dispersion parameter given the characteristics of the data which clearly showed overdispersion, but cannot be captured by the model. See Lord (2006) for a detailed discussion about the characteristics associated with these two issues.

Table 4
Performances comparison of NB, BPNN, BNN models

Training set size	MOEs	Training (fitting)			Testing (predicting)		
		NB	BPNN	BNN	NB	BPNN	BNN
60	MAD	0.99	0.85	0.82	1.28	1.00	0.97
	MSPE	1.76	1.17	1.05	2.96	1.42	1.35
70	MAD	0.96	0.82	0.78	1.58	1.15	1.18
	MSPE	1.63	1.10	0.95	4.24	1.75	1.79
80	MAD	1.04	0.85	0.80	1.86	1.41	1.42
	MSPE	2.04	1.12	0.98	6.53	2.52	2.48

Table 5
Additional performances comparison of NB, BPNN, BNN models

Additional data set	Training set size	MOEs	Training (fitting)			Testing (predicting)		
			NB	BPNN	BNN	NB	BPNN	BNN
	60	MAD	1.20	0.94	0.91	0.98	0.87	0.84
		MSPE	2.89	1.34	1.23	1.82	1.00	0.96
1	70	MAD	1.16	0.92	0.89	0.85	0.79	0.75
		MSPE	2.68	1.33	1.20	0.94	0.83	0.77
	80	MAD	1.14	0.92	0.88	0.80	0.86	0.73
		MSPE	2.56	1.28	1.14	0.87	1.02	0.80
	60	MAD	1.19	0.91	0.88	1.05	0.87	0.87
		MSPE	2.90	1.23	1.13	2.00	1.20	1.15
2	70	MAD	1.14	0.88	0.84	1.11	0.99	0.95
		MSPE	2.70	1.17	1.06	1.81	1.49	1.36
	80	MAD	1.14	0.88	0.84	1.40	1.17	1.04
		MSPE	2.73	1.17	1.06	2.57	2.05	1.74
	60	MAD	1.21	0.94	0.91	1.42	0.86	0.86
		MSPE	2.90	1.33	1.20	4.94	1.07	1.05
3	70	MAD	1.22	0.93	0.89	1.32	0.90	0.83
		MSPE	2.99	1.29	1.16	4.33	1.26	1.05
	80	MAD	1.10	0.91	0.87	1.35	0.84	0.76
		MSPE	2.22	1.26	1.12	5.26	1.15	0.96

that the inclusion of a regularization term into the BPNN models can impede on their nonlinear approximation (data fitting) ability. In our case, the better testing performance of the BPNN model is achieved at the cost of training or fitting performance. Marzban and Witt (2001) further noted that Bayesian inference has the potential to improve the generalization ability of neural networks without compromising their nonlinear approximation ability. This conclusion is supported by the training and testing performances of the BNN model shown in Tables 4 and 5.

The computation times of each model were also compared. All the computations were carried out on a desktop computer with Pentium(R) 3.00 GHz CPU and 512 MB memory. For the NB model, the computation time for any single run was less than 1 s; for the BPNN model, the average computation time for a single run was approximately 26 s; for the BNN model, the average computation time for a single run was around 28 s. Although both neural network models require significantly more computation resources than the NB model, this computation requirement can be handled easily by an ordinary desktop computer. Taking into account of the better fitting and predicting ability of neural network models, it is feasible and desirable to apply neural network models, especially the BNN models, to traffic safety forecasting studies.

6.2. Sensitivity analysis of the BNN model

Neural network models have been long criticized for not being able to generate interpretable parameters for each explanatory variable, and this is one of the major reasons that few neural network models have been used for modeling crash frequency. To minimize this problem, a method proposed by Fish and Blodgett (2003) was adopted in this study to analyze the sensitivity of each explanatory variable. This method has also been used by Delen et al. (2006) in their application of neural network models in accident injury severity study.

The basic idea of this method is that for each explanatory variable, one keeps all other explanatory variables unchanged and perturbs the current variable's value within a reasonable interval. At the same time, the corresponding variation of network output is recorded, and from this variation, one can find the effect of changing single explanatory variable on the network output. The idea of this method is simple, but it can be useful to minimize the black-box problem and help illustrate the training result of neural networks. It should be pointed out that the explanatory variables may not be independent of each

Table 6
Data used for the sensitivity analysis

Site ID	Length (mile)	ADT (vpd)	Right shoulder width (ft)	Lane width (ft)	Crash count
14	2.40	6168	4	11	3
88	1.15	428	0	10	1

other. Due to the complicated relationship between crash frequency and all explanatory variables, if one changes the value of any of the remaining explanatory variables, the relationship between crash frequency and the current explanatory variable may change accordingly.

Two sites were chosen for the sensitivity analysis as shown in Table 6. These two sites have different segment lengths, ADTs, right shoulder and lane widths. In this case, one can analyze the sensitivity of each explanatory variable under different conditions and compare them accordingly. The BNN model was estimated using an 80-observation training sample.

Figs. 4–7 show the results of the sensitivity analysis, from which the following observations can be drawn:

(1) From Fig. 4, one can see that there is an approximate linear relationship between crash frequency and segment length. The slopes of these two linear relationships are very similar, but site 14 has a significantly larger intercept that can probably be explained by its higher ADT. This outcome shows that segment length can be used as an offset variable rather than as an explanatory variable.
(2) Fig. 5 shows that the relationship between crash frequency and ADT can be established using a polynomial function. The two polynomial functions shown in Fig. 5 are almost the same except for the constant term; this relationship is similar to what has been reported in the literature (see Hauer, 1997). A reasonable explanation for the difference in the constant term is that site 14 is longer than site 88. Thus, for the same ADT, site 14 has higher crash exposure than site 88.
(3) Fig. 6 shows an interesting difference in the relationship between crash frequency and lane width. Site 14 has a near linear relationship while site 88 has a polynomial relationship. The figure suggests that increasing the lane width of site 88 from 10 to 11 ft may not be as useful as increasing the lane width for the same values at site 14. In addition, it shows that widening lanes at site 88 from 9 to 10 ft may not be desirable. This result appears to be counterintuitive,
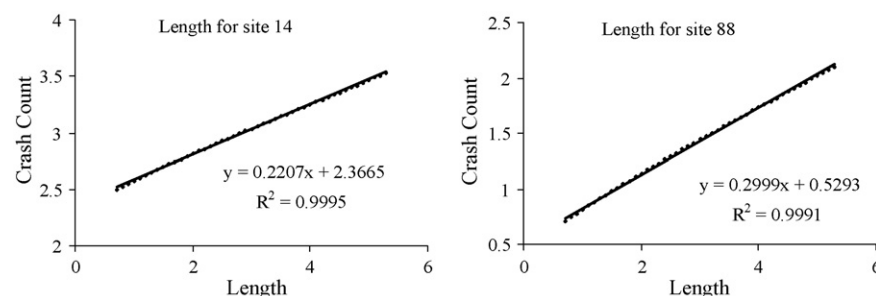

Fig. 4. Sensitivity analysis for the variable segment length for sites 14 and 88.
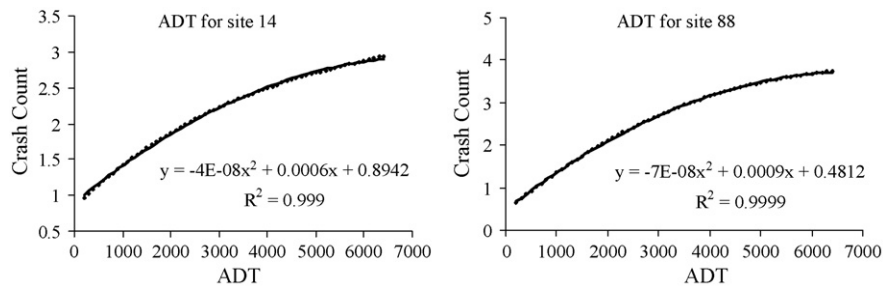
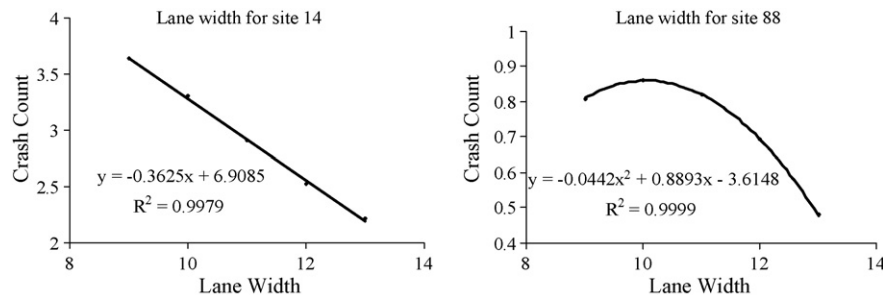Fig. 5. Sensitivity analysis for the variable ADT for sites 14 and 88.

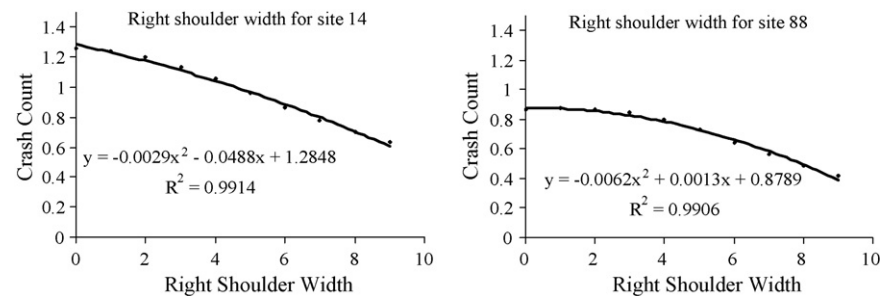Fig. 6. Sensitivity analysis for the variable lane width for sites 14 and 88.

Fig. 7. Sensitivity analysis for the variable right shoulder width for sites 14 and 88.

but may not be unusual given the wide variety relationships found by Hauer (2000) on this topic.

(4) Fig. 7 shows the relationship between right shoulder width and crash frequency. Both sites have similar polynomial relationships.

An interesting point about the sensitivity analysis is that it could also be used to develop AMFs. With BNN models, an AMF could be developed for each site individually or could be estimated for the entire dataset. The sensitivity analysis shows that the relationship between crashes and some explanatory variables follows a nonlinear function. With statistical regression models, the relationship between these variables can only follow an exponential function (e.g., $e^{\beta_i x_i}$ in GLMs) (Lord and Bonneson, 2006; Bonneson et al., 2006). However, the relationship could, in fact, possibly follow a nonlinear function without any specified form. As reported by Hauer et al. (2004b), a nonlinear function was established between crashes and different explanatory variables included in the statistical models they produced for

estimating the safety performance of four-lane urban highways. Another potential application of the sensitivity analysis is to help transportation safety analyst find appropriate functional forms to be used in NB models. Despite the initial findings described above, further work is needed to examine potential avenues for developing AMFs using BNN models.

7. Summary and conclusions

The objective of this study was to evaluate the application of BNN models for predicting motor vehicle crashes on transportation networks. This paper first described the fundamental principles of NB regression models commonly used in highway safety and the characteristics of neural network models. The review has shown that although neural network models have excellent function approximation abilities and do not require specifying a functional form linking the dependent variable to the explanatory variables, the over-fitting problem has significantly limited their application in highway safety.

By incorporating the Bayesian inference theory into neural network models, the over-fitting problem can technically be reduced, which makes them more suitable for modeling crash data.

In this study, the BNN and BPNN models were compared to the NB regression model. The models were applied to crash data collected on rural two-lane frontage roads located in central Texas. The dataset included 88 highway sections. Three sizes were used for training the three types of models: 60, 70, and 80, and the remaining data were used to evaluate the prediction capabilities. The MAD and MSPE were used to measure the performance of the three types of models.

The results of this study show that in most cases the NB model produced inferior prediction capabilities than the two neural network models according to the MAD and MSPE evaluation criteria, and its performance to fit the training data was also the poorest according to the MAD and MSPE criteria. Although the BPNN model sometimes can provide better prediction performance than the BNN model, in most cases its prediction performance was inferior to the BNN model based on the MAD and MSPE evaluation criteria. In addition, the fitting performance of the BPNN model was consistently worse than the BNN model, which suggests that the BNN model has better nonlinear approximation and generalization abilities than the BPNN model, and the over-fitting phenomenon has been alleviated without significantly compromising the nonlinear approximation ability. Based on the data and results from this study, BNN could be used effectively for predicting crash data.

In response to the criticism that neural network models are black-boxes and cannot produce interpretable parameters for each explanatory variable, the sensitivity analysis carried out in this study demonstrated how to extract the underlying relationship between crash frequency and explanatory variables from the trained neural networks. Although the sensitivity analysis method is empirical in nature and cannot be used equivalently as the statistical inference in NB models, it still is a useful tool that could be used for developing AMFs and helping transportation safety analysts find the best functional forms in regression modeling.

While the application of BNN models in this work offered positive results, it is suggested to apply this type of model to other datasets to validate the results obtained in this study. It is possible that when the training data size is large, the advantages of the BNN model over the BPNN model may be less significant. This is because large training data can generally better represent population characteristics. Thus, the chance of over-fitting may be reduced. However, for many transportation safety studies, it is difficult and costly to obtain large sample data. Therefore, BNNs still have a great potential to be widely used. Further work also includes evaluating these models when the data are characterized by both small sample mean values and small sample size (in terms of stability), the development of AMFs, and for comparing the safety performance of different highway design alternatives when the predictive accuracy is a controlling factor in the decision making process. The last two topics are closely tied to the upcoming Highway Safety Manual (Hughes et al., 2004). Finally, it is hoped that the outcome of this research will foster new research ideas for applying BNNs in highway safety.

Appendix A. Bayesian neural networks and evolutionary Monte Carlo algorithm

In the BNN model proposed by Liang (2005), a prior is used on both network connections and network parameters (weights). This means that both parameters are subject to certain probability distributions. Given the previous n samples, for certain network connection Λ and parameter θ, the output of the network, shown in Fig. 2 of the paper, can be defined as follows:

$$\hat{y}_i = f(x_i, \Lambda, \theta) = \alpha_0 I_{\alpha 0} + \sum_{k=1}^{P} x_{ik} \alpha_k I_{\alpha k}$$
$$+ \sum_{j=1}^{M} \beta_j I_{\beta j} \mathrm{fh} \left(\gamma_{j0} I_{\gamma j0} + \sum_{k=1}^{P} x_{ik} \gamma_{jk} I_{\gamma jk} \right) \qquad (A1)$$

where $I_\zeta = 0$, if the network link ζ is not connected, otherwise, $I_\zeta = 1$ (I_ζ represents $I_{\alpha 0}$, $I_{\alpha k}$, $I_{\beta j}$, $I_{\gamma j0}$ and $I_{\gamma jk}$ in Eq. (A1)). Λ denotes all the indicator functions I_ζ; $\alpha_k \sim N(0, s_\alpha^2)$, $k = 0, \ldots, P$ are the weights between the input layer, the bias, and the output layer with normal prior distribution; $\beta_j \sim N(0, s_\beta^2)$, $j = 1, \ldots, M$ are the weights between hidden layer and output layer; $\gamma_{jk} \sim N(0, \sigma_\gamma^2)$, $k = 0, \ldots, P$ and $j = 1, \ldots, M$ are the weights between the input layer, the bias, and the hidden layer; P is the input dimension; M is the maximum number of hidden neurons specified by the user. In addition, Liang (2005) assumed that $(y_i - \hat{y}_i) \sim N(0, \sigma^2)$, in which σ^2 is assumed to follow an inverse gamma distribution $IG(v_1, v_2)$. Based on the number of hidden neurons specified by the user, M, the number of connections for a fully connected network is given by $U = (M + 1) \times (P + 1) + M$. In the BNN proposed by Liang (2005), the number of connections is subject to a truncated Poisson prior distribution as shown in the following equation:

$$P(\Lambda) = \begin{cases} \dfrac{1}{Z} \dfrac{\lambda^m}{m!} & m = 3, \ldots, U \\ 0 & \text{otherwise} \end{cases} \qquad (A2)$$

where

$$Z = \sum_{m=3}^{U} \frac{\lambda^m}{m!}$$

Let θ denotes the following parameters $\{\alpha_k, \beta_j, \gamma_{jk}, \sigma^2\}$. Liang (2005) again assumed that all prior distributions are inde-

pendent. It can be shown that for given network connection Λ and parameter θ the prior distribution is:

$$P(((x_1, y_1), \ldots, (x_n, y_n))|\theta, \Lambda)$$

$$= \left\{ \prod_{i=1}^{n} \left[\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{1}{2\sigma^2}(y_i - \hat{y}_i)^2 \right\} \right] \right\}$$

$$\times \left\{ \frac{v_2^{v_1}}{\Gamma(v_1)} (\sigma^2)^{-v_1 - 1} \exp\left\{ \frac{-v_2}{\sigma^2} \right\} \right\}$$

$$\times \left\{ \prod_{k=0 \sim P}^{I_{\alpha k}=1} \left[\frac{1}{\sqrt{2\pi\sigma_\alpha^2}} \exp\left\{ -\frac{\alpha_k^2}{2\sigma_\alpha^2} \right\} \right] \right\}$$

$$\times \left\{ \prod_{j=1 \sim M}^{I_{\beta j}=1 \text{ and } \sum I_{\gamma jk}=1} \left[\frac{1}{\sqrt{2\pi\sigma_\beta^2}} \exp\left\{ -\frac{\beta_j^2}{2\sigma_\beta^2} \right\} \right] \right\}$$

$$\times \left\{ \prod_{j=1 \sim M}^{I_{\beta j}=1} \prod_{k=0 \sim P}^{I_{\gamma jk}=1} \left[\frac{1}{\sqrt{2\pi\sigma_\gamma^2}} \exp\left\{ -\frac{\gamma_{jk}^2}{2\sigma_\gamma^2} \right\} \right] \right\} \left\{ \frac{1}{Z} \frac{\lambda^m}{m!} \right\} \tag{A3}$$

And from Eq. (A3), it is easy to obtain the log-posterior distribution of θ and Λ as proposed by Liang (2005):

$$P(\theta, \Lambda|((x_1, y_1), \ldots, (x_n, y_n)))$$

$$= \text{constant} - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 - \frac{v_2}{\sigma^2} - \left(\frac{n}{2} + v_1 + 1 \right) \log \sigma^2$$

$$- \frac{1}{2} \sum_{k=0}^{P} I_{\alpha k} \left(\log \sigma_\alpha^2 + \frac{\alpha_k^2}{\sigma_\alpha^2} \right) - \frac{1}{2} \sum_{j=1}^{M} I_{\beta j} \delta \left(\sum_{k=0}^{P} I_{\gamma jk} \right)$$

$$\times \left(\log \sigma_\beta^2 + \frac{\beta_j^2}{\sigma_\beta^2} \right) - \frac{1}{2} \sum_{j=1}^{M} \sum_{k=0}^{P} I_{\beta j} I_{\gamma jk} \left(\log \sigma_\gamma^2 + \frac{\gamma_{jk}^2}{\sigma_\gamma^2} \right)$$

$$- \frac{m}{2} \log(2\pi) + m \log \lambda - \log(m!) \tag{A4}$$

Since, the log-posterior distribution of θ and Λ is now known, the prediction results based on this BNN model can be theoretically calculated using an equation similar to Eq. (8) in the main text. However, this requires the evaluation of an integral over all possible parameter θ and Λ, which can require extensive computer resources. In practice, some algorithms are used to obtain samples $(\theta_1, \Lambda_1), \ldots, (\theta_S, \Lambda_S)$ from the log-posterior distribution shown in Eq. (A4), and based on these samples an unbiased estimation of y_i is calculated using the following equation:

$$\hat{y}_i = \frac{1}{S} \sum_{j=1}^{S} f(x_i, \Lambda_j, \theta_j) \tag{A5}$$

The goal for training BNN models is to obtain $(\theta_1, \Lambda_1), \ldots, (\theta_S, \Lambda_S)$. Various algorithms can be used to train BNNs, such as the hybrid Monte Carlo (Neal, 1995), the tempered reversible jump MCMC (Liang, 2003), and the sequential Monte Carlo and the EMC (Liang, 2005 and references therein). Most of these algorithms are based on the Monte Carlo simulation

method. In this study, an EMC algorithm was used to sample from the log-posterior distribution shown in Eq. (A4). The EMC algorithm is fairly complicated and will not be described here. Interested readers are referred to Liang and Wong (2001) and Liang (2005) for a detailed description of this algorithm.

References

Abdelwahab, H.T., Abdel-Aty, M.A., 2002. Artificial neural networks and logit models for traffic safety analysis of toll plazas. Transport. Res. Rec. 1784, 115–125.

Bonneson, J.A., Zimmerman, K., Fitzpatrick, K., 2006. Road safety design synthesis. TTI Report 0-4703-P1. Texas Transportation Institute, College Station, TX.

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

Chin, H.C., Quddus, M.A., 2003. Applying the random effect negative binomial model to examine traffic accident occurrence at signalized intersections. Accid. Anal. Prev. 35 (2), 253–259.

Delen, D., Sharda, R., Bessonov, M., 2006. Identifying significant predictors of injury severity in traffic accidents using a series of artificial neural networks. Accid. Anal. Prev. 38 (3), 434–444.

Donnell, E.T., Mason, J.M., 2006. Predicting the frequency of median barrier crashes on Pennsylvania interstate highways. Accid. Anal. Prev. 38 (3), 590–599.

El-Basyouny, K., Sayed, T., 2006. Comparison of two negative binomial regression techniques in developing accident prediction models. In: Proceedings of the Transportation Research Board 85th Annual Meeting, Washington, DC (CD-ROM).

Fish, K.E., Blodgett, J.G., 2003. A visual method for determining variable importance in an artificial neural network model: an empirical benchmark study. J. Target. Meas. Anal. Market. 11 (3), 244–254.

Hauer, E., 1996. Identification of sites with promise. Transport. Res. Rec. 1542, 54–60.

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

Hauer, E., 2000. Lane width and safety. Draft document. Review of Literature for the Interactive Highway Safety Design Model, Toronto, Ont. (http://www.roadsafetyresearch.com/, accessed on December 14, 2006).

Hauer, E., Allery, B.K., Kononov, J., Griffith, M.S., 2004a. How to best rank site with promise. Transport. Res. Rec. 1897, 48–54.

Hauer, E., Council, F.M., Mohammedshah, Y., 2004b. Safety models for urban four-lane undivided urban segments. Transport. Res. Rec. 1897, 96–105.

Hiselius, L.W., 2004. Estimating the relationship between accident frequency and homogeneous and inhomogeneous traffic flows. Accid. Anal. Prev. 36 (6), 985–992.

Hornik, K., Stinchcombe, M., White, H., 1989. Multilayer feed-forward networks are universal approximators. Neural Netw. 2 (5), 359–366.

Hughes, W., Eccles, K., Harwood, D., Potts, I., Hauer, E., 2004. Development of a Highway Safety Manual. Appendix C: Highway Safety Manual Prototype Chapter: Two-Lane Highways. NCHRP Web Document 62 (Project 17-18(4)). Washington, DC (available at http://www.highwaysafetymanual.org/, accessed June 2006).

Kumara, S.S.P., Chin, H.C., 2003. Modeling accident occurrence at signalized tee intersections with special emphasis on excess zeros. Traffic Inj. Prev. 3 (4), 53–57.

Lee, J., Mannering, F., 2002. Impact of roadside features on the frequency and severity of run-off-roadway accidents: an empirical analysis. Accid. Anal. Prev. 34 (2), 149–161.

Liang, F., 2003. An effective Bayesian neural network classifier with a comparison study to support vector machine. Neural Comput. 15 (8), 1959–1989.

Liang, F., 2005. Bayesian neural networks for nonlinear time series forecasting. Stat. Comput. 15 (1), 13–29.

Liang, F., Wong, W.H., 2001. Real-parameter evolutionary Monte Carlo with applications to Bayesian mixture models. J. Am. Stat. Assoc. 96 (454), 653–666.

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. Accid. Anal. Prev. 38 (4), 751–766.

Lord, D., Bonneson, J.A., 2006. Estimating accident modification factors for rural frontage roads in Texas. In: Paper presented at the 86th Annual Meeting of the Transportation Research Board, Washington, DC.

Lord, D., Miranda-Moreno, L.F., 2006. 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. Working Paper. Zachry Department of Civil Engineering, Texas A&M University, College Station, TX.

Lord, D., Manar, A., Vizioli, A., 2005a. Modeling crash-flow-density and crash-flow-V/C ratio for rural and urban freeway segments. Accid. Anal. Prev. 37 (1), 185–199.

Lord, D., Washington, S.P., Ivan, J.N., 2005b. Poisson, Poisson-gamma and zero-inflated regression models of motor vehicle crashes: balancing statistical fit and theory. Accid. Anal. Prev. 37 (1), 35–46.

Lord, D., Washington, S.P., Ivan, J.N., 2007. Further notes on the application of zero inflated models in highway safety. Accid. Anal. Prev. 39 (1), 53–57.

Ma, J., Kockelman, K.M., 2006. Bayesian multivariate Poisson regression for models of injury count by severity. In: Proceedings of the Transportation Research Board 85th Annual Meeting, Washington, DC (CD-ROM).

Mackay, D.J.C., 1992. Bayesian methods for adaptive models. Ph.D. Dissertation. California Institute of Technology, Pasadena, California.

Maher, M.J., Summersgill, I., 1996. A comprehensive methodology for the fitting of predictive accident models. Accid. Anal. Prev. 28 (3), 281–296.

Marzban, C., Witt, A., 2001. A Bayesian neural network for severe-hail size prediction. Weather Forecast. 16 (5), 600–610.

2006. MATLAB Neural Network Toolbox 5. The MathWorks Inc., Natick, MA.

Miaou, S.P., 1994. The relationship between truck accidents and geometric design of road sections: Poisson versus negative binomial regressions. Accid. Anal. Prev. 26 (4), 471–482.

Miaou, S.-P., Lord, D., 2003. Modeling traffic crash-flow relationships for intersections: dispersion parameter, functional form, and Bayes versus empirical Bayes. Transport. Res. Rec. 1840, 31–40.

Miaou, S.P., Lum, H., 1993. Modeling vehicle accidents and highway geometric design relationships. Accid. Anal. Prev. 25 (6), 689–709.

Miaou, S.P., Song, J.J., 2005. Bayesian ranking of sites for engineering safety improvements: decision parameter, treatability concept, statistical criterion, and spatial dependence. Accid. Anal. Prev. 37 (4), 699–720.

Miaou, S.-P., Lu, A., Lum, H.S., 1996. Pitfalls of using R^2 to evaluate the goodness-of-fit of predictive models. Transport. Res. Rec. 1542, 6–13.

Milton, J., Mannering, F., 1998. The relationship among highway geometrics, traffic-related elements and motor-vehicle accident frequencies. Transportation 25 (4), 395–413.

Mitra, S., Washington, S., 2007. On the nature of over-dispersion in motor vehicle crash prediction models. Accid. Anal. Prev. 39, 459–468.

Mussone, L., Ferrari, A., Oneta, M., 1999. An analysis of urban collisions using an artificial intelligence model. Accid. Anal. Prev. 31 (6), 705–718.

Neal, R.M., 1995. Bayesian learning for neural networks. Ph.D. Dissertation. University of Toronto, Toronto, Ontario.

Oh, J., Lyon, C., Washington, S., Persaud, B., Bared, J., 2003. Validation of FHWA crash models for rural intersections: lessons learned. Transport. Res. Rec. 1840, 41–49.

Oh, J., Washington, S.P., Choi, K., 2004. Development of accident prediction models for rural highway intersections. Transport. Res. Rec. 1897, 18–27.

Oh, J., Washington, S.P., Nam, D., 2006. Accident prediction model for railway-highway interfaces. Accid. Anal. Prev. 38 (2), 346–356.

Okamoto, H., Koshi, M., 1989. A method to cope with the random errors of observed accident rates in regression analysis. Accid. Anal. Prev. 21 (4), 317–332.

Park, E.S., Lord, D., 2006. Multivariate Poisson-lognormal models for jointly modeling crash frequency by severity. In: Paper presented at the 86th Annual Meeting of the Transportation Research Board, Washington, DC.

Payne, R.W. (Ed.), 2000. The Guide to Genstat. Lawes Agricultural Trust, Rothamsted Experimental Station, Oxford, UK.

Persaud, B.N., Lord, D., Palminaso, J., 2002. Issues of calibration and transferability in developing accident prediction models for urban intersections. Transport. Res. Rec. 1784, 57–64.

Qin, X., Ivan, J.N., Ravishanker, N., Liu, J., 2005. Hierarchical Bayesian estimation of safety performance functions for two-lane highways using Markov Chain Monte Carlo modeling. J. Transport. Eng. 131 (5), 345–351.

Riviere, C., Lauret, P., Ramsamy, J.F.M., Page, Y., 2006. A Bayesian neural network approach to estimating the energy equivalent speed. Accid. Anal. Prev. 38 (2), 248–259.

SAS Institute Inc., 2002. Version 9 of the SAS System for Windows. SAS Institute Inc., Cary, NC.

Schluter, P.J., Deely, J.J., Nicholson, A.J., 1997. Ranking and selecting motor vehicle accident sites by using a hierarchical Bayesian model. Statistician 46 (3), 293–316.

Shankar, V., Mannering, F., Barfield, W., 1995. Effect of roadway geometrics and environmental factors on rural freeway accident frequencies. Accid. Anal. Prev. 27 (3), 371–389.

Shankar, V., Milton, J., Mannering, F., 1997. Modeling accident frequencies as zero-altered probability processes: an empirical inquiry. Accid. Anal. Prev. 29 (6), 829–837.

Song, J.J., Gosh, M., Miaou, S.-P., Malik, B., 2006. Bayesian multivariate spatial models for roadway traffic crash mapping. J. Multivar. Anal. 97 (1), 246–273.

Tunaru, R., 2002. Hierarchical Bayesian models for multiple count data. Aust. J. Stat. 31 (2/3), 221–229.

Vogt, A., Bared, J.G., 1998. Accident Models for Two-Lane Rural Roads: Segments and Intersections. Report No. FHWA-RD-98-133. Federal Highway Administration, Mclean, Virginia.

Warton, D.I., 2005. Many zeros does not mean zero inflation: comparing the goodness-of-fit of parametric models to multivariate abundance data. Environmetrics 16 (2), 275–289.

Zhang, Y., Ye, Z., Lord, D., 2006. Estimating the dispersion parameter of the negative binomial distribution for analyzing crash data using a Bootstrapped maximum likelihood method. In: Paper presented at the 86th Annual Meeting of the Transportation Research Board, Washington, DC.