

## **Predicting crash frequency using an optimized radial basis function neural network model**

Helai Huang, Ph.D., Professor,  
Urban Transport Research Center,  
School of Traffic and Transportation Engineering,  
Central South University, Changsha, Hunan, 410075 P.R. China.  
Tel: 86 731 82656631  
E-mail: huanghelai@csu.edu.cn

Qiang Zeng, Ph.D. Candidate,  
Urban Transport Research Center,  
School of Traffic and Transportation Engineering,  
Central South University, Changsha, Hunan, 410075 P.R. China.  
E-mail: qzeng1988@csu.edu.cn

Xin Pei,\* Ph.D., Lecturer,  
Department of Automation,  
Tsinghua University, Beijing, P.R. China.  
Tel: 861062795043  
E-mail: peixin@mail.tsinghua.edu.cn

S.C. Wong, Ph.D., Chair Professor,  
Department of Civil Engineering,  
The University of Hong Kong, Pokfulam Road, Hong Kong.  
Tel: 852 28591964  
E-mail: hhecwsc@hku.hk

Pengpeng Xu, Master Student,  
Urban Transport Research Center,  
School of Traffic and Transportation Engineering,  
Central South University, Changsha, Hunan, 410075 P.R. China.  
E-mail: xh150723@163.com

\* Corresponding author

This research was jointly supported by the Natural Science Foundation of China (No. 51108465, 71371192, 71301083), the Research Fund for Fok Ying Tong Education Foundation of Hong Kong (No. 142005), the Research Funds of Tsinghua University (No. 20151080412), and the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. 717512).

## ABSTRACT

With the enormous losses to society that result from highway crashes, gaining a better understanding of the risk factors that affect traffic crash occurrence has long been a prominent focus of safety research. In this study, we develop an optimized radial basis function neural network (RBFNN) model to approximate the nonlinear relationships between crash frequency and the relevant risk factors. Our case study compares the performance of the RBFNN model with that of the traditional negative binomial (NB) and back-propagation neural network (BPNN) models for crash frequency prediction on road segments in Hong Kong. The results indicate that the RBFNN has better fitting and prediction performance than the NB and BPNN models. After the RBFNN is optimized, its approximation performance improves, although several factors are found to hardly influence the frequency of crash occurrence for the crash data that we use. Furthermore, we conduct a sensitivity analysis to determine the effects of the remaining input variables of the optimized RBFNN on the outcome. The results reveal that there are nonlinear relationships between most of the risk factors and crash frequency, and they provide a deeper insight into the risk factors' effects than the NB model, supporting the use of the modified RBFNN models for road safety analysis.

**Keywords:** crash frequency prediction; radial basis function neural network; nonlinear relationship; sensitivity analysis.

## 1. Introduction

Highway crashes are estimated to result in about 1.24 million fatalities and as many as 50 million injuries annually around the world. They are predicted to be the fifth leading cause of death by 2030 ([World Health Organization 2013](#)). Given the enormous importance of highway safety in human societies, researchers have continually sought ways to better understand how the probability of crashes is affected by the relevant risk factors in the expectation of providing suggestions for laws, regulations and countermeasures aimed at reducing crash occurrence. In the absence of detailed driving data such as acceleration, braking and steering information, the associations between the risk factors and the number of crashes occurring at certain road entities, such as road segments or intersections, over a certain period (e.g. weeks, months or years; that is, crash frequencies) are investigated in most highway safety analyses. Moreover, predicting the crash frequencies at identified road segments, intersections or other facilities has become the technical core of roadway safety management in the *Highway Safety Manual* ([AASHTO 2010](#)).

Over the past decades, a large number of crash prediction models (CPMs) have been proposed. Statistical count models are the most popular CPMs, as they are able to explicitly illustrate the effects of observed explanatory variables on the frequency of crash occurrence and account for some characteristics in crash data, such as over-dispersion ([Miaou, 1994](#)), spatiotemporal correlations ([Shin and Washington 2012](#), [Wang et al. 2013](#), [Zeng and Huang 2014a](#)), and multilevel heterogeneities ([Anastasopoulos and Mannering 2009](#), [Huang and Abdel-Aty 2010](#), [Xu and Huang 2015](#)). However, in these statistical CPMs, the assumed generalized linear function and certain distributions of error terms may deteriorate their fitting/predicting performance and even cause over-fitting phenomena ([Xie et al. 2007](#)). [Lord and Mannering \(2010\)](#) and [Mannering and Bhat \(2014\)](#) have provided more detailed descriptions and assessments of these models.

Apart from statistical models, some artificial intelligence (AI) models have also been proposed for crash frequency prediction ([Chang 2005](#), [Li et al. 2008](#)). As a common class of AI models, neural network (NN) models have been successfully used in many fields of transportation research, including highway safety analysis ([Karlaftis and Vlahogianni 2011](#)). According to the results of previous studies, NNs exhibit better prediction performance than some traditional statistical models, such as the negative binomial (NB) model ([Chang 2005](#), [Xie et al. 2007](#)), suggesting it would be advantageous to use NNs to predict crash frequency.

NN models are information processing mechanisms inspired by biological nervous systems ([Haykin 2009](#)). Depending on their architecture, NN models can be divided into two categories, namely, feed forward and recurrent. In the former, processing units (also known as 'neurons' or 'nodes') are often grouped into layers (input, hidden, and output layers) and are connected from one layer to the next in one direction, from the input layer to the output layer ([Andina and Pham 2007](#)). Multilayer perceptron (MLP) and radial basis function neural network (RBFNN) are two classic types of feed forward NNs. Although both are universal approximators,

the numerical experiments demonstrate that RBFNN may outperform MLP with respect to function approximation and learning speed (Haykin 2009), suggesting that a RBFNN model describes more accurate relationships between crash frequency and risk factors. However, only BPNNs, a typical kind of MLPs trained by the back-propagation (BP) algorithm, were used to predict crash frequency in previous studies (Chang 2005, Xie et al. 2007).

In addition to the NN category, network structure, which mainly refers to the number of nodes in the hidden and input layers, is an important issue to be considered in NN model development, as it has a profound effect on the generalization performance (Haykin 2009). With the development of NN techniques, many advanced methods for optimizing model structure have been proposed (Ma and Khorasani 2003, Nielsen and Hansen 2008). The optimized NN models could not only effectively avoid over-fitting/under-fitting phenomena, but also identify the factors that scarcely contribute to outcome prediction. The NN pruning for function approximation (N2PFA) algorithm proposed by Setiono and Leow (2000) is used in this study to optimize the structure of the trained RBFNN, because it can quickly remove the redundant nodes and has few limitations on network architecture or training algorithms.

In this study, we aim to (1) develop an optimized RBFNN model to predict crash frequency, (2) identify the risk factors that appear from the evidence in the empirical study to be of little relevance to crash frequency and (3) compare the trained and optimized RBFNNs with the NB and BPNN models in terms of model fitting and predictive performance. Accordingly, the remainder of this paper is organized as follows. The next section briefly describes the general frameworks of statistical and AI modeling for crash frequency prediction. Section 3 specifies the NB, BPNN and RBFNN models and methods for RBFNN training and structure optimization. Section 4 introduces a case study for model demonstration and comparison. Finally, Section 5 presents our conclusions and recommendations for future research.

## 2. Crash Frequency Modeling

CPMs are the analytical methods used to learn the relationships between the crash frequency and the risk factors from the collected data sets. After the development of the methods, the expected crash frequency of a site, facility or roadway can be estimated for the given values of the factors. Methodologically, the relationships have been approximated from two different perspectives, i.e. statistical and AI modeling, as mentioned above.

To specify the general frameworks of these two modeling approaches, consider a set of crash data  $\{\mathbf{x}_n, Y_n | n=1, 2, \dots, N\}$ , where  $\mathbf{x}$  is a vector consisting of  $I$  risk factors, i.e.  $\mathbf{x} = [x_1, x_2, \dots, x_I]$ , and  $Y$  represents the observed crash count, while  $N$  is the number of observations in the collected data.

Statistically, the dependent variable,  $Y$ , is deemed to be a random variable and is often assumed to follow a Poisson or NB distribution for its non-integer characteristic. That is,

$$Y \sim \text{Poisson}(\lambda), \text{ or } \text{negative binomial}(\lambda).$$

A generalized linear relationship is assumed between the expected value of  $Y$ ,  $\lambda$ , and the independent variables  $\mathbf{x}$ , i.e.

$$\ln \lambda = \mathbf{x}\boldsymbol{\beta} + \varepsilon,$$

where  $\boldsymbol{\beta}$  is a vector consisting of estimable coefficients and  $\varepsilon$  is a disturbing term that accounts for certain features of the crash data such as over-dispersion, spatiotemporal correlations and cross-level heterogeneities.

Although statistical regression models are able to clearly present the effect that each observed factor has on the crash frequency, their performance greatly relies on the collected data. For example, the CPMs dealing with over-dispersion or their variations might result in biased inferences when encountering under-dispersed crash data<sup>1</sup>. In addition, the generalized linear function is susceptible, because the perfect functional form is unknown.

In the framework of AI techniques, a specific function form is not required. They (including NNs and support vector machines, etc.) essentially work as black boxes as shown in Fig. 1. Based on universal approximation theories, NN models are able to approximate the potential relationships between the crash frequency and the risk factors underlying the collected crash data with any desired accuracy providing enough hidden nodes, though they may sometimes bring about over-fitting.

[Insert Fig. 1 here]

### 3. Methodology

As in previous research, the most popular statistical and NN models in crash frequency analysis, NB and BPNN, are used as two benchmarks to compare the fitting and predictive performance with those of the RBFNN models. In this section, we first describe the characteristics of the NB, BPNN, and RBFNN models. We then introduce the training and structure optimization algorithms for the RBFNN model.

#### 3.1. Model specification

##### 3.1.1. NB model

Modified from the basic Poisson model, the NB model (also known as the Poisson-gamma model) was developed to deal with the over-dispersion commonly found in crash data. The crash count,  $Y_n$ , is assumed to follow an NB distribution

---

<sup>1</sup>Despite the rareness, crash data can sometimes be characterized by under-dispersion, such as those used in the research of Lord et al. (2010) and Oh et al. (2006).

(Washington 2012), that is,

$$P(Y_n = y_n) = \frac{\Gamma(1/\alpha + y_n)}{\Gamma(1/\alpha) + y_n!} \left( \frac{1/\alpha}{1/\alpha + \lambda_n} \right)^{1/\alpha} \left( \frac{\lambda_n}{1/\alpha + \lambda_n} \right)^{y_n}, y_n = 0, 1, 2, 3, \dots, \quad (1)$$

where  $\Gamma(\bullet)$  is the Gamma function and  $\alpha$  is the over-dispersion parameter.

Regarding the mean of  $Y_n$ ,  $\lambda_n$ , its natural logarithm is assumed to be linearly associated with the corresponding risk factors,  $\mathbf{x}_n$ , such that

$$\ln \lambda_n = \mathbf{x}_n \boldsymbol{\beta}', \quad (2)$$

in which  $\boldsymbol{\beta}'$  is a vector of coefficients to be estimated.

### 3.1.2. BPNN model

The structure of a common BPNN model for crash frequency prediction is shown in Fig. 2, where the nodes are fully connected. In the input layer, each risk factor is represented by a node. In addition, a node equaling one is constantly added and the weights of its connection with those hidden nodes are the biases. The only output node represents the expected crash frequency.  $w_{j,i}^{(1)}$  is the connection weight between the hidden node,  $j (j=1, \dots, J)$ , and the input node,  $i (i=0, 1, \dots, I)$ , while  $w_j^{(2)}$  is the weight connection between the output node and the hidden node  $j (j=1, \dots, J)$ .

A hyperbolic function,  $\tanh(\bullet)$ , is used as the transfer function for all hidden and output nodes. Then, the predicted crash frequency,  $\psi$ , is given by:

$$\psi = \sum_{j=1}^J \tanh \left( w_j^{(2)} \tanh \left( \sum_{i=0}^I w_{j,i}^{(1)} x_i \right) \right). \quad (3)$$

[Insert Fig. 2 here]

### 3.1.3. RBFNN model

The RBFNN model is based on interpolation theory, which argues that patterns (referred to as “observations” in statistical models) nonlinearly projected into a higher dimensional space are more linearly separable (Cover 1965). Specifically, an RBFNN model often consists of an input layer, a hidden layer, and an output layer, which correspond to the input space, hidden/feature space and output space, respectively. Patterns are first mapped from the input space into the hidden space by nonlinear functions, i.e. radial basis functions (RBFs) and then mapped from the hidden space to

the output space by a linear function. Generally, the more nodes there are in the hidden layer, meaning a higher dimensional hidden space, the better the approximation performance achieved. Please refer to [Haykin \(2009\)](#) and [Lowe \(1988\)](#) for more description of the inherent mechanism of RBFNN. [Fig. 3](#) shows the structure of an RBFNN model developed to predict crash frequency.

[Insert [Fig. 3](#) here]

In the hidden layer, the popular Gaussian function is used for all RBFs:

$$\varphi_k(\mathbf{x}) = \exp\left(-\frac{1}{2\sigma_1^2}\|\mathbf{x}-\boldsymbol{\mu}_k\|^2\right), \quad k=1,2,\dots,K, \quad (4)$$

where  $\boldsymbol{\mu}_k$  is the center of RBF  $\varphi_k(\bullet)$  ( $k=1,2,\dots,K$ ) and  $\|\mathbf{x}-\boldsymbol{\mu}_k\|$  denotes the Euclidian norm between the input,  $\mathbf{x}$ , and the center,  $\boldsymbol{\mu}_k$ . The spread of the function which controls the smoothness of the function approximation, is  $\sigma_1$ . The basis is denoted by  $w_0$  and  $w_k$  ( $k=1,2,\dots,K$ ) is the connection weight between the output node and  $\varphi_k(\bullet)$  ( $k=1,2,\dots,K$ ). Therefore, the predicted crash frequency,  $\psi$ , is given by:

$$\psi = w_0 + \sum_{k=1}^K w_k \exp\left(-\frac{1}{2\sigma^2}\|\mathbf{x}-\boldsymbol{\mu}_k\|^2\right). \quad (5)$$

### 3.2. RBFNN training

For a given number of hidden nodes, a hybrid learning/training method is used to approximate the real relationships between input attributes and crash frequency ([Haykin 2009](#)). First, the centers of RBFs are determined by a K-means cluster algorithm. Then, the basis and weights between the output node and RBFs are estimated using a recurrent least squares (RLS) algorithm.

#### 3.2.1. K-means cluster algorithm

Methodologically, clustering is an unsupervised approach, which tries to group a set of observations into a given number,  $K$ , of clusters and minimize the difference in observations within the same cluster:

$$\min J(C) = \frac{1}{2} \sum_{k=1}^K \sum_{C(n)=k} \sum_{C(n')=k} \|\mathbf{x}_n - \mathbf{x}_{n'}\|^2, \quad (6)$$

where  $k = C(n)$  ( $n=1,2,\dots,N; k=1,2,\dots,K$ ) is a many-to-one mapper/encoder. It

assigns observation,  $\mathbf{x}_n$ , into cluster  $k$ .

To search for the optimal encoder, the  $K$ -means cluster algorithm iteratively processes the following two steps, until the cluster assignment is not changed.

1. For a given encoder,  $C$ , the total variance of all clusters is minimized by the means  $\{\hat{\boldsymbol{\mu}}_k | k=1, 2, \dots, K\}$ :

$$\min_{\{\hat{\boldsymbol{\mu}}_k\}_{k=1}^K} \sum_{k=1}^K \sum_{C(n)=k} \|\mathbf{x}_n - \hat{\boldsymbol{\mu}}_k\|^2. \quad (7)$$

2. For the optimized cluster means  $\{\hat{\boldsymbol{\mu}}_k | k=1, 2, \dots, K\}$ , the encoder is optimized:

$$C(n) = \arg \min_{1 \leq k \leq K} \|\mathbf{x}_n - \hat{\boldsymbol{\mu}}_k\|^2. \quad (8)$$

After the algorithm is converged, the means  $\{\hat{\boldsymbol{\mu}}_k | k=1, 2, \dots, K\}$  are set as the centers of the RBFs in the NN model.

### 3.2.2. RLS algorithm

After the determination of RBFs' centers, patterns  $\{\mathbf{x}_n, Y_n | n=1, 2, \dots, N\}$  can be mapped into  $\{\boldsymbol{\Phi}_n, Y_n | n=1, 2, \dots, N\}$ , where  $\boldsymbol{\Phi}_n = [1, \varphi_1(\mathbf{x}_n), \varphi_2(\mathbf{x}_n), \dots, \varphi_K(\mathbf{x}_n)]'$ . The RLS algorithm updates the connection weights,  $\mathbf{w} = [w_0, w_1, w_2, \dots, w_K]'$ , by the following equations (please see [Haykin \(2009\)](#) for more details),

$$\mathbf{P}_n = \mathbf{P}_{n-1} - \frac{\mathbf{P}_{n-1} \boldsymbol{\Phi}_n \boldsymbol{\Phi}_n' \mathbf{P}_{n-1}}{1 + \boldsymbol{\Phi}_n' \mathbf{P}_{n-1} \boldsymbol{\Phi}_n}, \quad (9)$$

$$\mathbf{g}_n = \mathbf{P}_n \boldsymbol{\Phi}_n, \quad (10)$$

$$\alpha_n = Y_n - \mathbf{w}_{n-1}' \boldsymbol{\Phi}_n, \quad (11)$$

$$\mathbf{w}_n = \mathbf{w}_{n-1} + \mathbf{g}_n \alpha_n. \quad (12)$$

To initialize the algorithm, set  $\mathbf{w}_0 = \mathbf{0}$  and  $\mathbf{P}_0 = \lambda^{-1} \mathbf{I}$ , where  $\lambda$  is a very small positive number.

In the absence of useful prior knowledge we usually select the number of hidden nodes empirically or randomly. However, as a result the network may not be sufficiently trained. To achieve the aimed training performance we can add a certain number of hidden nodes and iteratively retrain the NN by the hybrid algorithm, because the training performance is positively correlated with the number of hidden nodes as mentioned above. In summary, the whole process of RBFNN training is



shown in Fig. 4.

[Insert Fig. 4 here]

### 3.3. Structure optimization

To improve the generalization capacity of the RBFNN model and to identify the almost irrelevant explanatory variables, the N2PFA algorithm, which has been successfully used to develop an optimized NN for crash injury severity prediction (Zeng and Huang 2014b), is proposed to prune the nodes that do not cause significant deterioration of the network's accuracy (Setiono and Leow 2000). The mean absolute deviations (MADs) of the training set,  $\mathbf{T}$ , and testing set,  $\mathbf{X}$ , that is  $p$  and  $q$ , are used to evaluate the network's fitting and predictive performance,

$$p = \frac{1}{M_1} \sum_{(x_m, Y_m) \in \mathbf{T}} |Y_m - \psi(m)|, \quad (13)$$

$$q = \frac{1}{M_2} \sum_{(x_m, Y_m) \in \mathbf{X}} |Y_m - \psi(m)|, \quad (14)$$

where  $M_1$  and  $M_2$  are the number of patterns in the training and testing sets, respectively. The N2PFA algorithm has been modified to fit its combination with the proposed RBFNN training method. The following steps describe the detailed pruning process.

1. Train the RBF network using the hybrid algorithm described above until the predetermined training performance is obtained.
2. Calculate  $p$  and  $q$  of the trained RBFNN and set  $p\_b = p$ ,  $q\_b = q$ ,  $ermax = \max\{p\_b, q\_b\}$ .
3. For each  $i$  ( $i = 1, \dots, I$ ), set the  $i$ th element  $x_i = 0$  for each attribute vector  $\mathbf{x}$  in the training set  $\mathbf{T}$  and calculate the fitting errors  $p_i$ .
4. Retrain the network with data set  $\mathbf{T}_l$ , which is  $\mathbf{T}$  after eliminating the  $l$ th elements of all input vectors where  $p_l = \min_i p_i$ , and compute  $p$  and  $q$  of the retrained network.
5. If  $p \leq (1 + \sigma_2)ermax$  and  $q \leq (1 + \sigma_2)ermax$ , then remove the input node  $l$ , set  $p\_b = \min\{p, p\_b\}$ ,  $q\_b = \min\{q, q\_b\}$ ,  $ermax = \max\{p\_b, q\_b\}$ ,  $\mathbf{T} = \mathbf{T}_l$ ,  $I = I - 1$  and go back to step 3. Otherwise, keep the previous network.

6. Retrain the network with  $K = K - 1$  and compute  $p$  and  $q$  of the retrained network.
7. If  $p \leq (1 + \sigma_2)ermax$  and  $q \leq (1 + \sigma_2)ermax$ , then set  $p\_b = \min\{p, p\_b\}$ ,  $q\_b = \min\{q, q\_b\}$ ,  $ermax = \max\{p\_b, q\_b\}$ ,  $K = K - 1$ , and go back to step 6. Otherwise, keep the previous network.

In the process above,  $p\_b$  and  $q\_b$  separately represent the minimum MADs of training and testing sets achieved so far. We use  $ermax$  as the parameter to determine whether a node can be removed or not and assign it the larger of  $p\_b$  and  $q\_b$  so as to remove as many redundant nodes as possible without sacrificing generalization accuracy. In addition,  $\sigma_2$  is a factor to control the chance that a node will be removed.

## 4. Case study

### 4.1. Data

A crash dataset obtained from the Traffic Information System (TIS) maintained by the Transport Department of Hong Kong is used to demonstrate the proposed RBFNN models and to compare them with the NB and BPNN models. This dataset contains 211 road segments evenly and widely distributed across Hong Kong. Geographical information system (GIS) techniques are used to map crashes to these segments, and the crash counts of the sites are aggregated by year for the 2002-2006 period. The road geometric and traffic information is also included in the dataset. [Table 1](#) illustrates the definitions and descriptive statistics of the variables used in the development of the model.

[Insert [Table 1](#) here]

The lane changing opportunity (LCO) variable refers to the length-weighted average number of possible lane-cuttings in a sub-segment with identical lane markings. For double continuous lines, no lane changing is allowed, thus,  $LCO = 0$ . For double lines with one continuous line and one broken line, lane changing is only allowed from the side of the broken line to the side of the continuous line, thus,  $LCO = 1$ . For a single broken line, lane changing between both adjacent lanes is allowed, thus,  $LCO = 2$ . [Pei et al. \(2012\)](#) provided a more detailed description of LCO.

According to [Table 1](#), the mean and variance (square of S.D.) of the crash frequencies are 7.64 and 40.32, respectively, which indicates over-dispersion in the

crash data. Therefore, a NB model is more suitable than a Poisson model. In the NB model, to account for the potential nonlinear relationship between the crash frequency and traffic volume, the natural logarithm of AADT and Length,  $\ln(\text{AADT})$  and  $\ln(\text{Length})$ , are modeled as other factors (Zeng and Huang 2014a).

Correlation tests and collinearity diagnoses for the risk factors are conducted.  $\ln(\text{AADT})$  and Lane,  $\ln(\text{AADT})$  and Park, Lane and LCO, SL (speed limit) and Shoulder, SL and Park are found to be significantly correlated, as the absolute values of their correlation coefficients are all over 0.6 (0.707, 0.651, 0.653, 0.692, 0.672, respectively). To reduce the model complexity and retain some important variables (such as  $\ln(\text{AADT})$ , etc.), Lane, Park, and Shoulder are excluded from the models. The results of the diagnoses indicate that there is no significant collinearity in the remaining factors.

#### **4.2. Results and discussions**

The NB model is estimated with Stata, a software program used to implement the statistical models. The training and structure optimization algorithms of the BPNN and RBFNN models are programmed in MATLAB. All of the variables are normalized for the convenience of network training. To implement the structure optimization algorithm and to compare the performance of these models, the prepared dataset is randomly divided into two parts, a training subset,  $\mathbf{T}$ , and a testing subset,  $\mathbf{X}$ , which account for approximately 80% and 20% of the dataset, respectively (Haykin 2009). During the RBFNN training, the network starts with a small number of hidden layer neurons and hidden units are then incrementally added until the mean square error (MSE) is not over 0.005. Accordingly, the BPNN contains the same number of hidden nodes as the trained RBFNN and the convergence criterion of its training is that the MSE is less than 0.005. In the N2PFA algorithm, we assume that  $\sigma = 0.025$ .

The results of the model comparison are shown in Table 2, while the mean and 95% confidence interval of the over-dispersion parameter in the NB model are 0.219 and [0.188,0.255] respectively, suggesting the crash data used is over-dispersed. However, in terms of MAD criterion, obviously, all NN models, including the BPNN (3.01/3.38), the trained (2.79/3.28), and optimized (3.05/3.19) RBFNNs, have smaller fitting/prediction errors for training and testing datasets in contrast to the NB model (3.51 and 3.64), which again demonstrates that NN models can give better fitting and predictive performance than traditional statistical models in crash frequency prediction. Moreover, the trained RBFNN's training/testing MADs and training time are less than those of the BPNN, demonstrating RBFNN's better performance for approximating the relationships between crash frequency and the risk factors. This is probably attributable to the RBFNN's greater capacity for approximating arbitrary non-linear functions without any prior knowledge, which has been found in many previous studies (Abdelwahab and Abdel-Aty 2002, Haykin 2009).

[Insert Table 2 here]

For the trained RBFNN, the testing MAD is 0.49 (about 17.6%) above the training MAD, suggesting that the model is over-fitted. After the network structure is optimized, the training MAD increases while the testing MAD decreases, and their difference is only 0.14 (about 4.6%). That is, the over-fitting has been significantly eliminated. In addition, seven input nodes, corresponding to Merge, Diverge, Inter, Median, Gradient, Curvature, and Width, and fifteen hidden nodes are removed from the trained RBFNN, which indicates that the original model has many redundant nodes and that the deleted factors have no significant effects on the crash frequency. The remaining factors are found to be significant in the NB model, with the estimation results shown in [Table 3](#).

[Insert [Table 3](#) here]

### **4.3. Sensitivity analysis**

Although the present study is focused on model development and comparison, a good understanding of the explanatory variables is also important to partially justify the model's validity. As in previous research ([Delen et al. 2006](#), [Xie et al. 2007](#)), a sensitivity analysis (SA) based on the optimized RBFNN is conducted to explore each remaining explanatory variable's effect on the crash frequency. When analyzing explanatory variables, a continuous variable varies between its mean plus and minus its certain number (0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, or 5) of standard deviations (S.D.s) within a reasonable interval. A categorical variable varies among all of its categories except the reference one, while the other variables are fixed on their means (continuous variables) or reference cases (categorical variables) ([Fishand Blodgett 2003](#), [Xie et al. 2007](#)). [Fig. 5](#) shows the results of SA for the remaining continuous variables.

[Insert [Fig. 5](#) here]

According to the results in [Fig. 5](#), the effects of three variables, AADT, SL, and LCO, are fluctuant, indicating that they are obviously nonlinearly related to crash frequency. Specifically, AADT increases or decreases crash frequency when it is smaller or bigger than its mean plus two S.D.s (61967 vehicles), which may be a result of the significant decrease in travel speed when the daily traffic volume reaches a certain level (such as 61967 vehicles). It should be pointed out that the nonlinear relationship between traffic volume and crash frequency has been confirmed by many traffic safety statistical analyses ([Labi 2011](#), [Qin et al. 2004](#)). Crash frequency decreases with higher speed limits when they are smaller than the mean plus 2.5 S.D.s (97 km/h), covering 95% of roads in Hong Kong. This result may be attributed to the well planned, constructed, and managed features that promote safety of roadways designed for higher speeds ([Milton and Mannering 1998](#)). The crash count will significantly increase if the speed limit is too high, conforming to traffic engineering intuition and many existing findings ([Aguero-Valverde and Jovanis 2008](#)). LCO is found to decrease the crash frequency between the mean (2.43) and the mean plus

four S.D.s (8.87) and to increase it at other values. This result may be generated by a combination of the permissible lane-cuttings' two functions. On the one hand, lane-cuttings increase vehicle interaction, such as overtaking, thereby raising the incidence of traffic conflict (Pei et al. 2012). On the other hand, they offer an opportunity for errant vehicles to recover or for vehicles to seek temporary refuge to avoid an errant oncoming vehicle, therefore reducing the risk of collisions.

Length and Rainfall have consistently positive effects on crash frequency. That is, more crashes tend to occur on longer road segments with more annual precipitation. These results are consistent with those in the NB model and many previous studies (AASHTO 2010, Zeng and Huang 2014a). From Fig. 5(b), we can see that there is an approximate linear relationship between crash frequency and segment length. It is consistent with the result of SA on the Bayesian NN developed by Xie et al. (2007). Despite the unidirectional effect, the different change rates of crash frequency under different variations of Rainfall indicate that it is still nonlinearly related to crash frequency.

With regard to the only remaining categorical variable, BS (bus stop) is found to have a positive (0.154) effect on the crash frequency in the optimized RBFNN, consistent with that in the NB model. This may be attributed to frequent pedestrian activity around bus stops and increased interaction between buses and other vehicles when entering or leaving bus bays (Pei et al. 2012). In fact, 93% of the pedestrian-involved crashes in the observed road segments occurred on roadways with bus stops.

## 5. Conclusions

In this study, we develop an optimized RBFNN model for crash frequency prediction. To learn the underlying relationships between crash frequency and risk factors, we use a hybrid method combining a K-means cluster algorithm and a RLS algorithm, and we propose a structure optimization algorithm for improving the generalization performance and recognizing the explanatory variables that may be redundant for the collected crash data. A crash data set obtained from the TIS maintained by the Transport Department of Hong Kong is used to demonstrate the proposed methods and to compare them with the NB and BPNN models.

Despite the crash data being over-dispersed, the results show that both the trained and the optimized RBFNN models outperform the NB and BPNN models in fitting/predictive performance, suggesting the advantages of the RBFNN over the NB model and the commonly used BPNN in the case study. Seven input nodes and fifteen hidden nodes are deleted in the optimized RBFNN, and its over-fitting is significantly relieved, demonstrating the ability of the structure optimization algorithm to empirically identify nearly irrelevant factors and to improve the generalization capacity.

Moreover, an SA is conducted on the optimized RBFNN model to determine the relationships between the explanatory variables and crash frequency. All continuous variables except for SL are found to be nonlinearly associated with the outcome. Although some of the results are consistent with those in the NB model, they illustrate

the effects of the risk factors more explicitly, further suggesting that the optimized RBFNN model is a good alternative for crash frequency prediction.

Compared to NB model and BPNN used in previous studies, the proposed RBFNN techniques achieve better fitting and predicting performance when modeling crash frequency without any prior knowledge. However, it is worth noting that the evidence for the superiority of the RBFNN is restricted to this case study. More field data sets are needed to confirm its universality. Moreover, comparisons between the RBFNN and the recently proposed statistical models such as the random parameter model may further justify its advantage. Nonetheless, as a universal approximation, the proposed RBFNN model may be useful in other aspects of highway safety analysis, such as predicting crash injury severity and crash frequencies by injury severity.

### **Acknowledgments**

This research was jointly supported by the Natural Science Foundation of China (No. 51108465, 71371192, 71301083), the Research Fund for Fok Ying Tong Education Foundation of Hong Kong (No. 142005), the Research Funds of Tsinghua University (No. 20151080412), and the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. 717512).

### **References**

- AASHTO, 2010. *Highway Safety Manual*, 1st edition.
- Abdelwahab, H.T., and Abdel-Aty, M.A., 2002. Artificial neural networks and logit models for traffic safety analysis of toll plazas. *Transportation Research Record*, 1784, 115-125.
- Aguero-Valverde, J., and Jovanis, P.P., 2008. Analysis of road crash frequency with spatial models. *Transportation Research Record*, 2061, 55-63.
- Anastasopoulos, P.C., and Mannering, F.L., 2009. A note on modeling vehicle accident frequencies with random-parameters count models. *Accident Analysis and Prevention*, 41(1), 153-159.
- Andina, D., and Pham, D.T., 2007. *Computational Intelligence: for Engineering and Manufacturing*. Springer, Boston.
- Chang, L., 2005. Analysis of freeway accident frequencies: negative binomial regression versus artificial neural network. *Safety Science*, 43(8), 541-557.
- Cover, T.M., 1965. Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition. *IEEE Transaction on Electronic Computers*, 14, 326-334.
- Delen, D., Sharda, R., and Bessonov, M., 2006. Identifying significant predictors of injury severity in traffic accidents using a series of artificial neural networks. *Accident Analysis and Prevention*, 38(3), 434-444.
- Fish, K.E., and Blodgett, J.G., 2003. A visual method for determining variable importance in an artificial neural network model: an empirical benchmark study. *Journal of Targeting, Measurement and Analysis for Marketing*, 11(3), 244-254.

- Haykin, S.S., 2009. *Neural networks and learning machines*, 3rd edition. Prentice Hall, New York.
- Huang, H., and Abdel-Aty, M., 2010. Multilevel data and Bayesian analysis in traffic safety. *Accident Analysis and Prevention*, 42(6), 1556-1565.
- Karlaftis, M.G., and Vlahogianni, E.I., 2011. Statistical methods versus neural networks in transportation research: differences, similarities, and some insights. *Transportation Research Part C: Emerging Technologies*, 19(3), 387-399.
- Labi, S., 2011. Efficacies of roadway safety improvements across functional subclasses of rural two-lane highways. *Journal of Safety Research*, 42(4), 231-239.
- Li, X., Lord, D., Zhang, Y., and Xie, Y., 2008. Predicting motor vehicle crashes using support vector machine models. *Accident Analysis and Prevention*, 40(4), 1611-1618.
- Lord, D., Geedipally, S.R., and Guikema, S.D., 2010. Extension of the Application of Conway-Maxwell-Poisson Models: Analyzing Traffic Crash Data Exhibiting Under-dispersion. *Risk Analysis*, 30(8), 1268-1276.
- Lord, D., and Mannering, F., 2010. The statistical analysis of crash-frequency data: a review and assessment of methodological alternatives. *Transportation Research Part A: Policy and Practice*, 44(5), 291-305.
- Lowe, D., 1988. Multi-variable functional interpolation and adaptive networks. *Complex Systems*, 2, 321-355.
- Ma, L., and Khorasani, K., 2003. A new strategy for adaptively constructing multilayer feedforward neural networks. *Neurocomputing*, 51, 361-385.
- Mannering, F.L., and Bhat, C.R., 2014. Analytic methods in accident research: methodological frontier and future directions. *Analytic Methods in Accident Research*, 1, 1-22.
- Miaou, S.P., 1994. The relationship between truck accidents and geometric design of road sections: Poisson versus negative binomial regressions. *Accident Analysis and Prevention*, 26(4), 471-482.
- Milton, J., and Mannering, F., 1998. The relationship among highway geometrics, traffic-related elements and motor-vehicle accident frequencies. *Transportation*, 25(4), 395-413.
- Nielsen, A.B., and Hansen, L.K., 2008. Structure learning by pruning in independent component analysis. *Neurocomputing* 71(10), 2281-2290.
- Oh, J., Washington, S.P., and Nam, D., 2006. Accident prediction model for railway-highway interfaces. *Accident Analysis and Prevention* 38 (2), 346-356.
- Pei, X., Wong, S.C., and Sze, N.N., 2012. The roles of exposure and speed in road safety analysis. *Accident Analysis and Prevention*, 48, 464-471.
- Qin, X., Ivan, J.N., and Ravishanker, N., 2004. Selecting exposure measures in crash rate prediction for two-lane highway segments. *Accident Analysis and Prevention*, 36(2), 183-191.
- Setiono, R., and Leow, W.K., 2000. Pruned neural networks for regression. In *PRICAI 2000 Topics in Artificial Intelligence*. Springer Berlin Heidelberg, 500-509.
- Shin, K., and Washington, S.P., 2012. Empirical Bayes method in the study of traffic

- safety via heterogeneous negative multinomial model. *Transportmetrica*, 8(2), 131-147.
- Wang, C., Quddus, M., and Ison, S., 2013. A spatio-temporal analysis of the impact of congestion on traffic safety on major roads in the UK. *Transportmetrica A: Transport Science*, 9(2), 124-148.
- Washington, S.P., Karlaftis, M.G., and Mannering, F.L., 2012. *Statistical and econometric methods for transportation data analysis*. CRC press.
- World Health Organization, 2013. *Global Status Report on Road Safety 2013: Supporting a Decade of Action*. World Health Organization, Geneva, Switzerland.
- Xie, Y., Lord, D., and Zhang, Y., 2007. Predicting motor vehicle collisions using Bayesian neural networks: an empirical analysis. *Accident Analysis and Prevention*, 39(5), 922-933.
- Xu, P., and Huang, H., 2015. Modeling crash spatial heterogeneity: random parameter versus geographically weighting. *Accident Analysis and Prevention*, 75, 16-25.
- Zeng, Q., and Huang, H., 2014a. Bayesian spatial joint modeling of traffic crashes on an urban road network. *Accident Analysis and Prevention*, 67, 105-112.
- Zeng, Q., and Huang, H., 2014b. A stable and optimized neural network model for crash injury severity prediction. *Accident Analysis and Prevention*, 73, 351-358.