# AN APPROACH TO PREDICT ROAD ACCIDENT FREQUENCIES: APPLICATION OF FUZZY NEURAL NETWORK


Lai Zheng
Ph.D. Candidate of Traffic Engineering, School of Transportation Science and Engineering, Harbin Institute of Technology, Harbin, China, e-mail: zhenglai1985@163.com

Xianghai Meng
Professor of Traffic Engineering, School of Transportation Science and Engineering, Harbin Institute of Technology, Harbin, China, e-mail: mengxianghai100@126.com



*Submitted to the 3rd International Conference on Road Safety and Simulation, September 14-16, 2011, Indianapolis, USA*



## ABSTRACT

Road accident prediction plays an important role in accessing and improving the road safety. Besides the conventional generalized linear regression, the prediction approaches based on fuzzy logic and neural networks have increasingly been proven to have a significant accident-predicting capability in recent years. However, fuzzy logic and neural network have their respective limitations. For example, it is difficult to construct a complete rule set for fuzzy logic and there is no general rule in determining the network structure for neural network. To overcome these limitations, the fuzzy neural network (FNN) is put forward. This approach has been applied for prediction in many areas, but no application exists in road accident prediction according to the authors' knowledge. Thus, this paper establishes a fuzzy neural network model (FNNM) for predicting accident frequencies. It is established based on a data set of 133 segments from urban arterials in Harbin city of China, which takes annual average daily traffic (AADT), lane width (LW), speed limit (SL) and traffic load (TL, calculated by volume/capacity) as input variables and accidents per kilometer per year (AF) as output variable. Comparisons among FNNM, fuzzy logic model (FLM) and BP neural network model (NNM) show the superiority of the FNNM in accuracy and flexibility. Finally, a sensitivity analysis is employed to identify the significant factors. The results show that AADT is the most significant factor in this model, followed by SL, TL and LW in order of their relative importance going from the most to the least significant.



**Keywords:** road accident prediction, fuzzy neural network, sensitivity analysis, urban arterial.


## INTRODUCTION

Road safety is always one of the major concerns of the whole society, since the death, injury and property loss caused by road accidents are considerable every year. One of the best ways to understand the occurrence of road accidents is to develop accident prediction models, which are also standard practice in assessing and improving the safety of roads for safety researchers and practitioners.

An accident prediction model is a mathematical model which describes the relationships between road accident frequencies and various traffic conditions, road geometric features, environment factors as well as the driver's behaviors. Considerable research, on accident prediction models, has been carried out in recent years, and these models may basically be grouped into four main approaches, namely multivariate analysis, empirical Bayes method, fuzzy logic and neural network (Caliendo et al., 2007).

Multivariate analysis usually models the road accident frequencies with multiple linear regression and generalized linear regression. The former is almost the earliest developed methods (Dionne et al., 1993; Okamoto and Koshi, 1989; Persaud and Dzbik, 1993), and now has been proven to be inadequate since its assumption of normally distributed errors and homoscedacity is not in accordance with the nature of accident occurrences. Thus, the Poisson regression model, negative binomial regression model and negative multinomial model based on generalized linear regression technique have been put forward (EI-Basyouny and Sayed (2006); Greibe, 2003; Lord and Persaud, 2000; Lord et al., 2005; Maher and Summersgill, 1996; Miaou et al., 1992; Miaou, 1994), and these models have been successfully applied and widespread adopted recently. Empirical Bayes method also has a significant accident-predicting capability, and this has been validated by Cafiso et al. (2010), EI-Basyouny and Sayed (2009) and Ozbay and Noyan (2006).

The fact that road accidents might not be a linear function of various dependent variables for prediction models has made large room for the using of non-linear approximators such as fuzzy logic and neural network. For example, Xiao et al. (1999) developed two fuzzy logic models for predicting the risk of accidents that occurred on wet pavements, and the two models were based on Mamdani inference method and Sugeno inference method, respectively. The result showed that the fuzzy logic models had superiority over both probabilistic model and nonlinear regression model. Meng et al. (2009) employed fuzzy logic to related urban road accident frequencies with various traffic and road conditions, and AADT and TL were recognized as the prominent influence factors by the model. Chang (2005) employed artificial neural network to analyze the freeway accident frequencies, and pointed out that the artificial neural network method did not require any pre-defined underlying relationship between dependent and independent variables. The study also demonstrated that the artificial neural network is a consistent alternative method for analyzing freeway accident frequency. Delen et al. (2006) used a series of artificial neural networks to model the potentially non-linear relationships between the injury severity levels and crash-related factors, and the artificial neural network models were found to have better predictive power comparing to traditional methods.

Albeit appearing to have a significant accident-predicting capability, researchers admitted that the fuzzy logic and neural network had some limitations, such as difficult to construct a complete fuzzy rule set for fuzzy logic, and time consuming and no general rule in determining the network structure for neural networks. To overcome the limitations, the FNN is put forward as a combination of the fuzzy logic and neural network, and this method has been applied for prediction in information, environment, energy and many other areas (Alotaibi et al., 2008; Alyisi and Franchini, 2011; Azamathulla et al., 2009; Wei et al., 2007). However, there is no application of FNN in road accident prediction till now according to the authors' knowledge (based on the open literature). Thus, this study will introduce this new approach to predict road accident frequencies and evaluate its application results.

The article is structured as follows: the basic structure of FNN is described in Section 2. Section 3 develops the accident prediction model based on FNN. Section 4 presents a performance evaluation of the proposed model by comparing it to other techniques and provides a sensitivity analysis of the input parameters. Section 5 presents some concluding remarks of this study.

## STRUCURE OF FNN

The FNN in this paper is developed based on adaptive neuro-fuzzy inference system (ANFIS), which integrates the best features of fuzzy inference systems and neural networks (Jang, 1993). A 5-layer network is employed to structure the FNN, and the first four layers are used to generate the premises of fuzzy rules, while the last layer is to generate the consequence. The neural network's learning algorithms are used to adjust the membership functions and associated parameters of consequence.

To simplify the operation a sample having two inputs and an output is considered, and the architecture of the FNN is shown in Figure 1 (Ekici and Aksoy, 2011).

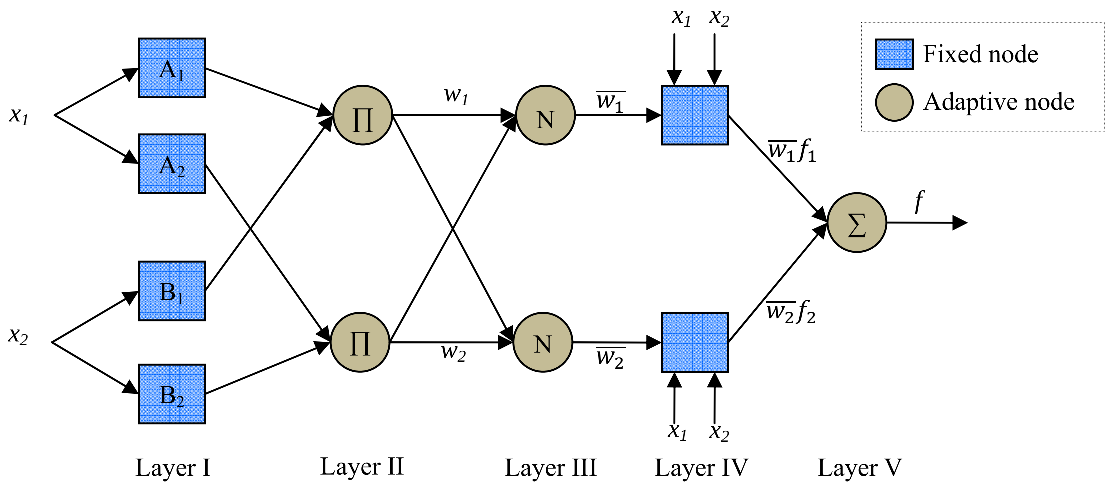



Figure 1 The architecture of FNN

*Layer I*: this layer is a fuzzy layer, in which $A_i$ and $B_i$ are fuzzy sets associated with inputs $x_1$ and $x_2$. The output of this layer is given by

$$O_i^1 = \begin{cases} \mu_{A_i}(x_1) \\ \mu_{B_i}(x_2) \end{cases}, i = 1, 2 \tag{1}$$

$$\mu_{A_i}(x_1) = \exp[-(x_1 - a_i^1)^2 / b_i^1] \tag{2}$$

$$\mu_{B_i}(x_2) = \exp[-(x_2 - a_i^2)^2 / b_i^2] \tag{3}$$

where:

| | |
|---|---|
| $O_i^1$: | the output of layer I |
| $\mu_{A_i}, \mu_{B_i}$: | membership functions in gauss type with maximum equals to 1 and minimum equals to 0 |
| $a_i^j, b_i^j$: | changeable parameters of the membership functions as well as parameters of the premise |

*Layer II*: this is a product layer with fixed nodes. ∏ indicates that the nodes play the role of a simple multiplier. The output of this layer is given by

$$O_i^2 = w_i = \mu_{A_i}(x_1)\mu_{B_i}(x_2),\ i=1,2 \quad (4)$$

where:

| | |
|---|---|
| $O_i^2$: | the output of layer II |
| $w_i$: | the weight of the *i*th rule |

*Layer III*: this is a normalized layer, whose nodes are fixed circles labeled as N. The *i*th node calculates the normalized value of the *i*th rule, given by

$$O_i^3 = \overline{w}_i = \frac{w_i}{\sum_i w_i},\ i=1,2 \quad (5)$$

where:

| | |
|---|---|
| $O_i^3$: | the output of layer III |
| $\overline{w}_i$: | the normalized weight of the *i*th rule |

*Layer IV*: it is a defuzzification layer with adaptive circle nodes, and this layer plays a role of simply product of the normalized value $\overline{w}_i$ and a first order polynomial. The two fuzzy if-then rules of this FNN are as follows:

Rule 1: if $x_1$ is $A_1$ and $x_2$ is $B_1$, then $f_1 = C_0^1 + C_1^1 x_1 + C_2^1 x_2$

Rule 2: if $x_1$ is $A_2$ and $x_2$ is $B_2$, then $f_2 = C_0^2 + C_1^2 x_1 + C_2^2 x_2$

Then the output of this layer is

$$O_i^4 = \overline{w}_i f_i = \overline{w}_i(C_0^i + C_1^i x_1 + C_2^i x_2),\ i=1,2 \quad (6)$$

where:

| | |
|---|---|
| $O_i^4$: | the output of layer IV |
| $f_i$: | consequence value of the *i*th rule |
| $C_0^i, C_1^i, C_2^i$: | changeable parameters of the consequence |

*Layer V*: it is the output layer of the system. The adaptive nodes labeled as ∑ calculate the overall output as the summation of all incoming signals from the 4th layer. The output of this layer is given by

$$O_i^5 = \sum_{i=1}^{2} \overline{w_i} f_i \qquad (7)$$

where:

$O_i^5$ : the output of layer V

It is seen that adaptive nodes in layer I and layer IV with changeable parameters ($a_i^j$ and $b_i^j$) and ($C_0^i$, $C_1^i$ and $C_2^i$) will be adjusted during the training process by neural network's learning algorithm.

## ACCIDENT PREDICTION MODEL BASED ON FNN

### Data Description

In order to develop the accident prediction model, a 5-year time period (1999-2004) data was collected, and the data included detailed information on accidents, traffic flow and road conditions of 133 main segments from urban arterials in Harbin city of China.

Accident data were collected from official records covering all police recorded accidents. For each accident, there was a description of the date and location of accident, weather conditions, type and severity of accident, number and type of vehicles involved, and number of person death and injured. From 1999 to 2004, there were 12100 accidents recorded at these segments, which were 2420 accidents per year.

Traffic flow were mainly extracted from the monitoring videos of traffic police, and for roads with no video camera, complementary manual counting and flow estimating were carried out. Logistical difficulties precluded data collection on all roads in Harbin city and this was the main reason behind limiting the data to 133 samples. The AADT was at last determined from the average daily traffic, and the AADT values were ranging from 1680 to 72660 vehicles per segment per day.

Road conditions were collected from the Harbin Municipal Bureau, and the information included length of segments, pavement width, number of lanes, and LW. It is mentioned that SL information can also be obtained from the bureau and the values were 40km/h and 60km/h for these arterials in Harbin city.

### Variables Selection

Road accidents relate to geometry, traffic, environment, vehicle and driver factors, and different prediction models focus on different factors. For example, Persaud and Dzbik(1993) just took traffic flow into consideration; Persaud et al. (2000) related the accident frequency with traffic flow and road geometry; both of Golob and Recker (2003) and Knuiman et al. (1993) considered traffic flow, weather and lighting conditions; Hauer (2004) took AADT, percentage of trucks, geometry, speed limit and access points as input variables of his prediction model; Caliendo et al. (2007) considered the sight distance besides the road, traffic and weather conditions.

Above all, traffic flow and road condition are the two main factors most of researchers adopted and proven to be efficient. According to the data available, AADT, LW, SL and TL are selected as input variables, and AF for each respective segment is selected as output variable. The statistical characteristics of the selected variables are shown in Table 1.

Table 1 Statistical values of the selected variables

| | Minimum | Maximum | Mean | Standard Deviation |
|---|---|---|---|---|
| AADT (vehicles per day) | 1680 | 72660 | 15923 | 13958 |
| LW (m) | 2.75 | 4.00 | 3.50 | 0.41 |
| SL (km/h) | 40.00 | 60.00 | 46.47 | 0.39 |
| TL | 0.04 | 1.49 | 0.47 | 0.32 |
| AF (accidents per km per year) | 0 | 39.40 | 9.66 | 9.49 |

## Model Training and Testing

The entire data set with 133 collected samples is divided into training and testing subsets randomly, which are used for learning and validating the model, respectively. To ensure the subsets covering all possible combinations, the training set is composed by 78 samples, and the testing set is by 55 samples. All the input variables of both training and testing data are normalized for better generalization.

ANFIS tool box in Matlab is used to build the model, and the FNNM for road accident prediction is shown in Figure 2. The building process consists of five steps.

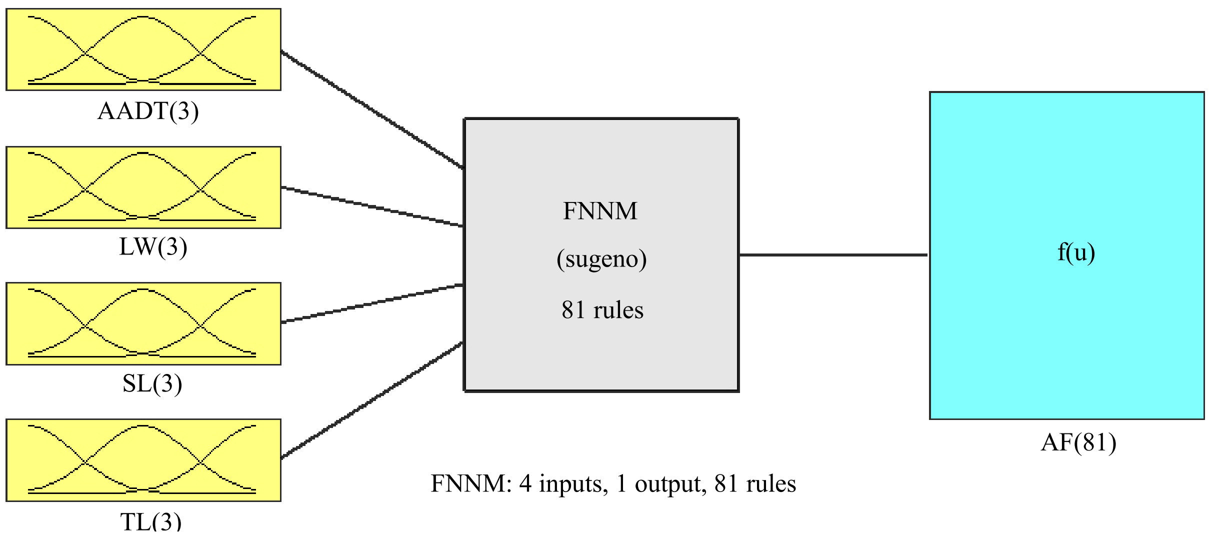



Figure 2 The architecture of FNNM

*Step I*: this step is to determine the optimal number of fuzzy sets for input variables of training data, here *k*-means clustering is employed and the mean of silhouette value is taken as the criterion to determine the number of clusters. Finally, 3 clusters for each input variable is determined, labeled as NL (negative large), ZO (zero), and PL (positive large).

*Step II*: this step is to initialize the parameters in premise ($a_i^j$ and $b_i^j$) and consequence ($C_0^k$, $C_1^k$, $C_2^k$, $C_3^k$ and $C_4^k$), and the initial values of the premise are shown in Table 2 and those of the consequence are all 0.

Table 2 Initial values of premise parameters

| j \ i | $a_i^j$ | | | $b_i^j$ | | |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 1 | 2 | 3 |
| 1 | 0 | 0.50 | 1.00 | 0.21 | 0.21 | 0.21 |
| 2 | 0.72 | 0.86 | 1.00 | 0.06 | 0.06 | 0.06 |
| 3 | 0.67 | 0.83 | 1.00 | 0.07 | 0.07 | 0.07 |
| 4 | 0 | 0.50 | 1.00 | 0.21 | 0.21 | 0.21 |

*Step III*: this step is to generate the fuzzy inference system, after trying both the grid partition method and sub. clustering method, the former is finally selected. Although this method will consume more time, it can reach higher accuracy. By this method, 81 effective fuzzy rules were determined (i.e., *k*=81).

*Step IV*: this step is to train the model by back-propagation method, and after 6000 times of iteration, the optimal network with minimum error is obtained.

*Step V*: the last step is to test the optimal model, and the result is shown in Figure 3.

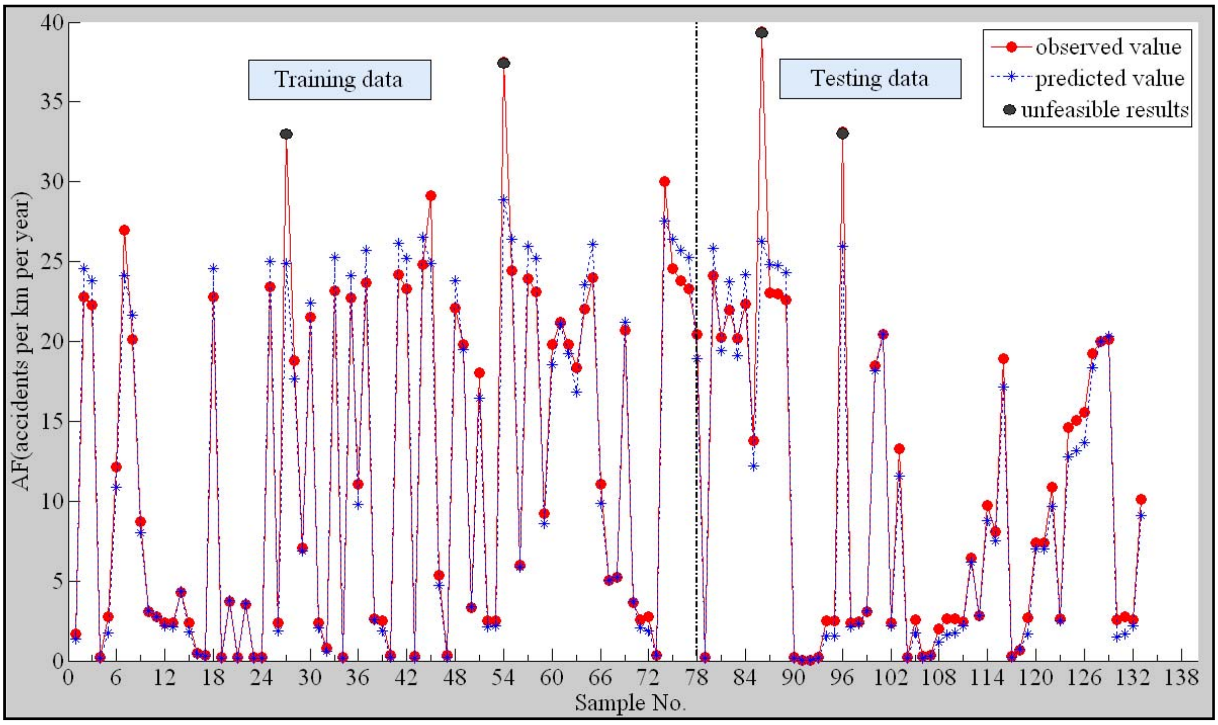
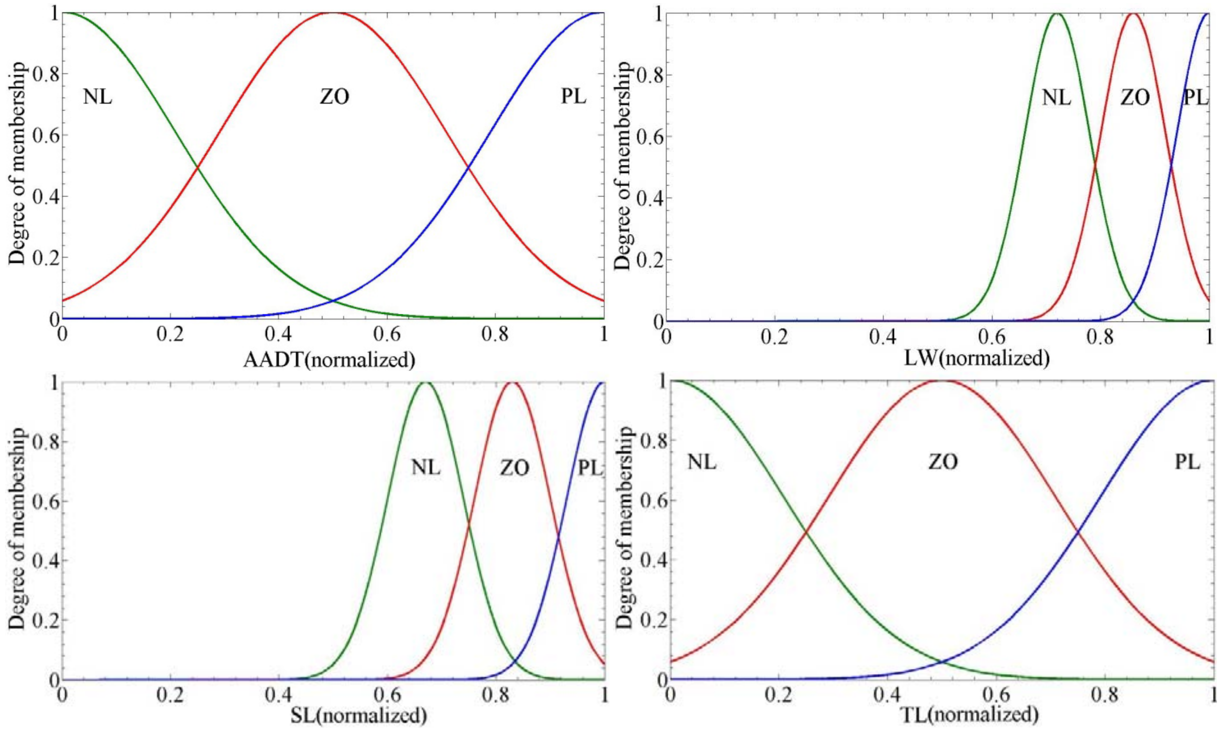
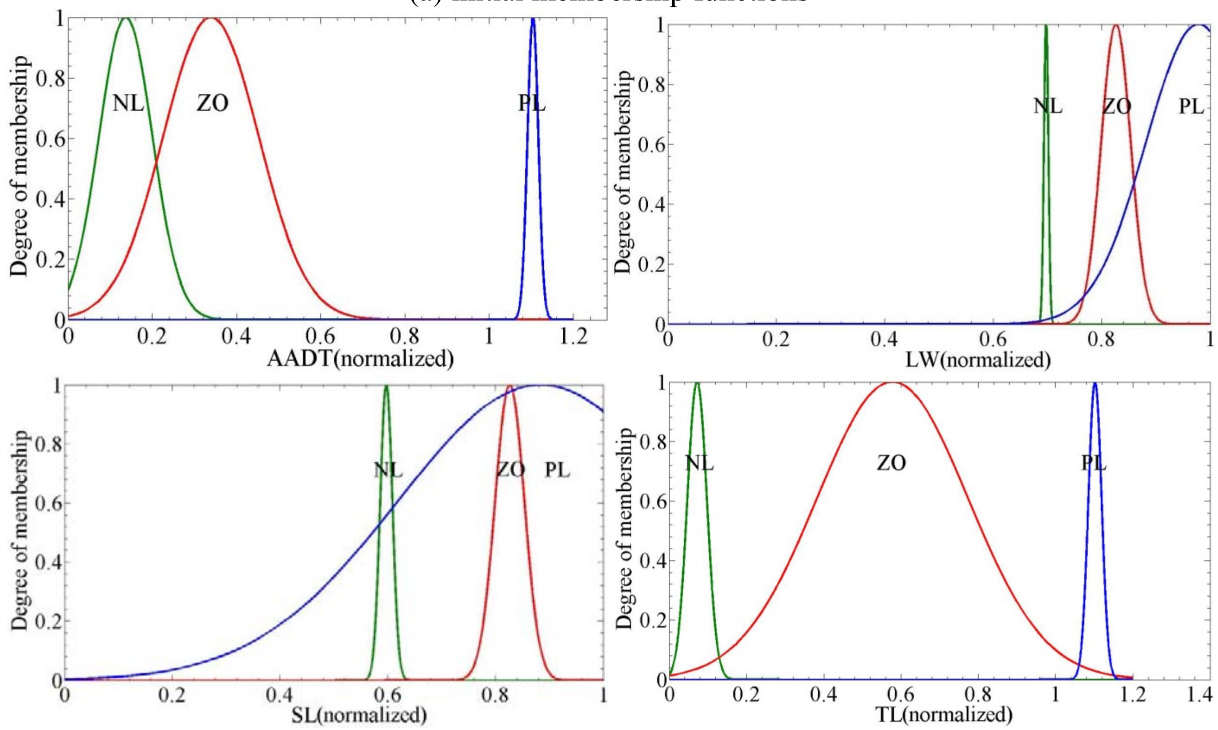



Figure 3 Prediction results of the FNNM




(a) initial membership functions




(b) final membership functions

Figure 4 Initial and final membership functions of the premise

After the training of the model, the parameters of both premise and consequence are adjusted automatically. The initial and final membership functions for the four input variables are shown in Figure 4. The final values of the premise parameters are shown in Table 3. The final values of the consequence parameters are shown in appendix.

Table 3 Final values of premise parameters

| $j$ \ $i$ | $a_i^j$ 1 | $a_i^j$ 2 | $a_i^j$ 3 | $b_i^j$ 1 | $b_i^j$ 2 | $b_i^j$ 3 |
|---|---|---|---|---|---|---|
| 1 | 0.1368 | 0.3392 | 1.104 | 0.0637 | 0.1131 | 0.0126 |
| 2 | 0.6977 | 0.8365 | 0.9789 | 0.0039 | 0.0372 | 0.0969 |
| 3 | 0.5973 | 0.8264 | 0.8842 | 0.0111 | 0.0271 | 0.2650 |
| 4 | 0.0719 | 0.5778 | 1.102 | 0.0251 | 0.1966 | 0.0168 |

From Table 3 and Figure 4, it can be seen that considerable changes occurred in the membership functions of all the four input variables during the training process. Taking the PL membership function of SL as an example, the central point ($a_3^3$) changed from 1 to 0.8842, and the width ($b_3^3$) changed from 0.07 to 0.265, which makes it looks much wider than the initial one. Moreover, it is found that the PL membership functions of both AADT and TL are out of the specified input range of 0 to 1, which means that both membership functions do not work during the training and testing process, and this may be the main reason why unfeasible outputs of the proposed model appear (i.e., samples of 27, 57, 86 and 96 whose errors are much larger than others). In the authors' opinion, the appearance of this strange phenomenon may be mainly caused by that there are not enough samples of extremely high risk segments with AF more than 30 accidents per km per year in this study.

## PERFORMANCE EVALUATION OF FNNM

### Comparisons

To demonstrate the performance of the proposed FNN prediction model, comparisons among FLM, NNM and FNNM are carried out. The FLM proposed by Meng et al. (2009) adopts the Mamdani style fuzzy inference system with 41 effective fuzzy rules, and the numbers of fuzzy sets are 3, 3, 2, 4 and 5 for AADT, LW, SL, TL and AF, respectively. The NNM is a 4-layer BP neural network, an input layer with 4 neurons, an output layer with 1 neuron, and two hidden layers with 12 neurons for each. The learning method of NNM is gradient descent with adaptive learning rate back-propagation algorithm. The same training data as that of the FNNM is used to train the FLM and NNM, respectively. Thirty randomly selected samples from the testing data are taken as the inputs and outputs for the three pre-trained models, and the samples and prediction results are listed in Table 4.

All the three prediction models are evaluated in terms of four performance measures: root mean square error (RMSE) which means the average deviation of the observed values to predicted values, the maximum relative error (MRE) which is measure for the largest error, the mean absolute percentage error (MAPE) which is measure for the average error and the goodness of fit ($R^2$) which generally takes value from 0 to 1, and the larger of the $R^2$ the regression points tend to align more accurately along the model curve. The measures are calculated by the following equations:

$$RMSE = \sqrt{\sum_{i=1}^{n}(Yd_i - Y_i)^2 / n} \tag{8}$$

$$MRE = \max(|Yd_i - Y_i| / Y_i \times 100) \tag{9}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}(|Yd_i - Y_i| / Y_i \times 100) \tag{10}$$

$$R^2 = 1 - (\sum_{i=1}^{n}(Yd_i - Y_i)^2) / (\sum_{i=1}^{n} Y_i^2) \tag{11}$$

where:

$n$: the number of samples

$Yd_i$: the observed value for the $i$th sample

$Y_i$: the predicted value for the $i$th sample

Table 4 Samples and prediction results for comparison experiments

| Sample No. | AADT (vehicles per day) | LW (m) | SL (km/h) | TL | AF(accidents per km per year) | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | Observed | FLM | NNM | FNNM |
| 1 | 2002 | 2.75 | 60 | 0.11 | 0.249 | 0.600 | 7.931 | 0.154 |
| 2 | 25988 | 3.25 | 60 | 1.37 | 24.114 | 27.000 | 24.016 | 25.820 |
| 3 | 18975 | 3.50 | 60 | 0.33 | 20.227 | 21.900 | 17.125 | 19.439 |
| 4 | 25455 | 3.50 | 60 | 0.45 | 21.938 | 22.200 | 22.381 | 23.714 |
| 5 | 18640 | 3.67 | 60 | 0.33 | 20.180 | 22.200 | 19.075 | 19.115 |
| 6 | 26970 | 3.67 | 60 | 0.47 | 22.349 | 22.200 | 22.665 | 24.206 |
| 7 | 15485 | 3.75 | 60 | 0.27 | 13.793 | 6.900 | 14.797 | 12.181 |
| 8 | 72660 | 2.95 | 60 | 0.63 | 39.400 | 6.900 | 39.071 | 26.278 |
| 9 | 26060 | 3.75 | 60 | 0.69 | 23.049 | 22.200 | 23.368 | 24.848 |
| 10 | 25626 | 3.93 | 60 | 0.57 | 22.991 | 24.600 | 22.993 | 24.782 |
| 11 | 3577 | 3.50 | 40 | 0.2 | 1.997 | 1.200 | 1.325 | 1.168 |
| 12 | 3831 | 3.50 | 40 | 0.22 | 2.631 | 1.200 | 1.465 | 1.638 |
| 13 | 4100 | 3.50 | 40 | 0.23 | 2.624 | 1.200 | 1.547 | 1.756 |
| 14 | 6460 | 3.50 | 40 | 0.37 | 2.458 | 1.200 | 2.604 | 2.161 |
| 15 | 12793 | 3.50 | 40 | 0.37 | 6.405 | 7.200 | 9.017 | 6.180 |
| 16 | 8535 | 3.50 | 40 | 0.49 | 2.840 | 6.900 | 3.726 | 2.818 |
| 17 | 14225 | 3.50 | 40 | 0.72 | 9.726 | 6.000 | 10.053 | 8.784 |
| 18 | 13460 | 3.50 | 40 | 0.77 | 8.055 | 6.300 | 8.908 | 7.510 |
| 19 | 17950 | 3.50 | 40 | 1.03 | 18.905 | 7.500 | 16.467 | 17.146 |
| 20 | 4998 | 3.75 | 40 | 0.14 | 0.266 | 0.600 | 1.505 | 0.208 |
| 21 | 16230 | 3.75 | 40 | 0.46 | 14.609 | 15.900 | 14.869 | 12.760 |
| 22 | 16404 | 3.75 | 40 | 0.47 | 15.035 | 16.200 | 15.264 | 13.152 |
| 23 | 16645 | 3.75 | 40 | 0.48 | 15.594 | 15.900 | 15.778 | 13.687 |
| 24 | 19068 | 3.75 | 40 | 0.54 | 19.216 | 21.900 | 19.229 | 18.343 |
| 25 | 20374 | 3.75 | 40 | 0.58 | 20.000 | 21.900 | 20.266 | 20.016 |
| 26 | 20680 | 3.75 | 40 | 0.59 | 20.119 | 21.900 | 20.469 | 20.319 |
| 27 | 3685 | 4.00 | 40 | 0.21 | 2.581 | 3.200 | 1.822 | 1.506 |
| 28 | 3773 | 4.00 | 40 | 0.22 | 2.771 | 3.200 | 1.832 | 1.708 |
| 29 | 6180 | 4.00 | 40 | 0.35 | 2.559 | 3.200 | 2.217 | 2.192 |
| 30 | 14381 | 4.00 | 40 | 0.41 | 10.095 | 8.900 | 10.326 | 9.082 |

The statistical values of RMSE, MRE, MAPE and $R^2$ of the models are given in Table 5. From Table 5, it is found that the performance of the proposed FNNM is the best according to MRE, which is the lowest compared to other models. The MRE of FNNM is 71.3%, while those of the FLM and NNM are 471.0% and 96.9%, respectively. The MAPE, RMSE and $R^2$ of FNNM are almost the same with NNM, and much better than FLM. The statistical indicators of FLM show that the performance of FLM is the poorest when compared to the other two models and are 6.615, 471.0%, 51.9% and 0.796 for RMSE, MRE, MAPE and $R^2$, respectively. A probable reason for the MRE of FLM reaching 471.0% is the incompletion of fuzzy rule sets. However, the performance measures also emphasize the fact that overall performance of the three models are all acceptable, since the $R^2$ of the models are all more than 0.7 in terms of goodness of fit, which are 0.796, 0.988 and 0.969 for FLM, NNM and FNNM, respectively.

Table 5 The statistical values of the models

| | RMSE | MRE | MAPE | $R^2$ |
|---|---|---|---|---|
| FLM | 6.615 | 471.0% | 51.9% | 0.796 |
| NNM | 1.754 | 96.9% | 20.7% | 0.988 |
| FNNM | 2.672 | 71.3% | 21.3% | 0.969 |

### Sensitivity Analysis

In order to reveal how the inputs work on output, as well as to identify the significant factors influencing the occurrence of accidents, sensitivity analysis is conducted. The basic idea is to perturb the inputs of the model by using the mean plus (or minus) a user-defined number of standard deviations, while all other inputs are fixed at their respective means, and the corresponding changing is calculated and recorded as an absolute percentage change (APC) above and below the mean of that output. The process is repeated for each input in the same way. Finally, a report was generated which summarizes the variation of output with respect to the variation of each input (Delan et al., 2006).

Table 6 Results of sensitivity analysis

| AADT(vehicles per day) | | LW(m) | | SL(km/h) | | TL | |
|---|---|---|---|---|---|---|---|
| E | APC | E | APC | E | APC | E | APC |
| -σ | 74.6% | -2σ | 6.7% | -40σ | 6.7% | -σ | 90.5% |
| +σ | 72.4% | -σ | 3.8% | -15σ | 19.0% | -0.5σ | 1.9% |
| +2σ | 75.2% | -0.5σ | 2.9% | -5σ | 24.8% | +σ | 3.8% |
| +3σ | 81.9% | +0.5σ | 1.9% | +5σ | 34.3% | +2σ | 7.6% |
| +4σ | 92.3% | +σ | 2.9% | +15σ | 19.0% | +3σ | 11.4% |
| +5σ | 84.6% | +2σ | 5.7% | +40σ | 27.8% | +4σ | 14.3% |
| Mean | 80.2% | | 4.0% | | 21.9% | | 21.6% |

E and σ are the respective mean and standard deviation for each variable; the tables filled with gray color are the values with inputs out of the specified range.

In this study, the defined numbers of standard deviations mentioned above are diverse for different input variables, since the standard deviations varied a lot (see Table 1) and the final inputs for the FNNM should be maintained above 0. To demonstrate the potential of the proposed model, some inputs with value above 1 which is out of the specified range are also taken into consideration. The results of sensitivity analysis are shown in Table 6. Examination of the sensitivity analysis results reveals that AADT is the most significant predictor for the proposed model, followed by SL, TL and LW in order of their relative importance going from the most to the least significant, and the average of the sensitivity values are 80.2%, 21.9%, 21.6% and 4.0% for AADT, SL, TL and LW, respectively. TL has a significantly larger sensitivity value when taking (E-σ) as input, while taking (E-0.5σ) as input the sensitivity value becomes much smaller. This means that TL is a significant predictor only in extremely smaller value area. The sensitivity analysis results also highlights that the proposed model has a good adaptability to a certain extent, since when the inputs are out of the specified range, the sensitivity values still keep the same trend as the values of inputs in the specified range.

## CONCLUSIONS

In this study, fuzzy neural network is applied to predict road accident frequencies as an alternative to more conventional accident prediction approaches. The proposed FNNM is established using a data set of 133 segments from urban arterials in Harbin city of China, with AADT, LW, SL and TL as input variables and AF as output variable. In order to evaluate the performance of the model, comparisons among FLM, NNM and FNNM are carried out, and statistical values of RMSE, MRE, MAPE and $R^2$ are employed as measures. Furthermore, a sensitivity analysis is carried out to identify the significant factors as well as to demonstrate the potential of the proposed FNNM.

Through these comparisons and analysis, it can be concluded that: (i) the FNNM is a consistent alternative to the NNM and much better than the FLM in terms of the four statistical measures. But just as mentioned earlier, the establishment of the NNM is a very complex and time consuming work, while that of the FNNM is much easier. In a word, the proposed FNNM is a more accurate, flexible and time saving model than NNM and FLM for the prediction of road accident frequencies. (ii) AADT is the most significant factor in the proposed model with the average sensitivity value as high as 80.2%. The order of the four inputs influencing the output is AADT, SL, TL and LW according to their relative importance going from the most to the least significant. However, TL becomes very significant when its value is less than 0.2, then the sensitivity value is as high as 90.5%, almost 10 times of its average value. This result is a bit surprising and needs further research. Another direction for future work is to focus on resolving the problem that the proposed FNNM cannot predict the extremely high accident frequencies, and a more extensive data set, taking more variables (i.e. road surface conditions) into consideration, would be helpful.

## ACKNOWLEDGEMENTS

The study is partially supported by Key Laboratory of Advanced Materials & Intelligent Control Technology on Transportation Safety authorized by Ministry of Communications of China. The authors also would like to thank Harbin Traffic Police Detachment and Harbin Municipal Bureau for providing the useful data relating to accidents, traffic flow and road conditions. The constructive comments from anonymous reviewers are also highly appreciated.

## REFERENCES


Alotaibi, F.D., Abdennour, A., and Ali, A.A. (2008). "A robust prediction model using ANFIS based on recent TETRA outdoor RF measurements conducted in Riyadh city-Saudi Arabia". *International Journal of Electronics and Communications* 62, 674-682.

Alyisi, S., and Franchini, M. (2011). "Fuzzy neural networks for water level and discharge forecasting with uncertainty". *Environmental Modelling& Software* 26, 523-537.

Azamathulla, H.M., Chang, C.K., Ghani, A.A., and Ariffin, J. (2009). "An ANFIS-based approach for predicting the bed load for moderately sized rivers". *Journal of Hydro-environment Research* 3, 35-44.

Cafiso, S., Graziano, A.D., Silvestro, G.D., Cava, G.L., and Persaud, B. (2010). "Development of comprehensive accident models for two-lane rural highways using exposure, geometry, consistency and context variables". *Accident Analysis and Prevention* 42, 1072-1079.

Caliendo, C., Guida, M., and Parisi, A. (2007). "A crash-prediction model for multilane roads", *Accident Analysis and Prevention* 39, 657-670.

Chang, L.Y. (2005). "Analysis of freeway accident frequencies: negative binomial regression versus artificial neural network". *Safety Science* 43, 541-557.

Delen, D., Sharada, 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, 434-444.

Dionne, G., Desjardins, D., Laberge-Nadeau, C., and Maag, U. (1993). "Medical conditions, risk exposure and truck drivers' accidents: an analysis with count data regression models". *37th Annual Meeting of the Association for the Advancement of Automotive Medicine*. San Antonio, Texas.

EI-Basyouny, K., and Sayed, T. (2006). "Comparison of two negative binomial regression techniques in developing accident prediction models". *Transportation Research Record* 1950, 9-16.

EI-Basyouny, K., and Sayed, T. (2009). "Accident prediction models with random corridor parameters". *Accident Analysis and Prevention* 41, 1118-1123.

Ekici, B.B., and Aksoy, U.T. (2011). "Prediction of building energy needs in early stage of design by using ANFIS". *Expert Systems with Applications* 38, 5352-5358.

Golob, T.F., and Recker, W.W. (2003). "Relationship among urban freeway accidents, traffic flow, weather, and lighting conditions". *Journal of Transportation Engineering* 129, 342-353.

Greibe, P. (2003). "Accident prediction models for urban roads". *Accident Analysis and Prevention* 35, 273-285.

Hauer, E. (2004). “Safety models for urban four-lane undivided road segments”. *83rd Annual Meeting Transportation Research Board*, Washington, DC.

Jang, J.R. (1993). “ANFIS: adaptive-network-based fuzzy inference system”. *IEEE Transaction on Systems, Man and Cybernetics* 23(3): 665-685.

Knuiman, M.W., Council, F.M., and Reinfurt, D.W. (1993). “Association of median width and highway accident rates”. *Transportation Research Record* 1401.

Lord, D., and Persaud, B. (2000). “Accident prediction models with and without trend: application of the generalized estimating equation”. *Transportation Research Record* 1717, 102-108.

Lord, D., Washington, S.P., and Ivan, J.N. (2005). “Poisson, Poisson-gamma and zero inflated regression models of motor vehicle crashes: balancing statistical fit and theory”. *Accident Analysis and Prevention* 37(1), 35-46.

Maher, M.J., and Summersgill, I. (1996). “A comprehensive methodology for the fitting of predictive accident models”. *Accident Analysis and Prevention* 28(3), 281-296.

Meng, X.H., Zheng, L., and Qing, G.M. (2009). “Traffic accidents prediction and prominent influencing factors analysis based on fuzzy logic”. *Journal of Transportation Systems and Information Technology* 9(2), 87-92.

Miaou, S.P., Hu, P.S., Wright, T., Rathi, A.K., and Davis, S.C. (1992). “Relationships between truck accidents and highway geometric design: a Poisson regression approach”. *Transportation Research Record* 1376, 10-18.

Miaou, S.P. (1994). “The relationship between truck accidents and geometric design of road sections: Poisson versus negative binomial regressions”. *73rd Annual Meeting Transportation Research Board*, Washington, DC.

Okamoto, H., and Koshi, M.A. (1989). “A method to cope with the random errors of observed accident rates in regression analysis”. *Accident Analysis and Prevention* 21, 317-332.

Ozbay, K., and Noyan, N. (2006). “Estimation of incident clearance times using Bayes Network approach”. *Accident Analysis and Prevention* 38, 542-555.

Persaud, B., and Dzbik, L. (1993). “Accident prediction models for freeways”. *Transportation Research Record* 1401, 55-60.

Persaud, B., Retting, R.A., and Lyon, C. (2000). “Guidelines for the identification of hazardous highway curves”. *Transportation Research Record* 1717, 14-18.

Wei, M., Bai, B., Sung, A.H., Liu, Q.Z., Wang, J.C., and Cather, M.E. (2007). “Predicting injection profiles using ANFIS”. *Information Sciences* 177, 4445-4461.

Xiao, J., Kulakowski, B.T., and EI-Gindy, M. (1999). "Prediction of risk of wet pavement accidents using fuzzy logic". *Transportation Research Record* 1717, 28-36.

**APPENDIX**

Final values of the consequence parameters

| | $C_0$ | $C_1$ | $C_2$ | $C_3$ | $C_4$ |
|---|---|---|---|---|---|
| *Output1* | 0.0005916 | 0.0047820 | 0.0042680 | 0.0013710 | 0.0064030 |
| *Output2* | 0.0018450 | 0.0080780 | 0.0071680 | 0.0042970 | 0.0107500 |
| *Output3* | 0.0003592 | 0.0009655 | 0.0008042 | 0.0009486 | 0.0012060 |
| *Output4* | 0.0000317 | 0.0001596 | 0.0001678 | 0.0000469 | 0.0002123 |
| *Output5* | 0.0000543 | 0.0001971 | 0.0001949 | 0.0001059 | 0.0002612 |
| *Output6* | 0.0000254 | 0.0000604 | 0.0000683 | 0.0000656 | 0.0000784 |
| *Output7* | 0.0009984 | 0.0031940 | 0.0042080 | 0.0009262 | 0.0042070 |
| *Output8* | 0.0010220 | 0.0028920 | 0.0038050 | 0.0013390 | 0.0038070 |
| *Output9* | 0.0008059 | 0.0017810 | 0.0023710 | 0.0020650 | 0.0023710 |
| *Output10* | 0.0003967 | 0.0029250 | 0.0022340 | 0.0008082 | 0.0033520 |
| *Output11* | 0.0015190 | 0.0064690 | 0.0048870 | 0.0033900 | 0.0073300 |
| *Output12* | 0.0012680 | 0.0032630 | 0.0025720 | 0.0034680 | 0.0038570 |
| *Output13* | 0.0000503 | 0.0002336 | 0.0002210 | 0.0000603 | 0.0002611 |
| *Output14* | 0.0001019 | 0.0003542 | 0.0003099 | 0.0001692 | 0.0003961 |
| *Output15* | 0.0000598 | 0.0001481 | 0.0001338 | 0.0001540 | 0.0001726 |
| *Output16* | 0.0009142 | 0.0034150 | 0.0037820 | 0.0008036 | 0.0037820 |
| *Output17* | 0.0013730 | 0.0035410 | 0.0039500 | 0.0015830 | 0.0039500 |
| *Output18* | 0.0006393 | 0.0014780 | 0.0016800 | 0.0014850 | 0.0016800 |
| *Output19* | 0.0018370 | 0.0117600 | 0.0082680 | 0.0033590 | 0.0124000 |
| *Output20* | 0.0102300 | 0.0522200 | 0.0368400 | 0.0210300 | 0.0552600 |
| *Output21* | 0.0006403 | 0.0019140 | 0.0014240 | 0.0016800 | 0.0021360 |
| *Output22* | 0.0001411 | 0.0005580 | 0.0005006 | 0.0001521 | 0.0005751 |
| *Output23* | 0.0002753 | 0.0010460 | 0.0008418 | 0.0004117 | 0.0010690 |
| *Output24* | 0.0000220 | 0.0000593 | 0.0000534 | 0.0000448 | 0.0000626 |
| *Output25* | 0.0061550 | 0.0295800 | 0.0299700 | 0.0052930 | 0.0332900 |
| *Output26* | 0.0116400 | 0.0455100 | 0.0406800 | 0.0145200 | 0.0489100 |
| *Output27* | 0.0008621 | 0.0020110 | 0.0021720 | 0.0017780 | 0.0021720 |
| *Output28* | 0.0002310 | 0.0011160 | 0.0009945 | 0.0004194 | 0.0014920 |
| *Output29* | 0.0043520 | 0.0099490 | 0.0088000 | 0.0070250 | 0.0132000 |
| *Output30* | 0.0009349 | 0.0022840 | 0.0018960 | 0.0022650 | 0.0028430 |
| *Output31* | 0.0001118 | 0.0001717 | 0.0002194 | 0.0000696 | 0.0002295 |
| *Output32* | 0.0003492 | 0.0005006 | 0.0005895 | 0.0002953 | 0.0006620 |
| *Output33* | 0.0000678 | 0.0001494 | 0.0001735 | 0.0001613 | 0.0001937 |
| *Output34* | 0.0056990 | 0.0083080 | 0.0111100 | 0.0032860 | 0.0111100 |
| *Output35* | 0.0195200 | 0.0238300 | 0.0316000 | 0.0125600 | 0.0316000 |
| *Output36* | 0.0030070 | 0.0064480 | 0.0085570 | 0.0073370 | 0.0085570 |
| *Output37* | 0.0002419 | 0.0010700 | 0.0007852 | 0.0003869 | 0.0011780 |
| *Output38* | 0.0026200 | 0.0068840 | 0.0052080 | 0.0048200 | 0.0078110 |
| *Output39* | 0.0056640 | 0.0119000 | 0.0091530 | 0.0125000 | 0.0137300 |
| *Output40* | 0.0001023 | 0.0002564 | 0.0002662 | 0.0000838 | 0.0002820 |
| *Output41* | 0.0005565 | 0.0010690 | 0.0011120 | 0.0006409 | 0.0012020 |
| *Output42* | 0.0003508 | 0.0006758 | 0.0006279 | 0.0006685 | 0.0007654 |
| *Output43* | 0.0024630 | 0.0056710 | 0.0062730 | 0.0018340 | 0.0062730 |
| *Output44* | 0.0156900 | 0.0274500 | 0.0311800 | 0.0165400 | 0.0311800 |
| *Output45* | 0.0053650 | 0.0098790 | 0.0110500 | 0.0091610 | 0.0110500 |
| *Output46* | 0.0016830 | 0.0069410 | 0.0047860 | 0.0025350 | 0.0071780 |

Appendix (continued)

| | $C_0$ | $C_1$ | $C_2$ | $C_3$ | $C_4$ |
|---|---|---|---|---|---|
| *Output47* | 0.1835000 | 0.5770000 | 0.4239000 | 0.3364000 | 0.6359000 |
| *Output48* | 0.0125700 | 0.0251600 | 0.0191100 | 0.0259500 | 0.0286600 |
| *Output49* | 0.0002833 | 0.0008235 | 0.0007993 | 0.0002425 | 0.0008457 |
| *Output50* | 0.0009686 | 0.0023230 | 0.0021930 | 0.0010930 | 0.0023880 |
| *Output51* | 0.0002977 | 0.0005115 | 0.0004993 | 0.0004328 | 0.0005402 |
| *Output52* | 0.01706000 | 0.0478200 | 0.0499000 | 0.0127100 | 0.0496900 |
| *Output53* | 4.50200000 | 6.7700000 | 7.3570000 | 4.0070000 | 7.7310000 |
| *Output54* | 0.02987000 | 0.0498400 | 0.0538600 | 0.0465500 | 0.0538600 |
| *Output55* | 0.00000137 | 0.0000028 | 0.0000025 | 0.0000019 | 0.0000037 |
| *Output56* | 0.00004056 | 0.0000784 | 0.0000693 | 0.0000579 | 0.0001037 |
| *Output57* | 0.00001027 | 0.0000202 | 0.0000163 | 0.0000189 | 0.0000244 |
| *Output58* | 0.00014270 | 0.0001203 | 0.0001614 | 0.0000603 | 0.0001615 |
| *Output59* | 0.00060870 | 0.0004933 | 0.0006615 | 0.0002633 | 0.0006624 |
| *Output60* | 0.00001132 | 0.0000101 | 0.0000131 | 0.0000064 | 0.0000133 |
| *Output61* | 0.00667700 | 0.0056260 | 0.0075500 | 0.0028980 | 0.0075500 |
| *Output62* | 0.03989000 | 0.0320900 | 0.0431100 | 0.0171000 | 0.0431100 |
| *Output63* | 0.00046300 | 0.0003938 | 0.0005253 | 0.0002424 | 0.0005253 |
| *Output64* | 0.00000111 | 0.0000034 | 0.0000027 | 0.0000016 | 0.0000004 |
| *Output65* | 0.00017670 | 0.0002776 | 0.0002054 | 0.0002515 | 0.0003081 |
| *Output66* | 0.00043880 | 0.0006883 | 0.0005085 | 0.0006493 | 0.0007627 |
| *Output67* | 0.00002362 | 0.0000265 | 0.0000333 | 0.0000126 | 0.0000334 |
| *Output68* | 0.00013150 | 0.0001429 | 0.0001716 | 0.0000856 | 0.0001754 |
| *Output69* | 0.00008574 | 0.0001125 | 0.0001125 | 0.0001126 | 0.0001219 |
| *Output70* | 0.00067500 | 0.0006781 | 0.0008566 | 0.0003241 | 0.0008566 |
| *Output71* | 0.00350800 | 0.0038280 | 0.0047250 | 0.0022400 | 0.0047250 |
| *Output72* | 0.00183400 | 0.0023120 | 0.0024930 | 0.0023440 | 0.0024930 |
| *Output73* | 0.00000424 | 0.0000152 | 0.0000105 | 0.0000060 | 0.0000157 |
| *Output74* | 0.00033680 | 0.0005479 | 0.0003998 | 0.0004720 | 0.0005995 |
| *Output75* | 0.00062500 | 0.0009513 | 0.0006996 | 0.0008834 | 0.0010490 |
| *Output76* | 0.00000865 | 0.0000165 | 0.0000174 | 0.0000595 | 0.0000175 |
| *Output77* | 0.00006864 | 0.0001169 | 0.0001192 | 0.0000718 | 0.0001216 |
| *Output78* | 0.00012580 | 0.0001635 | 0.0001688 | 0.0001620 | 0.0001733 |
| *Output79* | 0.00050870 | 0.0008434 | 0.0009101 | 0.0003082 | 0.0009101 |
| *Output80* | 0.03495000 | 0.0310100 | 0.0394500 | 0.0176200 | 0.0394500 |
| *Output81* | 0.01038000 | 0.0130100 | 0.0138100 | 0.0133100 | 0.0138100 |

$C_0$, $C_1$, $C_2$, $C_3$ and $C_4$ are parameters of consequence