ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

Indian Journal of Science and Technology, *Vol 12(21), DOI: 10.17485/ijst/2019/v12i21/143108, June 2019*

Efficiency of Fuzzy Bayesian Inference in Predicting the Frequency of PDO Crashes on Urban Highways

Ehsan Abbasi* and Mansour Hadjihosseinlou

Khajeh Nasir Toosi University of Technology, Iran; ehsan.abbasi395@gmail.com, mansour@kntu.ac.ir

Abstract

Objective: In this article, the effect rate of speed and volume of traffic on occurrence of PDO crashes on urban highways are investigated using Fuzzy Bayesian inference and PDO accident data of Tehran urban highways is used as case study. **Methods/Findings:** To fuzzify the variables, values of turning points of Triangular Membership Functions (TMFs) are estimated using Bayesian inference and MCMC algorithms. To produce rules in each model, one or more variables are deemed effective in occurrence of crashes. The evaluated frequency of crashes by developed models is compared with the frequency of observed crashes. The results of comparison represent the accuracy of each model. The model with highest value of ^{2} is the best model and the variables deemed effective for that model are those which have got effect on crashes occurrence. The results of comparison between the effect of elements of traffic volume indicates that after speed, volume of Light non-passenger car Vehicles (LVs) is more effective than volume of Heavy Vehicles (HVs) and Passenger Cars (PCs) on crashes occurrence on urban highways. After that the part of volume of PCs is more prominent than volume of HVs in likelihood of PDO accident. **Application:** After prioritization of variables in terms of influence on occurrence of crashes, the authors employed the model best fitting the data with highest value of goodness of fit to do the sensitivity analyze. Sensitivity analyze specifies the effect rate of each variable on occurrence of crashes.

Keywords: Crashes, Fuzzy Bayesian Inference, Light Non-Passenger Car Vehicles, Sensitivity Analysis, Urban Highways, Speed

1. Introduction

The volume of traffic is comprised of different elements which could be categorized in three groups including PCs, HVs and LVs. HVs include minibuses, buses, trucks and trailers and LVs include taxis, pickups and motorcycles. The effect of traffic volume on likelihood of property and more severe crashes in roads safety is investigated in different researches through different statistical and mathematical methods[1–8]. The part of HVs in various segments of urban and rural roads network is studied many times more than the other elements of traffic volume. Furthermore, the influence of the presence of HVs in traffic flow on road safety has been studied in different methods.

Recently, Fuzzy Bayesian inference based models have been known as a perfect and powerful tool for modeling processes, which are complex with conventional quantitative techniques, or when the acquired information from modeling processes are qualitative or indefinite. Fuzzy Bayesian inference could be used for developing models and dealing with complicated systems.

In[9] conducted an analysis of bus-involved accident data to assess the safety impact of introducing bus priority measures in Metropolitan Melbourne, Australia. They developed two models; mixed-effects negative binomial and neural network based on the back-propagation approach to explore key traffic, transit and route factors associated with bus accident frequency.

The results of mixed-effects negative binomial model indicated that bus accident frequency increases with traffic volume, route length, service frequency and stop density. The developed a non-parametric Classification and Regression Tree (CART) model to develop the

*Author for correspondence

empirical relationship between injury severity results and driver/vehicle characteristics, highway geometric factors, environmental features, and accident variables. The results of their research showed that drinking-driving, seatbelt use, vehicle and collision type, contributing circumstance and driver/vehicle action, frequency of vehicles involved in the accident and accident location were the most effective factors of injury severity outcomes in truck crashes. The safety problem of HVs has ever been of important concern for managers and officials of road safety and community, and has involved many researchers of traffic safety area such as. In[10–13] researches on traffic safety, researcher intends to predict traffic crashes due to potential conflict situation regarding road geometry, traffic volume composition, weather condition, drivers' characteristics or behaviors. Such intention usually appears in format of a model which relates various potential factors to output variable that is normally frequency or likelihood of crashes both property and more severe crashes.

Recent scholars have utilized different methods to investigate the road safety under flow of HVs. Part of them studied the operational features and design criterions to predict factors affecting the road safety among many potential factors that each scholar studied on and concluded about its effect. This methodology intends to assess the dimension and elements of HVs performance on urban or rural highways to let them drive on roads facilitated with standard geometry[14–17].

In[18,19] in their research estimated the level of accident severity at intersections using environmental and traffic factors through a back propagation neural network model and the generalized linear mixed model that used an analytical approach. Counting passing flow shows the number of conflicts (potential accident) occurs on road intersections under study. In[20] investigated the frequency of single vehicle crashes in terms of geometric factors comprising horizontal and vertical curves, traffic flow related factors, weather condition, cross-sectional factors, and roadside features. Results of their research implied that rainfall during the accident is positively associated with single vehicle crashes, but real-time visibility is negatively associated. The presence of a road shoulder along mountainous highways is associated with less frequency of single vehicle crashes.

The factors that lead to violation from speed limit for large truck drivers of Taiwan. The results of their research indicated that the factors which influenced speeding offense were not related to job experience. Rather, the driver's demographics including age and education, mental condition i.e. sleep quality, and driving status i.e. yearly distance driven and driving late at night were significantly associated with violation from speed limit. In[21] studied the effect of cell phones on truck accident rates. Cell phones were found to have a significant effect on these rates. A nonlinear model was evaluated by a set of exact specification requirements. The model suggested a non-linear impact of cell phone usage on truck accident rates. Hence, cell phones have a positive effect on crashes but at a decreasing rate.

Keeping a proper level of friction is a crucial maintenance practice, because of the effect it has on roadway safety. In[22] came up with a fuzzy logic inference system in their research to predict the rate of vehicle crashes based on traffic level, speed limit, and surface friction. In fuzzy controllers were used to develop the model. The results of research provided a decision support model for highway agencies to monitor the network's friction and make sound judgments to remove flaws based on crash risk. Furthermore, the model could be implemented in the vehicle periphery to warn drivers of slippery locations in ice or rainy periods. The accident data of Tehran urban highways in the period from 2012 to 2017 are used to model crashes on urban highways in this study. These data were gathered by the traffic and transportation organization of Tehran. For collecting the accident statistics, 100 kilometers of highways are divided into 100 parts with constant length equal to 1 kilometer. In addition, the frequency of crashes in each part is collected for different hours of the day which were divided into first traffic peak hours, second the day non-peak hours and third night non-peak hours. Thus there are 300 data available for 300 subparts.

The objective of this research is to determine the effect rate of speed, volume of PCs, HVs and LVs on likelihood of crashes on urban highways. To conduct so, Fuzzy Bayesian Network for modeling crashes on urban highways is used based on the gathered data of crashes and independent variables. The stages of developing models include Fuzzification of input and output variables, Rule production, Composition or Aggregation of diagrams and Defuzzification. After developing the models for predicting the frequency of crashes on urban highways, the evaluated frequency of PDO (property damage only) crashes by the models are compared with the frequency of observed crashes for each combination of effective factors. The results of comparison represent the accuracy of each model which is developed based on the role of variables that considered effective in occurrence of PDO crashes on urban highways.

2. Methodology

Fuzzy modeling is a four staged procedure including fuzzification of input and output variables, rule production, composition and aggregation of diagrams and defuzzification. To fuzzify the variables, the data available as scatter diagrams are used as prior distribution. This research presents a novel method to learn from data by exploiting a rule foundation based on Fuzzy Inference System (FIS) in which the turning points of fuzzy Membership Functions (MFs) are determined by Bayesian inference and Markov Chain Monte Carlo (MCMC) algorithms.

Linguistic terms low-medium-high are usually applied for developing Fuzzy models in modeling crashes and a part could be simultaneously the member of more than one fuzzy set. Information flow through a fuzzy model requires that the input variables go through three major transformations before exiting the system as output information, which are known as fuzzification, rule production, composition and defuzzification. After fuzzification of input and output variables and establishing the rules based on the role of variables deemed effective in accident occurrence, the composition of diagrams is processed. The next step is defuzzification for obtaining the crisp output from assembled fuzzy output. For so, the Centroid method is applied. In this method, the center of area of the aggregation diagram is calculated. This centroid is the crisp output that is the frequency of predicted crashes. Through a similar process, for all values of volume and speed, the frequency of crashes is predicted. For doing calculations, MATLAB program, the Fuzzy Toolbox is utilized.

2.1 Fuzzification of the Variables and Producing Rules

Fuzzification encompasses two stages; obtaining the MFs for input and output variables and linguistic depicting of these functions. Triangular and trapezoidal MFs are applicable for modeling crashes with great deal of variations.

2.1.1 Data

Statistics of Property Damage Only (PDO) crashes and independent variables of the models are collected for 100 sections of Tehran urban highways in 300 subsections. The passenger car equivalent factors of vehicle types are given in Table 1.

The total volume (PC equivalent) of LVs and HVs is calculated as equation (1):

$$V_{te} = e_1 N_1 + \cdots + e_n N_n \tag{1}$$

where, V_{te} is the total equivalent volume and e_i and N_i are equivalent factor and number of vehicles of type i respectively.

The scatter diagrams of input and output.

2.1.2 Producing Rules

For each combination of factors, the corresponding rules are producible. In this research a novel method of learning from data by exploiting the rule based FISs is proposed in which the parameters are assessed by Bayesian inference and MCMC algorithms.

In the first case it is assumed that speed is the only efficient factor on occurrence of PDO crashes on urban highways. Hence the rule foundation applied to develop the FIS is as following:

1. If speed in high, then frequency of PDO crashes is high
2. If speed in low, then frequency of PDO crashes is low

2.1.3 Bayesian Parameter Estimation by MCMC Algorithm

The MFs of speed and frequency of PDO crashes are.

In the method used here, MCMC is applied to estimate the posterior distribution. Therefore, this method can work on any likelihood or prior distribution. In this study linguistic expert view inserted to the standard Bayesian

Table 1. The passenger car equivalent factors of vehicle types

Heavy vehicles	Motorcycle	Intracity bus	Intercity bus	Minibus	Pickup	Taxi	Passenger car
2.5	0.5	5	2.5	2	1	2	1

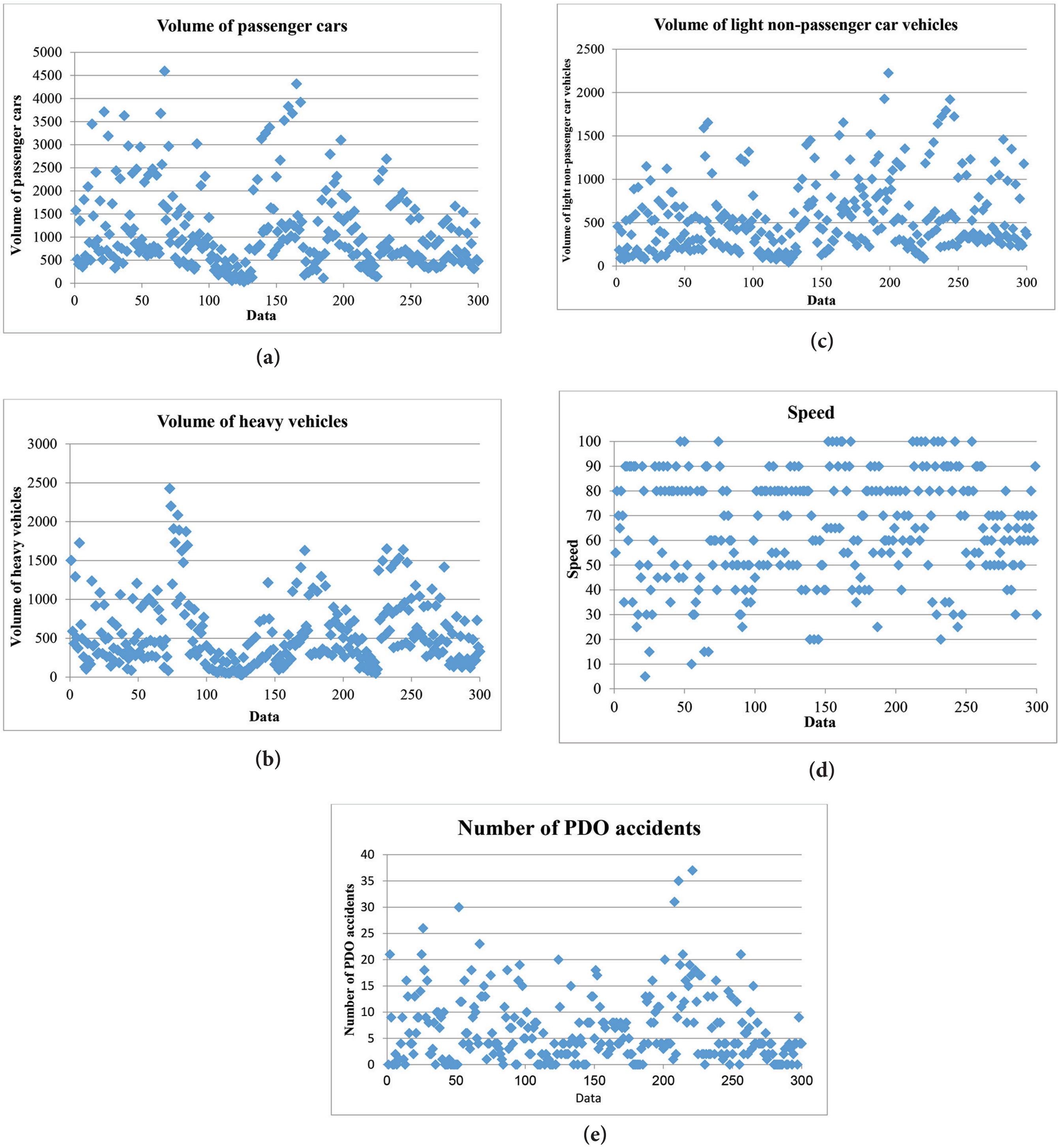


Figure 1. Scatter diagrams of input and output variables of models
(a) Scatter diagram of input variable volume of passenger cars
(b) Scatter diagram of input variable volume of heavy vehicles
(c) Scatter diagram of input variable volume of light non passenger car vehicles
(d) Scatter diagram of input variable speed
(e) Scatter diagram of output variable number of PDO accidents

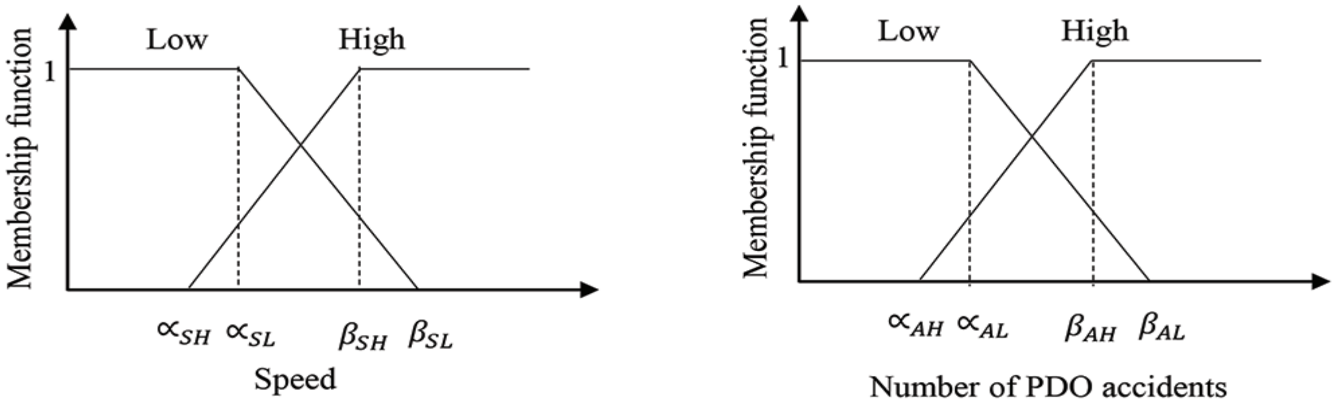


Figure. 2. Membership functions of speed and number of PDO accidents in case 1

base, therefore the end-to-end entire system works on crisp data that may have probabilistic uncertainties. Linguistic rule base makes range information more convenient for experts. The novelty of this study is transforming this issue into a Bayesian parameter approximation case as MFs are evaluated by MCMC technique for available data-set.

The objective of this study is predicting frequency of PDO crashes based on the variables speed, volume of PCs, HVs and LVs; For variables volume of PCs, LVs and HVs with more dispersion, to define sufficient number of rules to cover most real cases, 3 fuzzy memberships (low, medium and high) are considered; while variables speed and frequency of PDO crashes with less dispersion are represented by 2 fuzzy memberships (low and high). In this work the fuzzy rules are known based on the variables considered effective on occurrence of PDO crashes and just determining parameters of MFs is required. 300 data points are produced based on rule foundation aforementioned by Triangular Membership Functions (TMFs) and defuzzification method of centroid. The values of turning points of TMFs are assessed by MCMC method, so evaluating 8 parameters is required. The likelihood is calculated by following equation:

$$P(y_n|g(x;\theta),x_n) = \frac{1}{(2\pi)^{\frac{N}{2}}.\sigma^N} \times \exp\left[-\frac{1}{2\sigma^{2N}}(y_N - g(X_N;\theta))^T(y_N - g(X_N;\theta))\right] \quad (2)$$

where, y_N is a realization of the random variable Y_N .θ={$\emptyset_i$: i=1…n}. is set of n parameters which should be evaluated using Bayesian method and a presumed prior p (θ) which leads to posterior distribution:

$$P(\theta|y_N, X_N) \propto P(y_N|g(x;\theta), X_N)p(\theta) \quad (3)$$

Drawing samples from Eq. (3) through Gibbs sampler is important. There is no additional noise and the measurement error is very small. Prior P (θ) is an independent uniform distribution through the parameter range for each φ_i as:

$$P(\theta) = P(\varphi_1, \varphi_2, \ldots, \varphi_n) = P(\varphi_1)P(\varphi_1)\ldots P(\varphi_n) = \frac{1}{A^n} \quad (4)$$

where, A is highest posterior density estimate for the parameter φ_i.

Initial Markov Chain Monte Carlo simulation is run with 300 data and the overlapping detection for 8 parameters in the case 1 are represented in. The overlapping plots of all parameters in cases 2 to 10 are similar and the Highest Density Intervals (HDIs) are presented in Table 2 for case 1. For all simulation cases, 300 data points as informative priors are applied to all the parameters in Bayesian assessment step.

The left top chart in represents the time-series plots for 3 separate Markov chains Monte Carlo through parameter space for 5000 repetitions. The density charts on the right top are intersecting which implies the right posterior distribution. The ninety-five percent Highest Density Intervals (HDIs) were drawn on x-axis that indicates the likeliest domain of parameter. For parameter, the estimated value is 50, and highest density intervals for 3 chains are between 49.955 and 50.124, which indicate the ability of Markov chain Monte Carlo technique for right assessing the parameter in the tightest range possible with available 300 data points. The overlapping features are

Table 2. Estimation of values of tips and bases of membership functions using Bayesian inference and MCMC for case 1.

		α_{SL}	β_{SL}	α_{SH}	β_{SH}	α_{AL}	β_{AL}	α_{AH}	β_{AH}
Chain 1	HDI_L	49.972	74.945	44.955	69.992	4.998	8.967	4	13.98
	HDI_H	50.113	75.154	45.016	70.009	5.005	9.01	4	14.19
Chain 2	HDI_L	49.969	74.944	44.952	69.981	4.99	8.955	4	13.89
	HDI_H	50.118	75.158	45.02	70.012	5.01	9.018	4.001	14.24
Chain 3	HDI_L	49.955	74.941	44.949	69.983	4.988	8.953	3.998	13.879
	HDI_H	50.124	75.16	45.024	70.01	5.013	9.02	4.004	14.241
Evaluated value		50	75	45	70	5	9	4	14

examined with the auto-correlation charts of 3 chains on the left bottom.

The autocorrelation drops with lags for upcoming data so other techniques such as chain thinning are not needed to reach a reasonable posterior distribution. The Gelman Rubin statistic and shrink factor is usually used to determine the variance of chains relative to the variance within the chains. As is clear from a value of variance is near 1. This shows the convergence of the chains. The Z-scores in overlapping detection charts are the Geweke's diagnostic. These are from normal distribution, therefore most of the values fluctuate in the range [-1,1].

2.2 Aggregation of Diagrams and Defuzzification

After fuzzification of input and output variables and producing rules, combination of diagrams is processed. Composition of diagrams is done according to the rules. After composition and aggregating the diagrams, the next step is defuzzification. In defuzzification, for each mixture of input variables and resulted aggregation diagram, the crisp output, that is the specified number of PDO crashes, is obtained using the method of centriod. For combination of diagrams and defuzzification, MATLAB software, Fuzzy Toolbox is applied.

2.3 Evaluating the Accuracy of Models

For evaluating the efficiency of each model statistical indexes such as goodness of fit and relative error statically Nash–Sutcliffe criterions are used.

$$R^2 = \frac{E_0 - E}{E_0} \times 100 \tag{5}$$

$$E_0 = \sum_i \left(X_{observed} - \bar{X}_{predicted} \right)^2 \tag{6}$$

$$E = \sum_i (X_{observed} - \hat{X}_{predicted})^2 \tag{7}$$

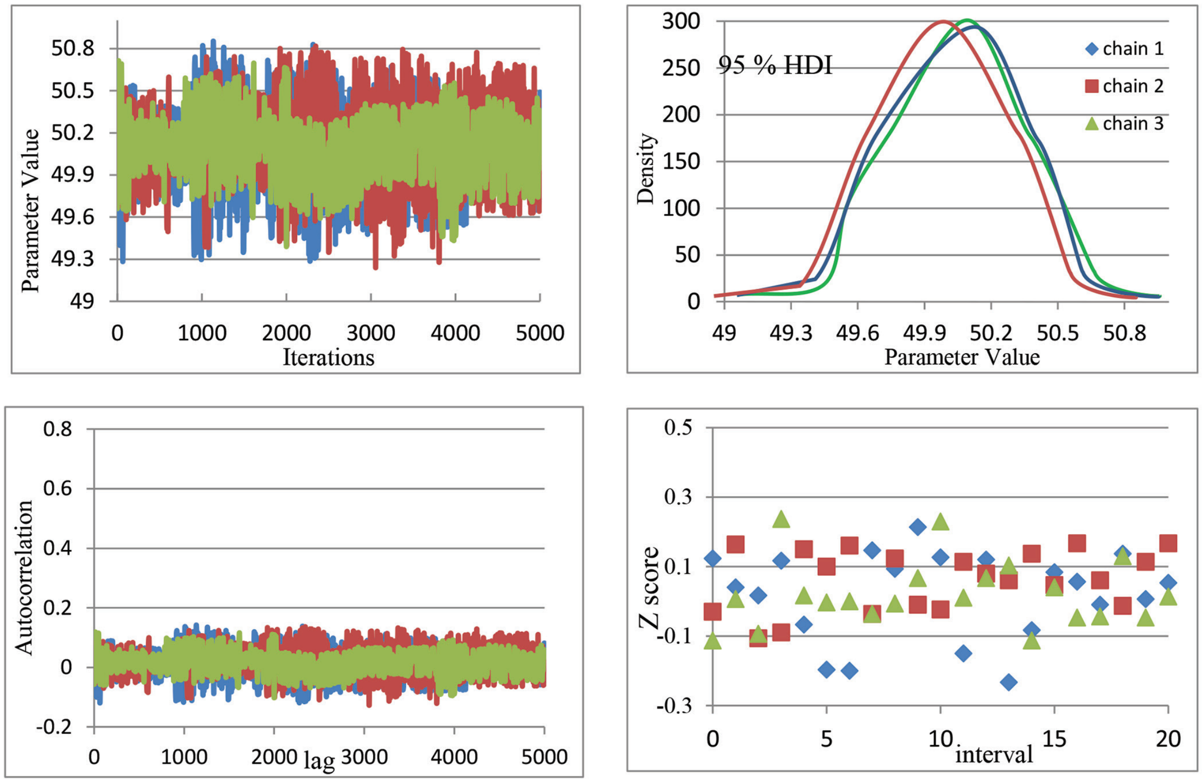
$$\bar{X}_{predicted} = \left(\sum_{i=1}^{n} X_{observed} \right) / n \tag{8}$$

where, $X_{observed}$ = value of observed accident rate, $\bar{X}_{predicted}$ = average of predicted accident rate, $\hat{X}_{predicted}$ = value of predicted accident rate and n is the number of samples.

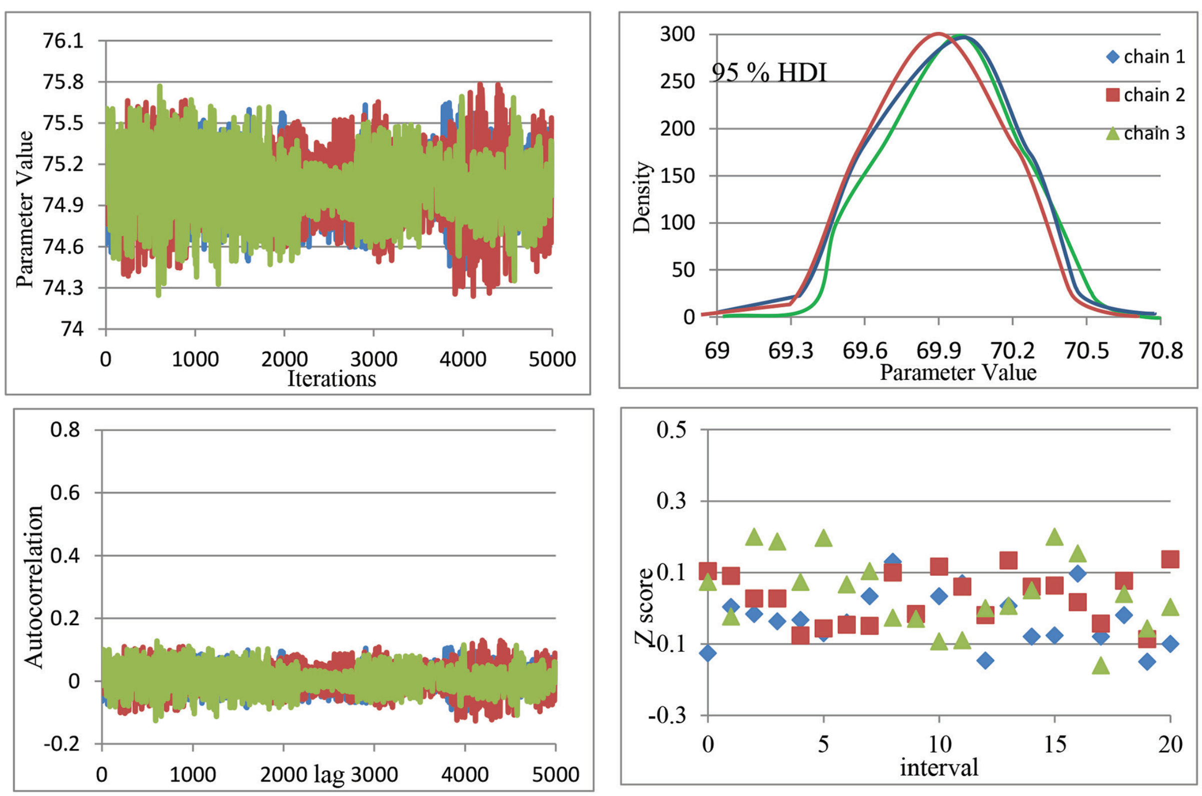
R^2 represents the goodness of fit which varies within the range [0,1]. Moreover, the value of Root Mean Squared Error (RMSE) is applied for estimating the error between predicted and observed frequency of PDO crashes.

$$RMSE = \left[\frac{1}{n} \sum_{i=1}^{n} \left(X_{observed} - X_{predicted} \right) \right]^{\frac{1}{2}} \tag{9}$$

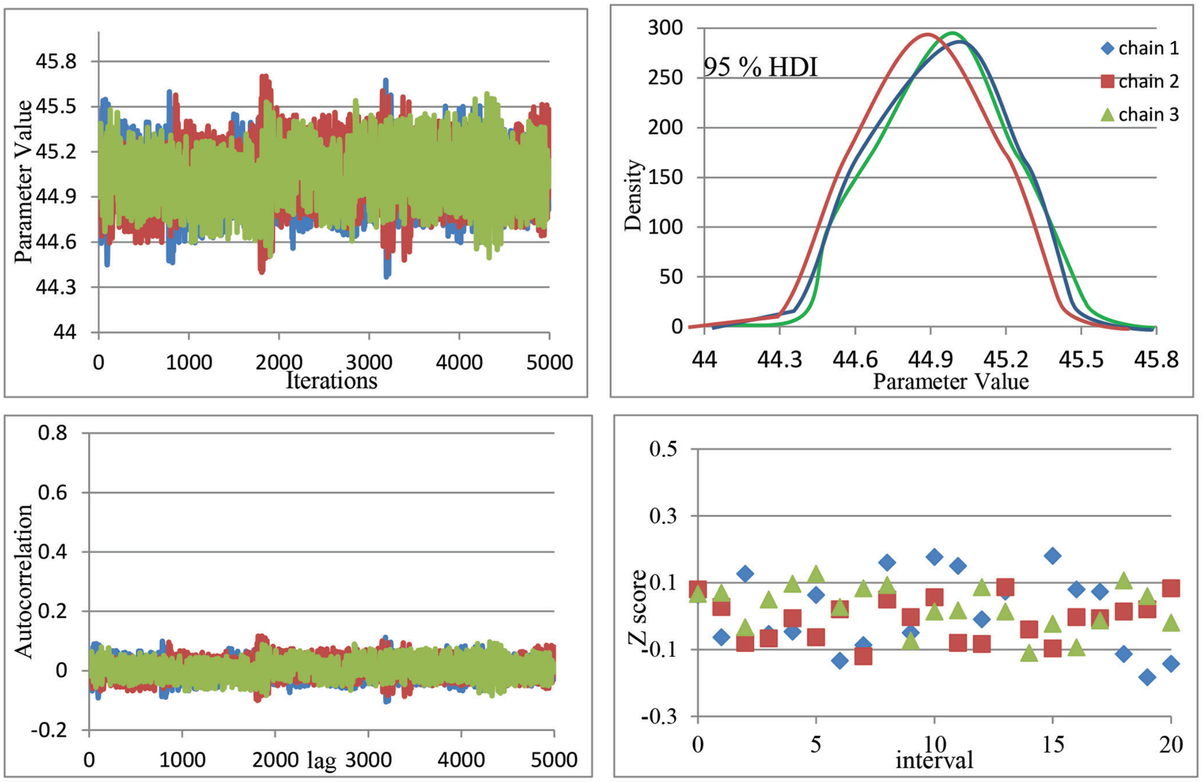
The model with highest value of [2] or lowest value of is the best model and the variables considered effective for that model are the traffic flow related variables, i.e. speed and three predefined categories of traffic volume that do play role in occurrence of PDO crashes. This technique of identification of effective factors on likelihood of traffic crashes is one of the contributions of this article. The contribution goes further when the effect of these factors is ranked and it is specified that which variable is more effective than the other. Furthermore, the researchers want to determine the impact rate of each factor on occurrence of crashes on urban highways. For this job sensitivity analysis is done, as the number of each effective factor is increased by one unit and the best model predict the rate of change in frequency of crashes.


50.8
50.5
50.2
49.9
49.6
49.3
49
Parameter Value
0 1000 2000 3000 4000 5000
Iterations
300
250
200
150
100
50
0
Density
95 % HDI
chain 1
chain 2
chain 3
49 49.3 49.6 49.9 50.2 50.5 50.8
Parameter Value
0.8
0.6
0.4
0.2
0
-0.2
Autocorrelation
0 1000 2000 lag 3000 4000 5000
0.5
0.3
0.1
-0.1
-0.3
Z score
0 5 10 interval 15 20

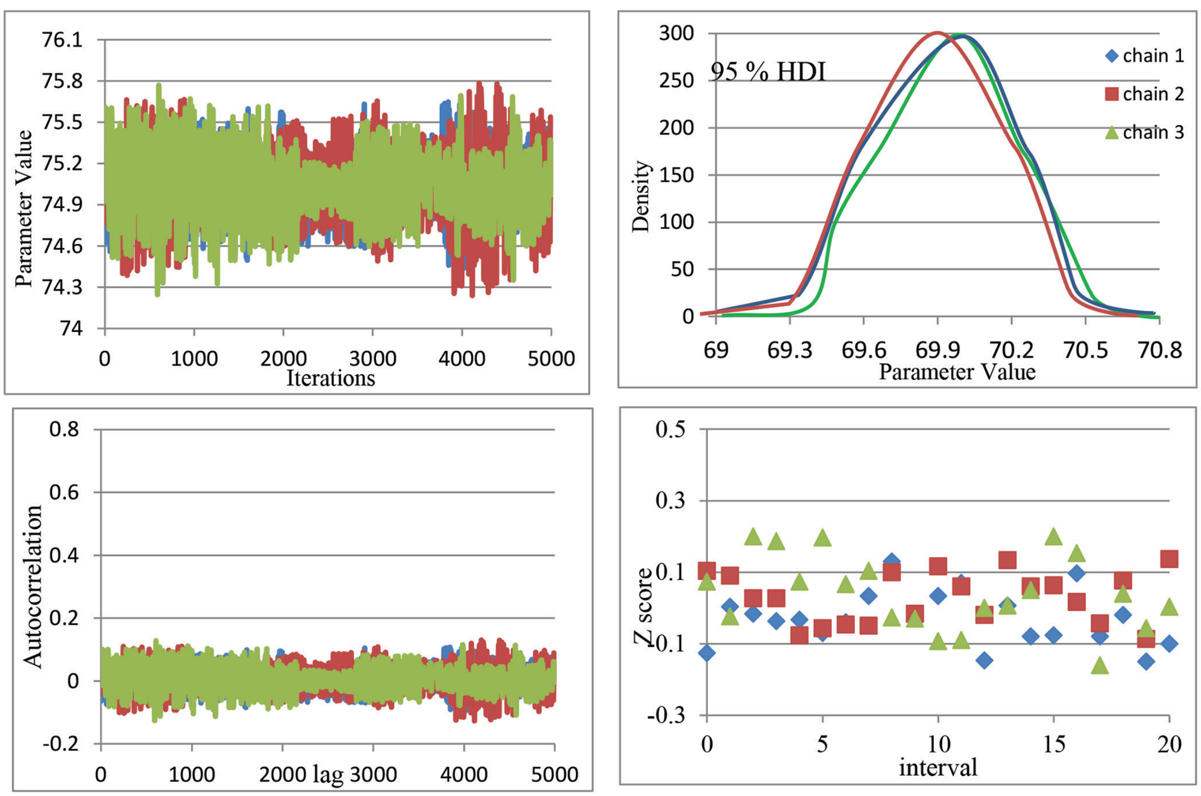
(a)


76.1
75.8
75.5
75.2
74.9
74.6
74.3
74
Parameter Value
0 1000 2000 3000 4000 5000
Iterations
300
250
200
150
100
50
0
Density
95 % HDI
chain 1
chain 2
chain 3
69 69.3 69.6 69.9 70.2 70.5 70.8
Parameter Value
0.8
0.6
0.4
0.2
0
-0.2
Autocorrelation
0 1000 2000 lag 3000 4000 5000
0.5
0.3
0.1
-0.1
-0.3
Z score
0 5 10 15 20
interval

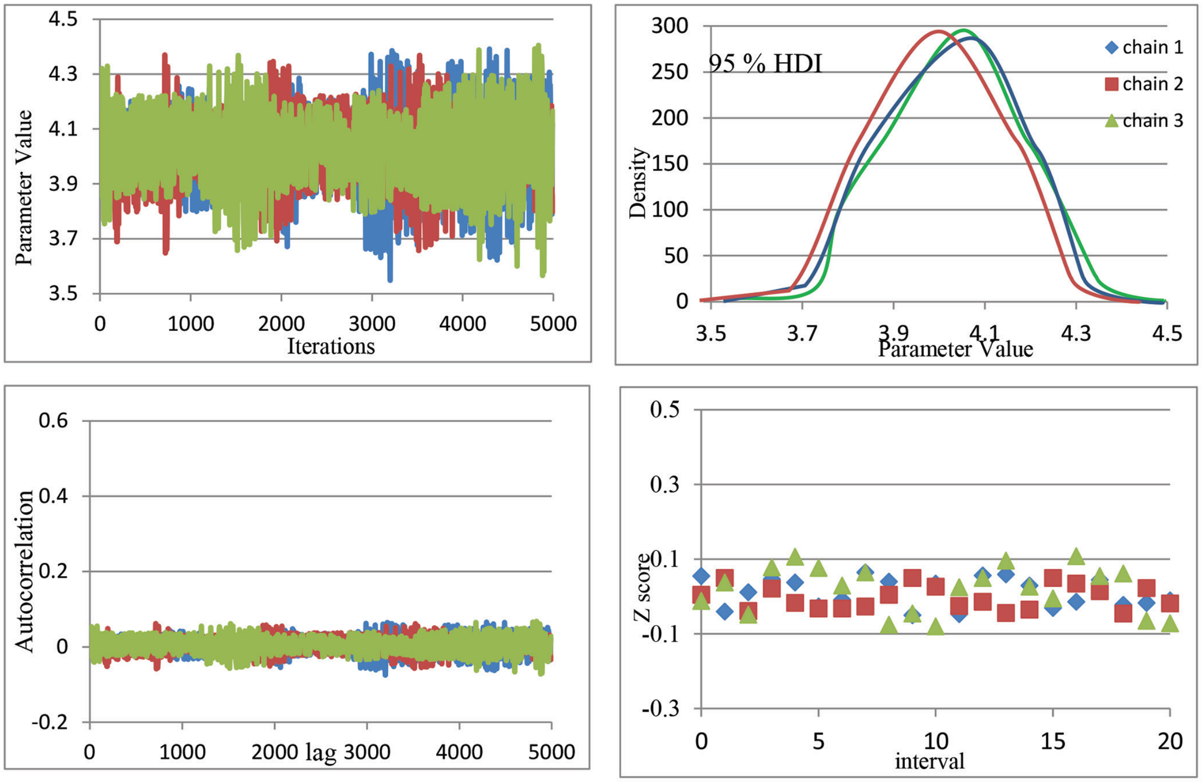
(b)


Parameter Value
Iterations
95 % HDI
chain 1
chain 2
chain 3
Density
Parameter Value
Autocorrelation
lag
Z score
interval

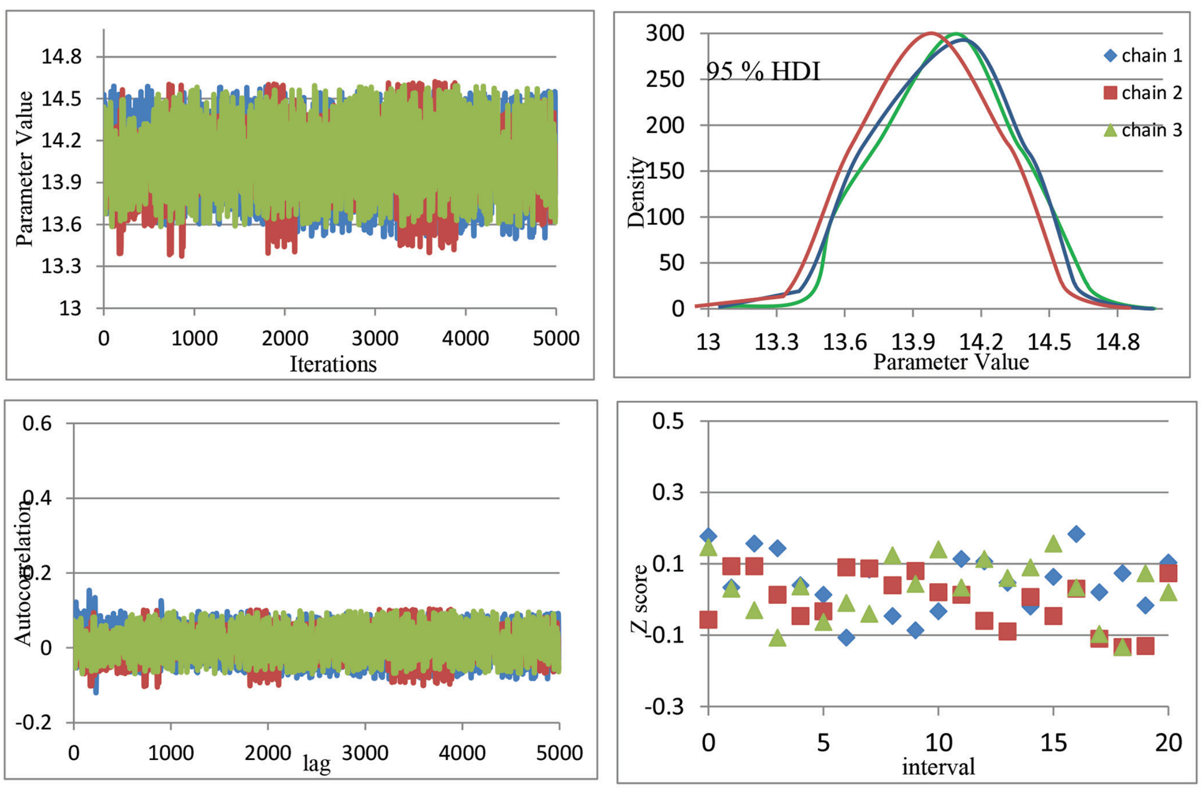
(c)


Parameter Value
Iterations
95 % HDI
chain 1
chain 2
chain 3
Density
Parameter Value
Autocorrelation
lag
Z score
interval

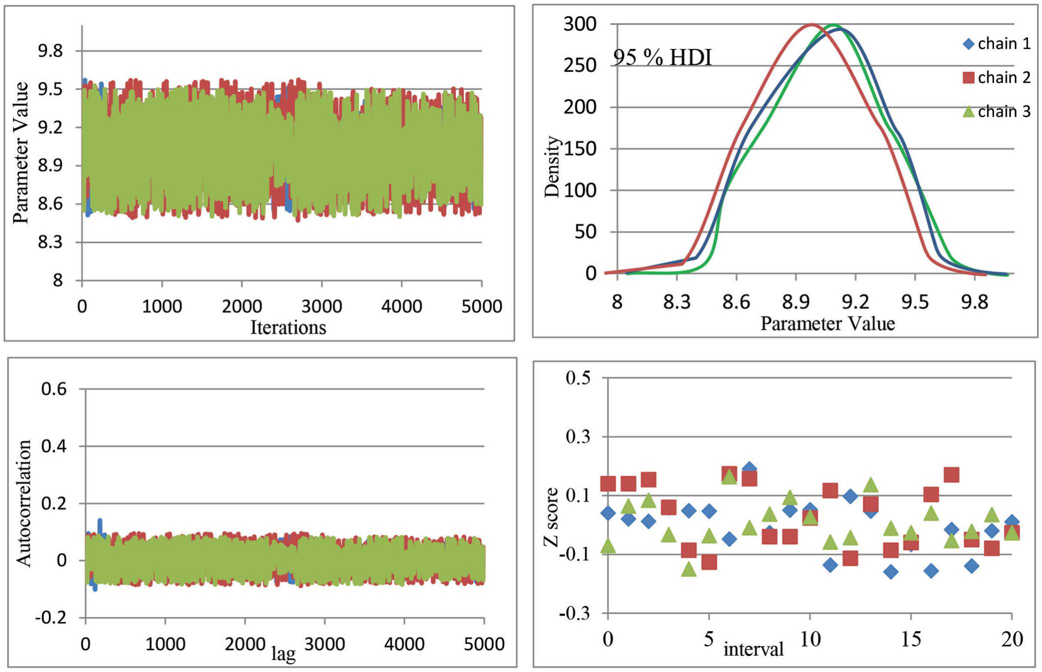
(d)


Parameter Value
Iterations
95 % HDI
Density
Parameter Value
chain 1
chain 2
chain 3
Autocorrelation
lag
Z score
interval

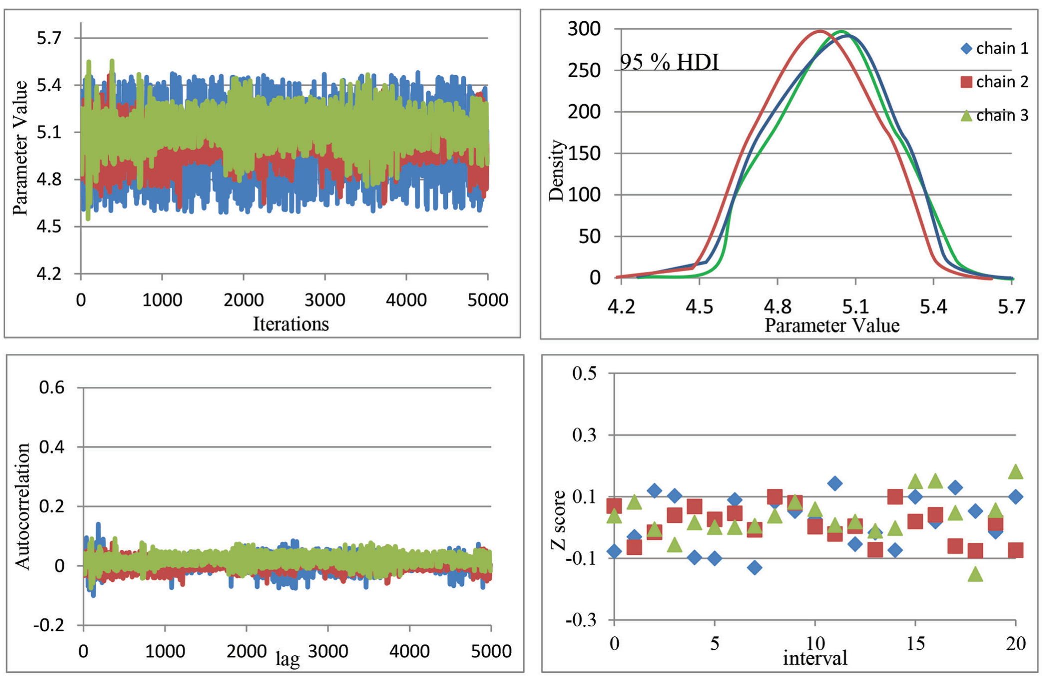
(e)


Parameter Value
Iterations
95 % HDI
Density
Parameter Value
chain 1
chain 2
chain 3
Autocorrelation
lag
Z score
interval

(f)



(g)



(h)

Figure 3. Bayesian estimation of tips and bases of fuzzy membership function for case 1.
(a) Parameter α_{SI}.
(b) Parameter β_{SI}.
(c) Parameter α_{SH}.
(d) Parameter β_{SH}.
(e) Parameter α_{AH}.
(f) Parameter β_{AH}.
(g) Parameter β_{AL}.
(h) Parameter α_{AL}.

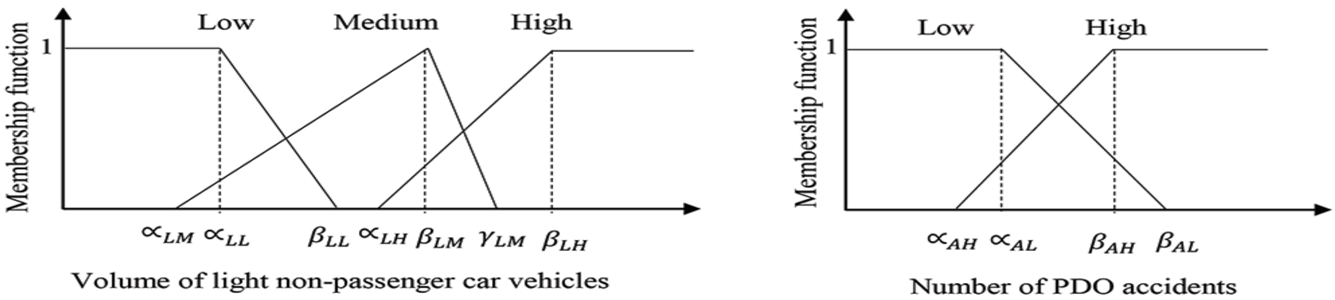


Figure. 4. Membership functions of volume of light non-passenger car vehicles and number of PDO accidents in case 2

3. Results

3.1 Parameter Estimation in Fuzzification

For the first case, speed is considered to be the only efficient factor on occurrence of PDO crashes on urban highways. The results of Bayesian estimation of turning points of fuzzy MFs for the first case are tabulated in Table 2.

In the second case it is assumed that volume of LVs is the only efficient factor on occurrence of PDO crashes on urban highways. Therefore, the rule base applied to develop the FIS is as:

1. If volume of LVs is low, then frequency of PDO crashes is low
2. If volume of LVs is medium, then frequency of PDO crashes is low (or high)
3. If volume of LVs is high, then frequency of PDO crashes is high

The MFs of volume of LVs and PDO crashes are as in.

Results of Bayesian estimation of the turning points of fuzzy MFs for case 2 are in Table 3.

For cases 3 and 4 the rule foundation applied to develop the FIS is as case 2 for volume of PCs and HVs respectively. The MFs of volume of PCs, HVs and PDO crashes are as in.

The estimated values of turning points of TMFs for cases 3 and 4 are presented in Table 4.

Developed membership functions for cases 1 to 4 are represented in figure 6

Table 3. Estimation of the turning points of MFs using Bayesian inference and MCMC for case 2

Case 2		α_{LL}	β_{LL}	α_{LM}	β_{LM}	γ_{LM}	α_{LH}	β_{LH}
Chain 1	HDI_L	207.72	404.18	236.91	395.12	782.54	535.6	1220.1
	HDI_H	214.04	417.09	244.54	406.21	799.67	545.78	1238.9
Chain 2	HDI_L	206.14	403.09	236.12	394.25	781.04	534.1	1219.6
	HDI_H	214.67	418.34	243.87	407.56	802.45	546.66	1239.8
Chain 3	HDI_L	205.88	401.13	235.24	394.15	779.21	532.92	1217.8
	HDI_H	215.08	419.10	244.43	407.98	804.02	547.08	1242.1
Evaluated value		210	410	240	400	790	540	1230
Case 2		α_{AL}	β_{AL}	α_{AH}	β_{AH}			
Chain 1	HDI_L	6.989	10.91	5.976	14.899			
	HDI_H	7.013	11.149	6.005	15.186			
Chain 2	HDI_L	6.978	10.883	5.978	14.876			
	HDI_H	7.019	11.195	6.002	15.198			
Chain 3	HDI_L	6.966	10.801	5.945	14.864			
	HDI_H	7.03	11.205	6.032	15.225			
Evaluated value		7	11	6	15			

Table 4. Estimation of tips and bases of membership functions using Bayesian inference and MCMC for cases 3 and 4.

Case 3	α_{PL}	β_{PL}	α_{PM}	β_{PM}	γ_{PM}	α_{PH}	β_{PH}
Evaluated value	350	520	290	670	1230	950	1820
Case 3	α_{AL}	β_{AL}	α_{AH}	β_{AH}			
Evaluated value	6	9	8	14			
Case 4	α_{HL}	β_{HL}	α_{HM}	β_{HM}	γ_{HM}	α_{HH}	β_{HH}
Evaluated value	270	410	250	570	1020	590	905
Case 4	α_{AL}	β_{AL}	α_{AH}	β_{AH}			
Evaluated value	7	10	5	17			

diagram, the crisp output, that is the specified frequency of PDO crashes, is obtained. For combination of diagrams and defuzzification, MATLAB software, Fuzzy Toolbox is applied. Accordingly, for each input data set, one specified frequency of PDO crashes is estimated by fuzzy Bayesian inference based models. This process is implemented for cases 1 to 4. The results of comparison between the frequencies of predicted and observed crashes for cases 1 to 4 is as Table 5. The corresponding diagrams are represented.

3.3 Process Continued for Cases 5 to 7

Based on the results obtained up to now, for the next case (5) the variable volume of LVs is added to speed, then corresponding rules, for their direct effect on the

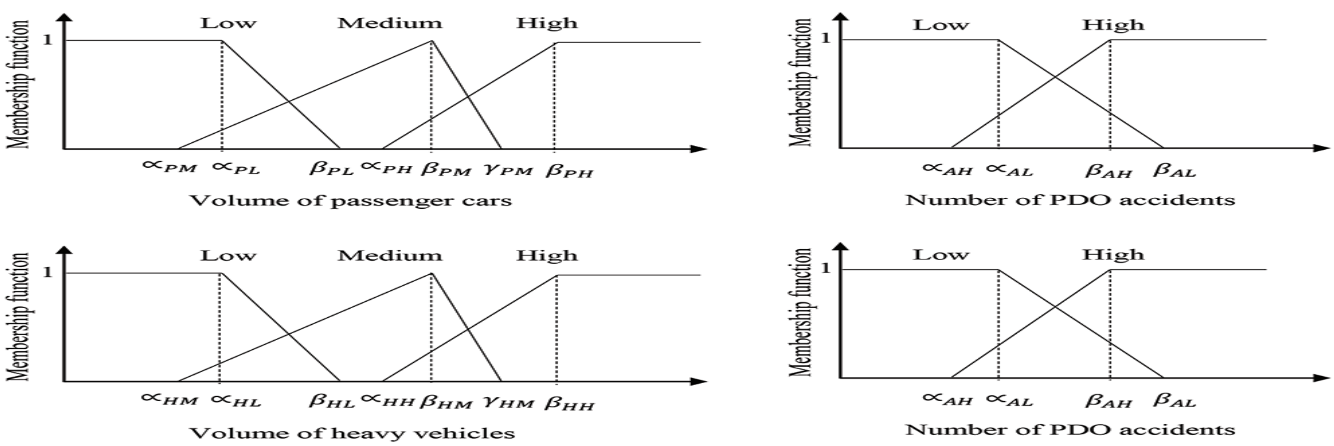


Figure. 5. Membership functions of volume of passenger car (case 3), heavy vehicles (case 4) and number of PDO accidents

Table 5. Results of comparison between the predicted and observed number of PDO accidents for cases 1 to 4

Cases	R^2	*RMSE*
1	0.56	0.029
2	0.52	0.032
3	0.48	0.0342
4	0.41	0.0384

3.2 Aggregation and Defuzzification

After fuzzification of input and output variables and producing rules, combination of diagrams is processed. Composition of diagrams is done according to the rules. After composition and aggregating the diagrams, the next step is defuzzification. In defuzzification, for each mixture of input variables and resulted aggregation

frequency of PDO crashes are produced and as previous fuzzification of each parameter is done using Bayesian inference by MCMC algorithms and at last defuzzification gives the crisp predicted frequency of crashes for each input data set.

For case 5 the rule foundation applied to develop the FIS is as following:

1. If average speed is low and volume of LVs is low; then frequency of PDO crashes is low
2. If average speed is low and volume of LVs is medium; then frequency of PDO crashes is low
3. If average speed is low and volume of LVs is high; then frequency of PDO crashes is low (or high)
4. If average speed is high and volume of LVs is low; then frequency of PDO crashes is high (or low)

Table 6. Estimation of tips and bases of membership functions using Bayesian inference and MCMC for case 5.

Case 5	α_{LL}	β_{LL}	α_{LM}	β_{LM}	γ_{LM}	α_{LH}	β_{LH}	
Evaluated value	150	370	310	630	870	850	1220	
Case 5	α_{SL}	β_{SL}	α_{SH}	β_{SH}	α_{AL}	β_{AL}	α_{AH}	β_{AH}
Evaluated value	40	80	45	70	6	11	7	15

Table 7. Estimation of the tips and bases of membership functions using Bayesian inference and MCMC for cases 6 and 7.

Case 6	α_{PL}	β_{PL}	α_{PM}	β_{PM}	γ_{PM}	α_{PH}	β_{PH}	
Evaluated value	320	570	350	750	1320	1050	1780	
Case 6	α_{SL}	β_{SL}	α_{SH}	β_{SH}	α_{AL}	β_{AL}	α_{AH}	β_{AH}
Evaluated value	55	78	47	76	5	10	6	15
Case 7	α_{HL}	β_{HL}	α_{HM}	β_{HM}	γ_{HM}	α_{HH}	β_{HH}	
Evaluated value	270	440	240	550	1020	630	1105	
Case 7	α_{SL}	β_{SL}	α_{SH}	β_{SH}	α_{AL}	β_{AL}	α_{AH}	β_{AH}
Evaluated value	49	68	53	82	4	10	6	14

Table 8. Results of comparison between predicted and observed number of PDO accidents for cases 5 to 7

Cases	R^2	*RMSE*
5	0.58	0.0282
6	0.53	0.031
7	0.5	0.033

5. If average speed is high and volume of LVs is medium; then frequency of PDO crashes is high
6. If average speed is high and volume of LVs is high; then frequency of PDO crashes is high

The MFs of volume of speed, LVs and PDO crashes are as in.

The estimated values of turning points of TMFs in case 5 are presented in Table 6.

For cases 6 and 7 the rule foundation applied for developing the FIS is as case 5 for volume of PCs and HVs. The estimated values of turning points of TMFs in cases 6 and 7 are tabulated in Table 7.

Developed MFs for cases 5 to 7 are represented in.

3.4 Aggregation and Defuzzification for cases 5 to 7

As previously stated, after fuzzification of input and output variables and producing rules, combination of diagrams and defuzzification is processed. In defuzzification, for each set of input variables and developed aggregation diagram a crisp output is obtained in MATLAB software, Fuzzy Toolbox. This process is implemented for cases 5 to 7 similar to what is done before for cases 1 to 4. The results of comparison between the numbers of predicted and observed crashes for cases 5 to 7 is as Table 8.

3.5 Process Continued for Cases 8 and 9

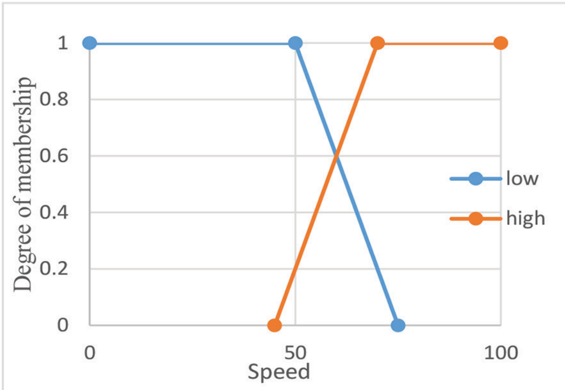
Based on the results obtained, for the next case (8) the variable volume of PCs is added to speed and volume of LVs, then corresponding rules, for their direct effect on the frequency of PDO crashes are produced and as previous fuzzification of each parameter is done using Bayesian inference by

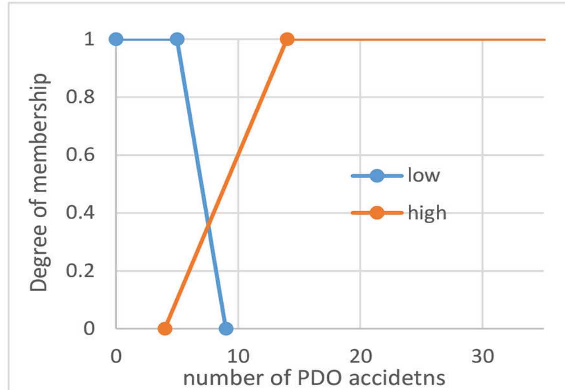
MCMC algorithms. Then defuzzification gives the crisp predicted frequency of crashes for each input data set.

For case 8 the rule foundation applied to develop the FIS is as following:

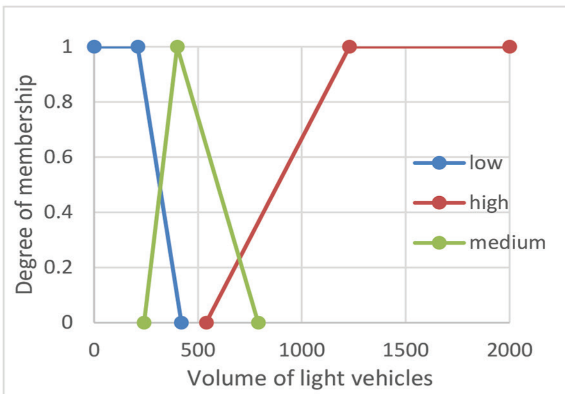
1. If speed is low, volume of LVs is low and volume of PCs is low, then frequency of PDO crashes is low.
2. If speed is low, volume of LVs is low and volume of PCs is medium, then frequency of PDO crashes is low.

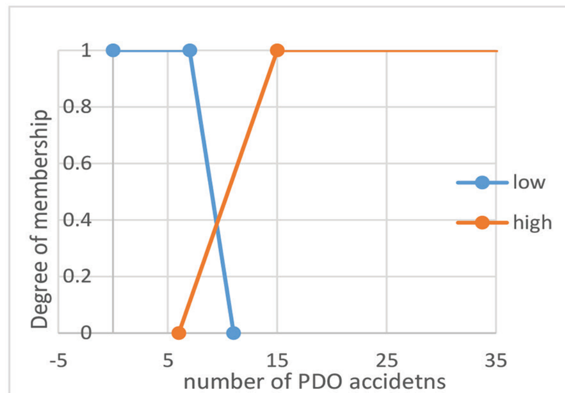
Case 1



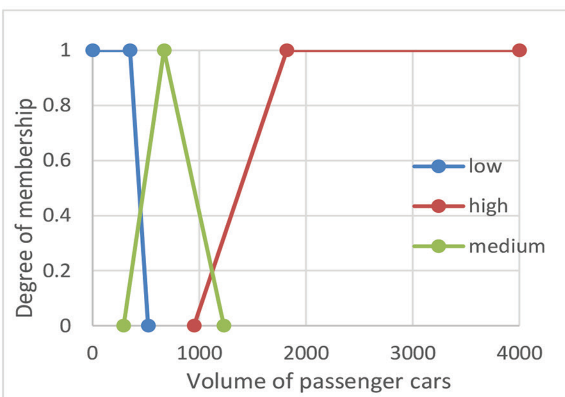


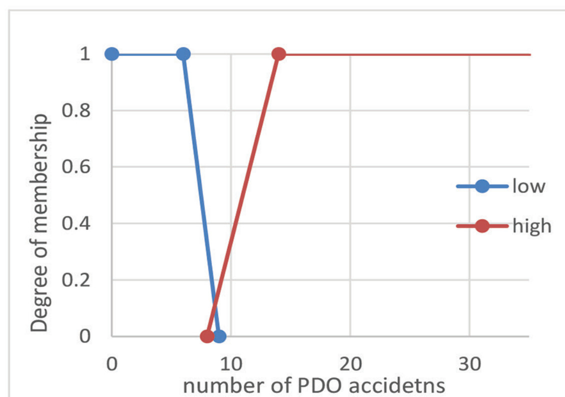
Case 2



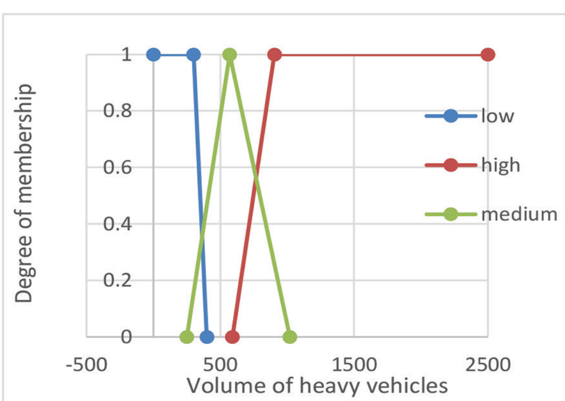


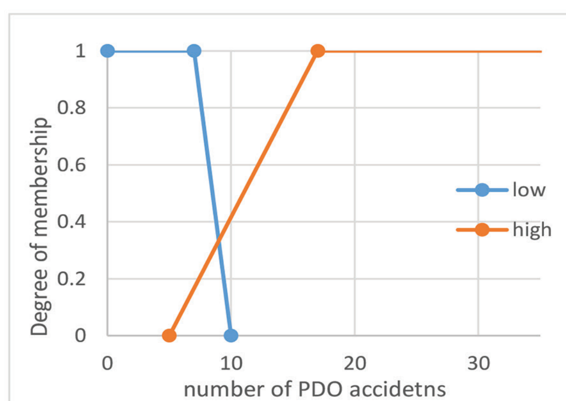
Case 3



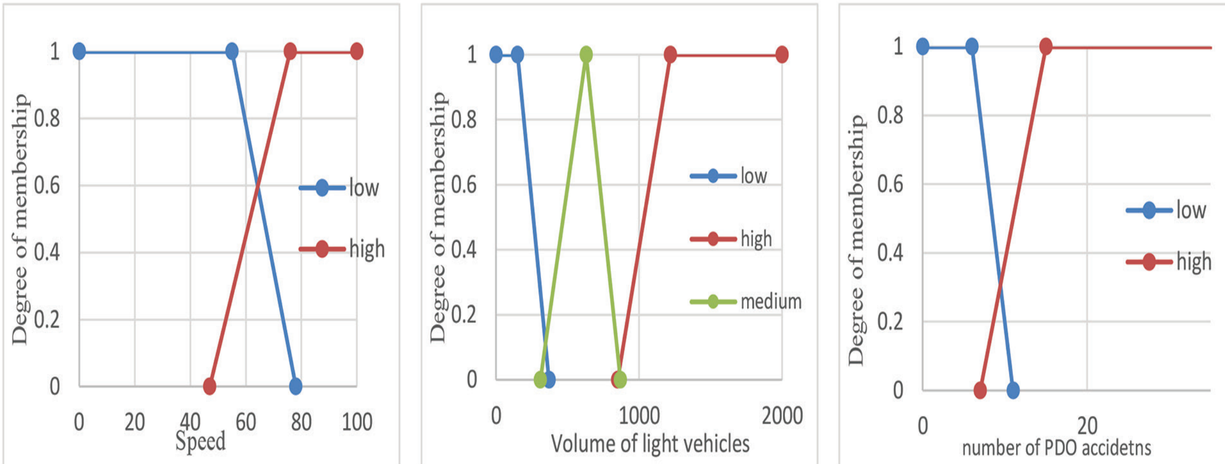


Case 4

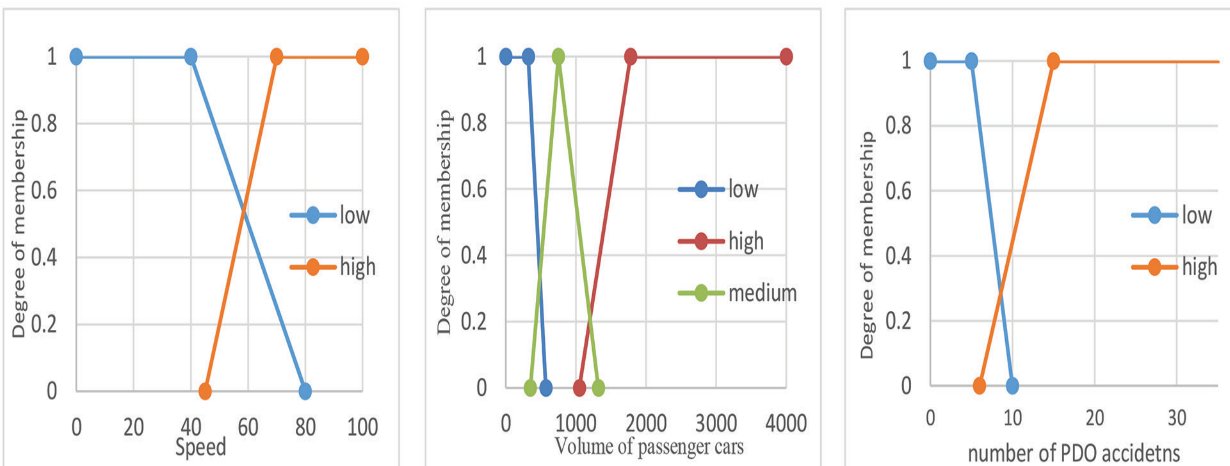




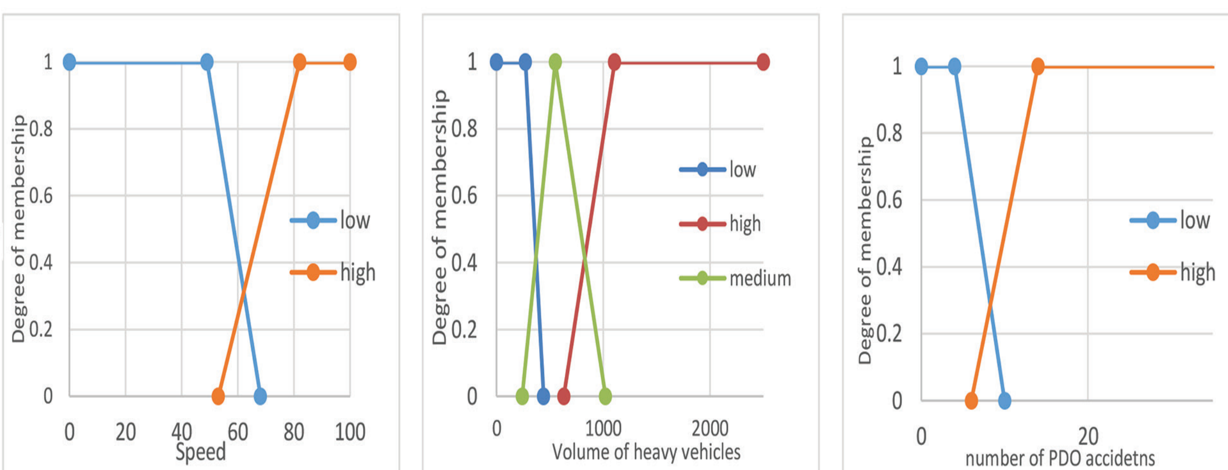
Case 5



Case 6



Case 7



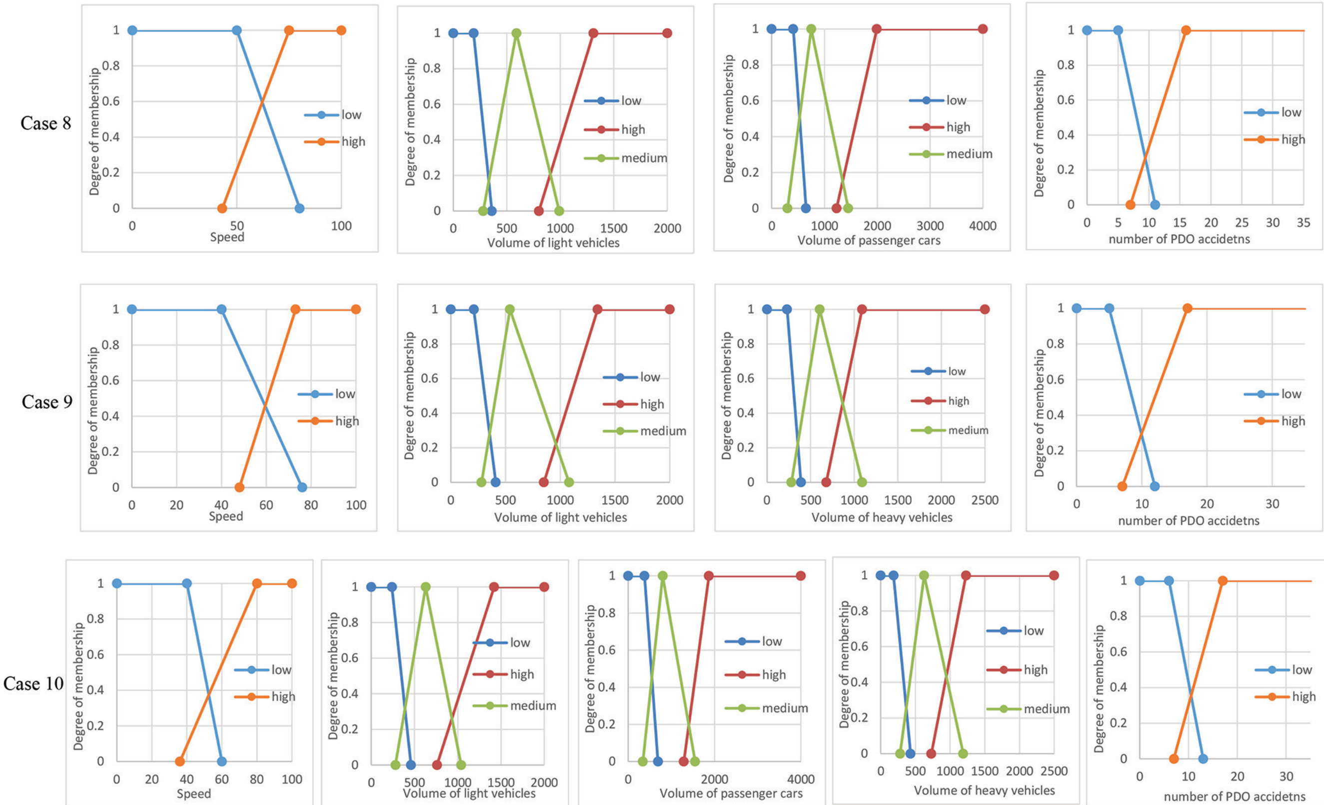


Figure. 6. Developed membership functions for cases 1-10.

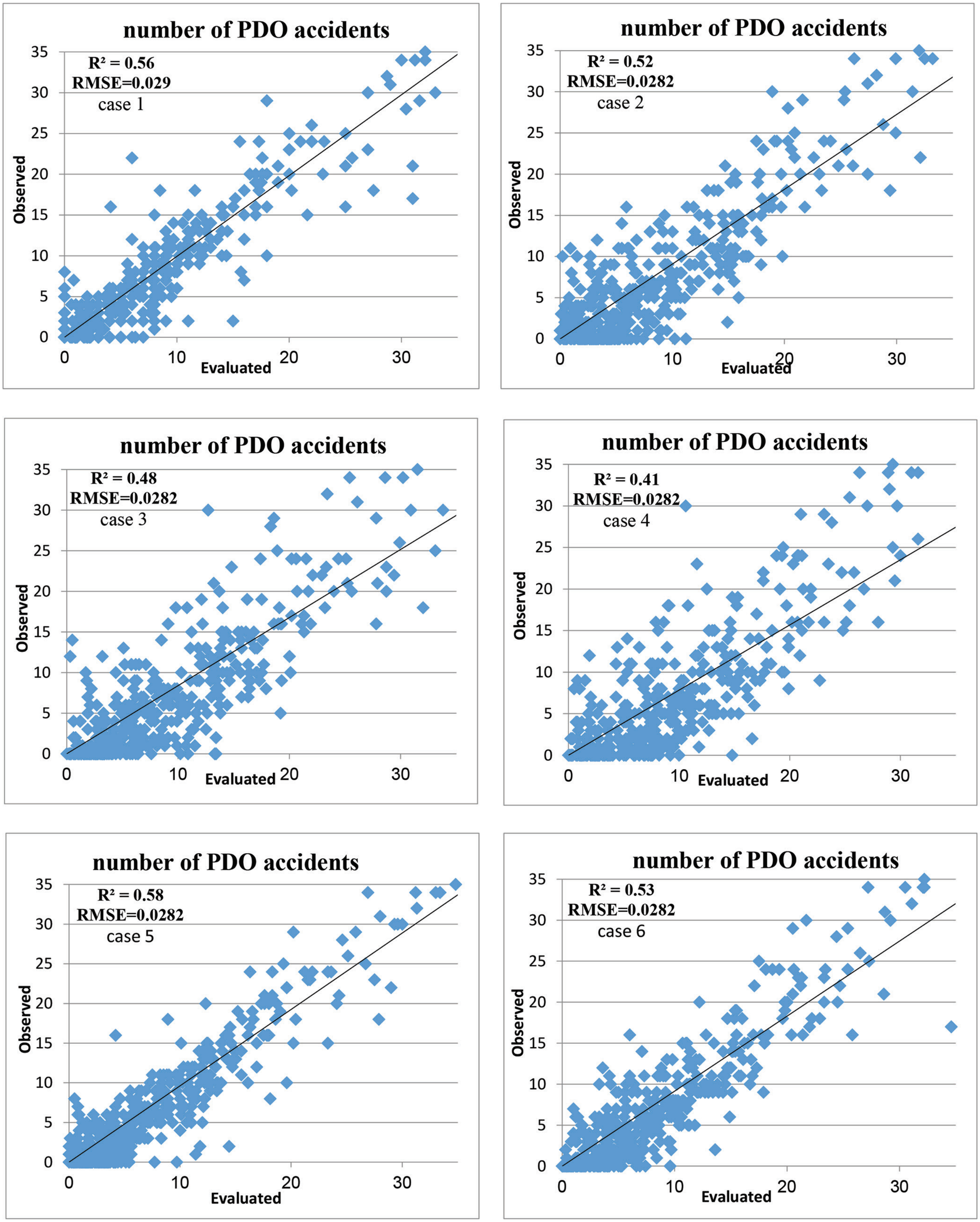
Table 9. Estimation of tips and bases of membership functions using Bayesian inference and MCMC for case 8.

Case 8	α_{LL}	β_{LL}	α_{LM}	β_{LM}	γ_{LM}	α_{LH}	β_{LH}	
Evaluated value	190	360	280	590	990	800	1310	
Case 8	α_{PL}	β_{PL}	α_{PM}	β_{PM}	γ_{PM}	α_{PH}	β_{PH}	
Evaluated value	410	650	300	750	1450	1230	1990	
Case 8	α_{SL}	β_{SL}	α_{SH}	β_{SH}	α_{AL}	β_{AL}	α_{AH}	β_{AH}
Evaluated value	50	80	43	75	5	11	7	16

Table 10. Estimation of tips and bases of membership functions using Bayesian inference and MCMC for case 9.

Case 9	α_{LL}	β_{LL}	α_{LM}	β_{LM}	γ_{LM}	α_{LH}	β_{LH}	
Evaluated value	210	410	280	540	1080	850	1340	
Case 9	α_{HL}	β_{HL}	α_{HM}	β_{HM}	γ_{HM}	α_{HH}	β_{HH}	
Evaluated value	230	390	280	605	1090	680	1090	
Case 9	α_{SL}	β_{SL}	α_{SH}	β_{SH}	α_{AL}	β_{AL}	α_{AH}	β_{AH}
Evaluated value	40	76	48	73	5	12	7	17

3. If speed is low, volume of LVs is low and volume of PCs is high then frequency of PDO crashes is low.
4. If speed is low, volume of LVs is medium and volume of PCs is low, then frequency of PDO crashes is low.


number of PDO accidents
R² = 0.56
RMSE=0.029
case 1
Observed
Evaluated
number of PDO accidents
R² = 0.52
RMSE=0.0282
case 2
Observed
Evaluated
number of PDO accidents
R² = 0.48
RMSE=0.0282
case 3
Observed
Evaluated
number of PDO accidents
R² = 0.41
RMSE=0.0282
case 4
Observed
Evaluated
number of PDO accidents
R² = 0.58
RMSE=0.0282
case 5
Observed
Evaluated
number of PDO accidents
R² = 0.53
RMSE=0.0282
case 6
Observed
Evaluated

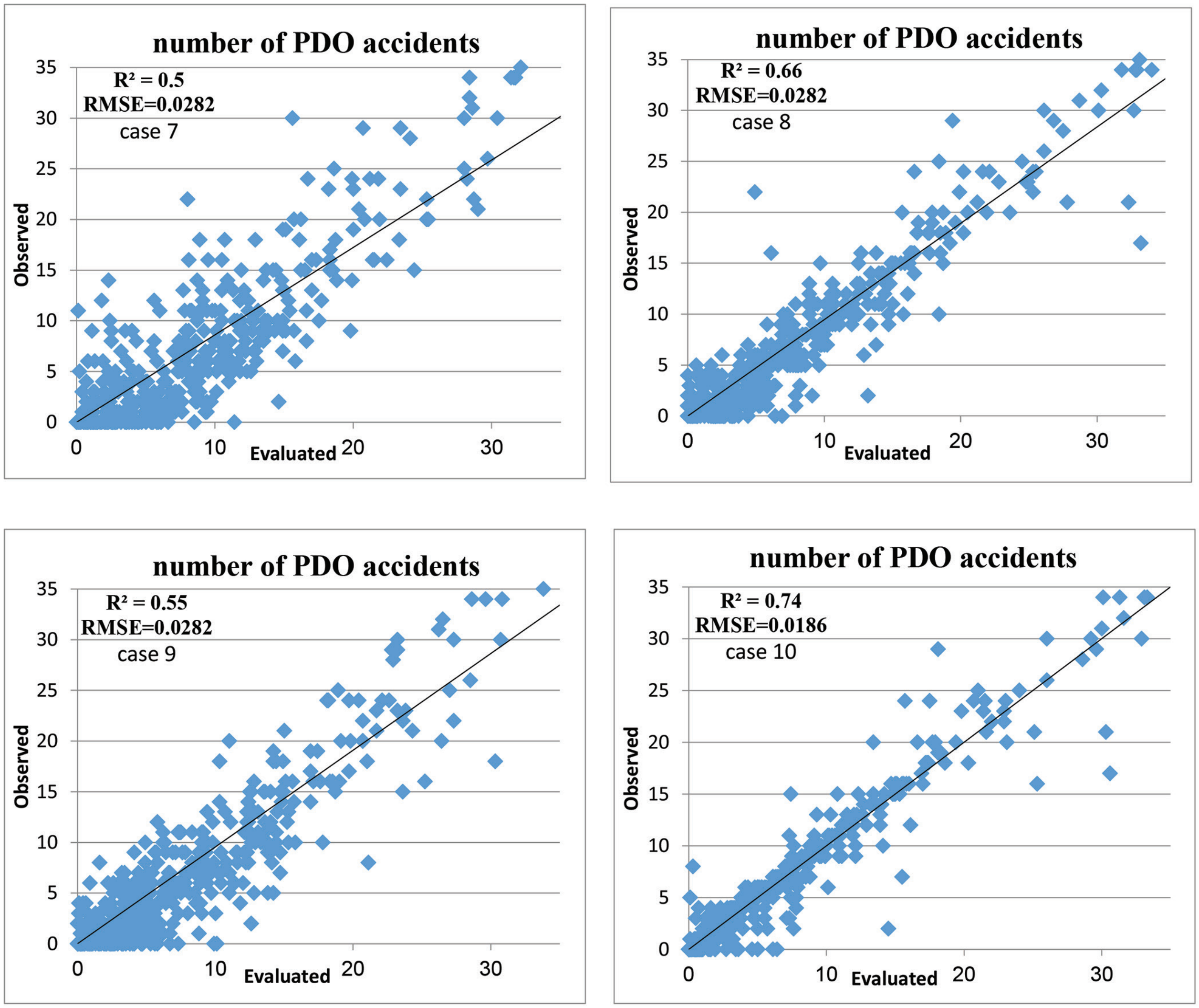


Figure 7. Evaluating the accuracy of models based on comparison between the numbers of predicted and observed accidents for cases 1 to 10

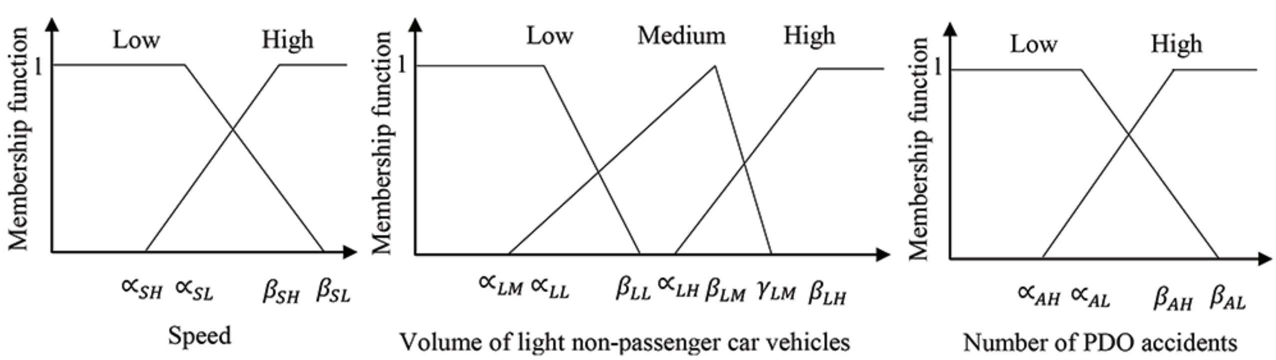


Figure. 8. Membership functions of speed, volume of light non-passenger car vehicles and number of PDO accidents for case 5

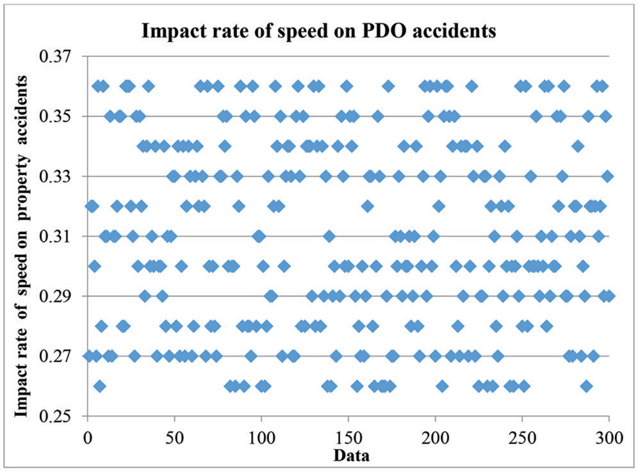


Figure. 9. Impact rate of speed on likelihood of accidents on urban highways

5. If speed is low, volume of LVs is medium and volume of PCs is medium, then frequency of PDO crashes is low.
6. If speed is low, volume of LVs is medium and volume of PCs is high, then frequency of PDO crashes is low (or high).
7. If speed is high and volume of LVs is high and volume of PCs is low, then frequency of PDO crashes is high.
8. If speed is high and volume of LVs is high and volume of PCs is medium, then frequency of PDO crashes is high.

Table 11. Results of comparison between predicted and observed number of PDO accidents for cases 8 and 9

Cases	R^2	*RMSE*
8	0.66	0.0234
9	0.55	0.03

9. If speed is high and volume of LVs is high and volume of PCs is high, then frequency of PDO crashes is high.

The estimated values of turning points of TMFs in case 8 is tabulated in Table 9.

For case 9 the rule foundation applied for developing the FIS is as case 8 for volume of HVs instead of PCs. The estimated values of turning points of TMFs in case 9 is tabulated in Table 10.

Developed MFs for cases 8 and 9 are represented in.

3.6 Aggregation and Defuzzification for Cases 8 and 9

As previous, after fuzzification of input and output variables, producing rules and combination of diagrams, in defuzzification, for each set of input variables and developed aggregation diagram a crisp output is obtained using the method of centroid in MATLAB software, Fuzzy Toolbox for cases 8 and 9 similar to what is done for previous cases. The results of comparison between the numbers of predicted and observed crashes for cases 8 and 9 is tabulated in Table 11. The corresponding diagrams are represented in.

3.7 Process Continued for Case 10

Based on recent results, for the last case the variable volume of HVs is added to speed, volume of LVs and PCs, corresponding rules are produced and fuzzification of each parameter is done using Bayesian inference by MCMC algorithms. After combination of diagrams, defuzzification gives the crisp predicted frequency of crashes for each data set.

For case 10 the rule foundation applied to develop the FIS is as following:

Table 12. Estimation of tips and bases of membership functions using Bayesian inference and MCMC for case 10

Case 10	α_{LL}	β_{LL}	α_{LM}	β_{LM}	γ_{LM}	α_{LH}	β_{LH}	
Evaluated value	240	460	280	630	1040	760	1420	
Case 10	α_{PL}	β_{PL}	α_{PM}	β_{PM}	γ_{PM}	α_{PH}	β_{PH}	
Evaluated value	380	690	340	800	1550	1290	1870	
Case 10	α_{HL}	β_{HL}	α_{HM}	β_{HM}	γ_{HM}	α_{HH}	β_{HH}	
Evaluated value	185	430	280	625	1190	730	1230	
Case 10	α_{SL}	β_{SL}	α_{SH}	β_{SH}	α_{AL}	β_{AL}	α_{AH}	β_{AH}
Evaluated value	40	60	36	80	6	13	7	17

Table 13. Results of comparison between the numbers of predicted and observed PDO accidents for cases

Case	Effective factors (considered)	Number of rules	R^2	*RMSE*
1	speed	2	0.56	0.029
2	Volume of LVs	3	0.52	0.032
3	Volume of PCs	3	0.48	0.0342
4	Volume of HVs	3	0.41	0.0384
5	1. speed	6	0.58	0.0282
6	1. speed	6	0.53	0.031
7	1. speed	6	0.5	0.033
	2. Volume of HVs			
8	1. speed	18	0.66	0.0234
	2. Volume of LVs			
	3. Volume of PCs			
9	1. speed	18	0.55	0.03
	2. Volume of LVs			
	3. Volume of HVs			
10	1. speed	54	0.74	0.0186
	2. Volume of LVs			
	3. Volume of HVs			
	4. Volume of PCs			

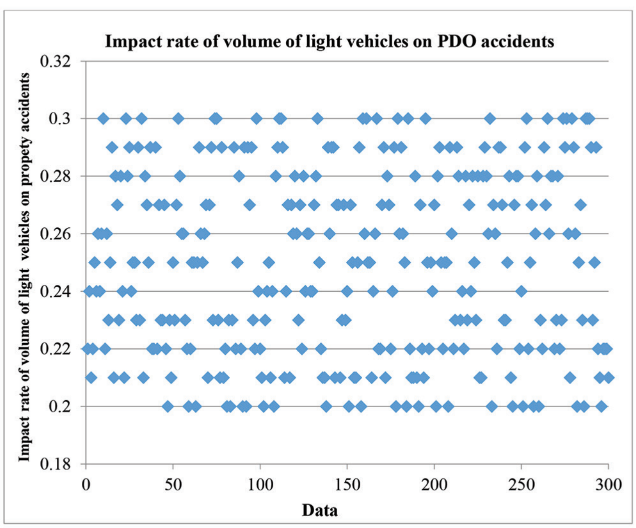


Figure. 10. Impact rate of volume of light non passenger car vehicles on likelihood of PDO accidents on urban highways

If speed is low, volume of LVs is low, volume of PCs is low and volume of HVs is low then frequency of PDO crashes is low.

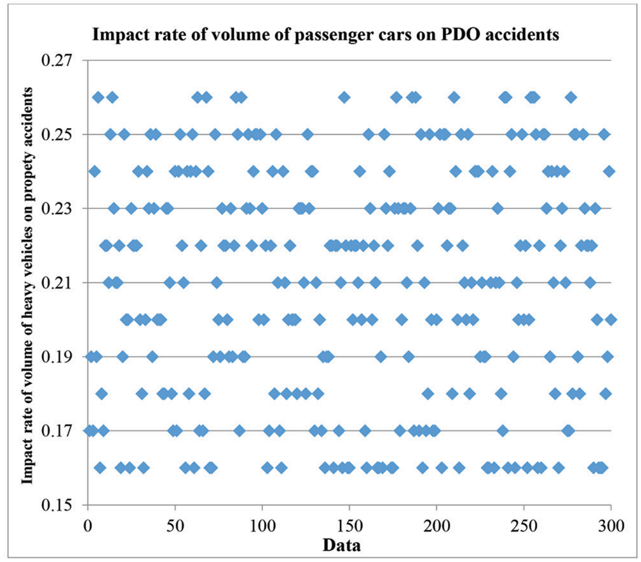


Figure. 11. Impact rate of volume of passenger cars on likelihood of PDO accidents on urban highways

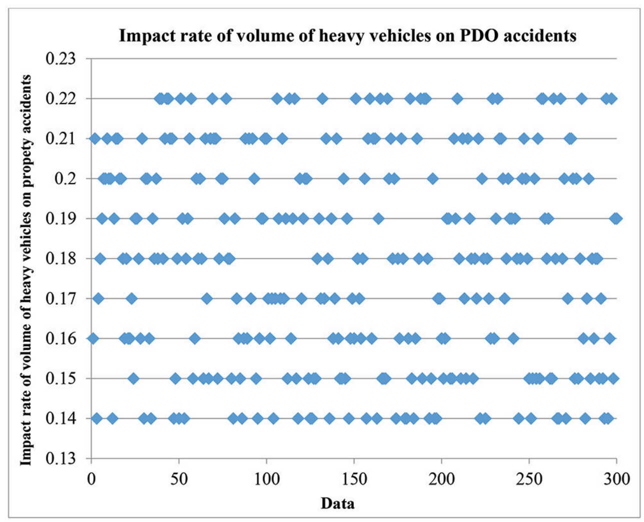


Figure. 12. Impact rate of volume of heavy vehicles on likelihood of PDO accidents on urban highways

If speed is low, volume of LVs is medium, volume of PCs is low and volume of HVs is low then frequency of PDO crashes is low.

If speed is low, volume of LVs is medium, volume of PCs is high and volume of HVs is high then frequency of PDO crashes is medium.

If speed is low, volume of LVs is high, volume of PCs is low and volume of HVs is low then frequency of PDO crashes is low.

Table 14. Impact rate of input variables on occurrence of PDO accidents on urban highways.

Impact rate of variables on occurrence of PDO crashes	Speed	Volume of light non passenger car vehicles	Volume of heavy vehicles	Volume of passenger cars
	0.3	0.24	0.2	0.22

If speed is low, volume of LVs is high, volume of PCs is high and volume of HVs is high then frequency of PDO crashes is high.

If speed is high, volume of LVs is low, volume of PCs is low and volume of HVs is low then frequency of PDO crashes is low.

If speed is high, volume of LVs is high, volume of PCs is high and volume of HVs is medium then frequency of PDO crashes is high.

If speed is high, volume of LVs is high, volume of PCs is high and volume of HVs is high then frequency of PDO crashes is high.

The estimated values of turning points of TMFs in case 10 are tabulated in Table 12.

Developed MFs for case 10 are represented in.

To sum up, results of comparison between the numbers of predicted and observed PDO crashes for all investigated cases are tabulated in Table 13.

4. Sensitivity Analysis and Discussion

The Final part of this research is sensitivity analysis in which the impact rate of each variable on occurrence of PDO crashes on urban highways is determined. To do this the best model which presented the most accurate results for PDO crashes is applied. The most accurate model for predicting the frequency of PDO crashes is case 10 with goodness of fit equal to 0.74. To determine the effect rate of variables on likelihood of crashes, each variable in data set (300 data) is increased by one unit and the change rate in the estimated frequency of crashes is investigated. As the average of speed, volume of LVs, HVs and PCs in data set is 65, 630, 565 and 1095 respectively, it is more accurate to compare the impact of one-unit augmentation in speed with 9.7, 8.7 and 16.8 unit increase in the value of LVs, HVs and PCs respectively on likelihood of crashes. Clearly 9.7, 8.7 and 16.8 are equal to 630, 565 and 1095 divided by 65 respectively.

Each data set comprises 300 data for each variable. The results of sensitivity Analysis for input variables of speed, volume of LVs, HVs and PCs are depicted. To measure the impact rate of each variable on likelihood of PDO crashes, the impact rate for all 300 data are averaged to determine the specific value. The specific values determined for all variables on occurrence of crashes on urban highways are presented for PDO crashes in Table 14.

As displayed in Table 14, the effect rate of speed on likelihood of PDO crashes on urban highways is equal to 0.3. It indicates that when the value of speed increases by one unit, for example from 58 km/h to 59 km/h, the frequency of PDO crashes is increased by 0.3 unit or more clearly when the moving speed is increased by ten units, for example from 58 km/h to 68 km/h, the frequency of PDO crashes augments to about 3 units. Then impact rate of the volume of PCs on likelihood of PDO crashes on urban highways is equal to 0.24. It indicates that when the volume of PCs augment to 9.7 unit, e.g. from 605 pc/h to 614.7 pc/h, the frequency of PDO crashes increases by 0.24 unit or more clearly when the volume of LVs is increased by 97 units, for example from 605 pc/h to 702 pc/h, the frequency of PDO crashes increases 2.4 units. The effect rate of volume of HVs on occurrence of PDO crashes is 0.2, indicating as volume of HVs augment to 8.7 units, e.g. from 550 to 558.7, frequency of PDO crashes increases 0.2 unit on urban highways or more tangibly, if the volume of HVs is increased by 87 units, for example from 550 to 637 units, frequency of PDO crashes augment to 2 units. Eventually the effect rate of volume of PCs on likelihood of PDO crashes is 0.22. It implies that if the volume of PCs augments to 16.8 units in an hour, the frequency of PDO crashes is increases by 0.22 units.

More tangibly, when the volume of PCs is increased by 168 units, for example from 1105 pc/h to 1273 pc/h, the frequency of PDO crashes increases by about 2 units.

In addition to sensitivity analysis, the values of impact rate in Table 4 confirm the prioritization of variables in terms of their measure of influence on occurrence of PDO crashes. It is clear that speed have the most influence on likelihood of crashes in comparison with volume of PCs (0.22), HVs (0.2) and LVs (0.24) because of the bigger value of effect rate (0.3). After that, volume of LVs is more influential in occurrence of PDO crashes on urban highways. It is noteworthy to remind that the average of speed, volume of LVs, HVs and PCs in data set is 65, 630, 565 and 1095 respectively. Therefore, it is logical to compare the effect of one-unit increase in speed with 9.7, 8.7 and 16.8 unit increase in the value of LVs, HVs and PCs respectively on likelihood of PDO crashes. Where 9.7, 8.7 and 16.8 are equal to 630, 565 and 1095 divided by 65 respectively.

5. Conclusion

In this research, fuzzy Bayesian inference based models to predict the frequency of PDO crashes on urban highways are developed. Then, the estimated frequency of PDO crashes by developed models is contrasted with the frequency of observed crashes for each case that is specific combination of effective factors. The results of comparison indicate each model accuracy. The best model has got maximum value of[2] goodness of fit or minimum value of RMSE.

The results of research indicate that when speed, volume of LVs, volume of HVs and volume of PCs are considered to influence on the frequency of crashes straightly, the rules develop the model to predict frequency of crashes most exactly by $R^2 = 0.74$ and RMSE = 0.0186 which implies that all variables considered in the research are effective on occurrence of PDO crashes. The complementally results showed that the variable speed is more prominent than the volume elements of traffic in likelihood of PDO crashes. This is concluded when the value of R^2 for case 1 is compared to that for cases 2, 3 and 4. After speed, the variable volume of LVs is more effective on occurrence of PDO crashes than volume of HVs and PCs. This is concluded when the value of R^2 for case 5 is compared to that for cases 6 and 7. Finally the role of PCs is more influential in occurrence of PDO crashes than HVs. This is clear when the value of R^2 for cases 8 and 9 are compared together.

After prioritization of variables in terms of influence on occurrence of PDO crashes, the model best fitting the data with highest value of goodness of fit were employed to do the sensitivity analyze. Sensitivity analyze specifies the effect rate of effective factors on crashes occurrence. The results imply that the effect rate of speed on likelihood of PDO crashes on urban highways is equal to 0.3. It indicates that when the value of speed increases by 1 unit, the number of PDO crashes augment to 0.3 units. The effect rate of the volume of LVs on occurrence of PDO crashes on urban highways is equal to 0.24. It represents that when the value of LVs is increased by 9.7 units, the number of PDO crashes augment to 0.24 units. The effect rate of volume of HVs on likelihood of PDO crashes is 0.2 indicating as volume of HVs augments to 8.7 units, frequency of PDO crashes is increased by 0.2 units. At last the effect rate of volume of PCs on occurrence of PDO crashes is 0.22 indicating as the frequency of PCs in traffic flow of urban highway augments to 16.8 units in an hour, the frequency of PDO crashes is increased by 0.22 unit.

6. References

1. Chen S, Chen F, Wu J. Multi-scale traffic safety and operational performance study of large trucks on mountainous interstate highway. Accident Analysis and Prevention. 2011, 43, pp. 429-438. https://doi.org/10.1016/j.aap.2010.09.013 PMid:21094341
2. Ayati E, Abbasi E. Investigation on the role of traffic volume in crashes on urban highways. Safety research. 2011, 42, pp. 209-214. https://doi.org/10.1016/j.jsr.2011.03.006 PMid:21855692
3. Kaplan S, Prato CG. Risk factors associated with bus accident severity in the United States: A generalized ordered logit model. Journal of Safety Research. 2012, 43, pp. 171-180. https://doi.org/10.1016/j.jsr.2012.05.003 PMid:22974682
4. Chang L, Chien J. Analysis of driver injury severity in truck-involved crashes using a non-parametric classification tree model. Safety Science. 2013, 51 (1), pp. 17-22. https://doi.org/10.1016/j.ssci.2012.06.017
5. Feng SH, Zhenning L, Yusheng C, Zhang G. Risk factors affecting fatal bus accident severity:
6. Their impact on different types of bus drivers. Accident Analysis and Prevention. 2016, 86, pp. 29-39. https://doi.org/10.1016/j.aap.2015.09.025 PMid:26513334
7. Green CP, Heywood JS, Navarro M. Traffic crashes and the London congestion charge. Journal of Public Economics. 2016, 133, pp. 11-22. https://doi.org/10.1016/j.jpubeco.2015.10.005
8. Castillo-Manzano JI, Castro-Nuño M, Fageda X. Exploring the relationship between truck load capacity and traffic crashes in the European Union. Transportation Research Part E. 2016, 88, pp. 94-109. https://doi.org/10.1016/j.tre.2016.02.003
9. Tseng CM, Yeh MS, Tseng LY, Liu HH, Lee MC. A comprehensive analysis of factors leading to speeding offenses among large-truck drivers. Transportation Research Part F. 2016, 38, pp. 171-181. https://doi.org/10.1016/j.trf.2016.02.007

10. Goh K, Currie G, Sarvi M, Logan D. Factors affecting the probability of bus drivers being at-fault in bus-involved crashes. Accident Analysis and Prevention. 2014, 66, pp. 20-26. https://doi.org/10.1016/j.aap.2013.12.022 PMid:24486771
11. Grytnes R, Shibuya H, Dyreborg J, Grøn S, Cleal B. Too individualistic for safety culture? Non-traffic related work safety among heavy goods vehicle drivers. Transportation Research Part F. 2016, 40, pp. 145-155. https://doi.org/10.1016/j.trf.2016.04.012
12. Evgenikos P, Yannis G, Folla K, Bauer R, Machata K, Brandstaetter C. Characteristics and causes of heavy goods vehicles and buses crashes in Europe. Transportation Research Procedia. 2018, 14, pp. 2158- 2167. https://doi.org/10.1016/j.trpro.2016.05.231
13. Boyce WS. Does truck driver health and wellness deserve more attention? Journal of Transport and Health. 2016, 3, pp. 124-128. https://doi.org/10.1016/j.jth.2016.02.001
14. Zhu X, Srinivasan S. Modeling occupant-level injury severity: An application to large-truck crashes. Accident Analysis and Prevention. 2011, 43, pp. 1427-1437. https://doi.org/10.1016/j.aap.2011.02.021 PMid:21545876
15. Abdelwahab H, Abdel-Aty M. Investigating the effect of light truck vehicle percentages on head-on fatal traffic crashes. Journal of Transportation Engineering. 2004, 130 (4), pp. 429-437. https://doi.org/10.1061/(ASCE)0733-947X(2004)130:4(429)
16. Daniel J, Chien SIJ. Truck safety factors on urban arterials. Journal of Transportation Engineering. 2004, 130 (6), pp. 742-752. https://doi.org/10.1061/(ASCE)0733-947X(2004)130:6(742)
17. Mohamed N, Mohd-Yusoff MF, Othman I, Zulkipli ZH, Osman MR, Voon WS. Fatigue-related crashes involving express buses in Malaysia: will the proposed policy of banning the early-hour operation reduce fatigue-relatedcrashes and benefit overall road safety? Accident Analysis and Prevention. 2012, 45, pp. 45-49. https://doi.org/10.1016/j.aap.2011.09.025 PMid:22239931
18. Edwards JRD, Davey J, Armstrong AK. Profiling contextual factors which influence safety in heavy vehicle industries. Accident Analysis and Prevention. 2014, 73, pp. 340-350. https://doi.org/10.1016/j.aap.2014.09.003 PMid:25269101
19. Mussonea S, Bassanib M, Mascib P. Analysis of factors affecting the severity of crashes in urban road intersections. Accident Analysis and Prevention. 2017, 103, pp. 112-122. https://doi.org/10.1016/j.aap.2017.04.007 PMid:28432882
20. Chen GX, Fang Y, Guo F, Hanowski RJ. The influence of daily sleep patterns of commercial truck drivers ondriving performance. Accident Analysis and Prevention. 2016, 91, pp. 55-63. https://doi.org/10.1016/j.aap.2016.02.027 PMid:26954762 PMCid:PMC4828254
21. Ruslia R, Haquea MM, King M, Wong Shaw Voon WS. Single-vehicle crashes along rural mountainous highways in Malaysia: An application of random parameters negative binomial model. Accident Analysis and Prevention. 2017, 102, pp. 153-164. https://doi.org/10.1016/j.aap.2017.03.002 PMid:28314189
22. Fowles R, Loeb PD, Clarke W. The cell phone effect on truck crashes: A specification error approach. Transportation Research Part E. 2013, 50, pp. 18-28. https://doi.org/10.1016/j.tre.2012.10.002
23. Najafi SH, Flintsch G, Khaleghian S. Fuzzy logic inference-based Pavement Friction Management and real-time slippery warning systems: A proof of concept study. Accident Analysis and Prevention. 2016, 90, pp. 41-49. https://doi.org/10.1016/j.aap.2016.02.007 PMid:26914521