

# Traffic Accident Analysis With Or Without Bus Priority

Lin li<sup>1</sup>, Tan Wenfang<sup>2</sup>, Miao Yuqi<sup>3</sup>, Zhou Xiyuan<sup>4</sup>, Gao Rujing<sup>5</sup>

<sup>123</sup>Nanjing Forestry University, Nanjing Jiangsu Province, 210037, China

<sup>45</sup>Jinling High School Hexi Campus, Nanjing Jiangsu Province, 210019, China

<sup>1</sup>LinLi401@njfu.com.cn

<sup>2</sup>1253729669@qq.com

<sup>3</sup>1911083635@qq.com

<sup>4</sup>743695641@qq.com

<sup>5</sup>2225656860@qq.com

**Abstract.** Based on the empirical analysis of the bus accident data, this paper discusses the impact of bus priority on road safety. Based on the conclusions of road safety performance and the results of traffic accident research, the public traffic priority measures are proposed. Through the empirical analysis of the types of accidents, it has been found that the proportion of bus collision accidents involved in the bus priority measures is significantly reduced, which indicates that the bus will give priority to the collision with other vehicles. Mixed-effect negative binomial (MENB) regression model and BPNN neural network model were used to consider bus accidents that have more extensive impact on accident rates in some sections of the road. The results show that the safety benefits are more significant when providing bus priority measures. The sensitivity analysis performed on the BPNN model shows that the predicted values of the accident frequencies are basically the same between the two models. The performance of MENB model results shows that it is advantageous to use a mixed-effects modeling method to predict accident counts in practice because it can take into account the effects of specific factors. Therefore, research on the priority status of buses will help improve road safety. When implementing road programs, road management agencies should give priority to bus priority measures.

## 1. Introduction

There are various types of bus priority projects internationally, each of which is basically different in the amount of road space or time (or a combination of both) allocated to transportation vehicles. Regardless of its form, there is ample evidence that the priority of public transport will improve the travel time of public transport and increase the reliability and attractiveness of public transport.

Based on the empirical analysis of bus accident data in Nanjing, this paper discusses the road safety impact of bus priority. The focus of this study is to understand the impact of public transport priority implemented in Nanjing's "Smart Bus" BRT system. Smart Bus is a freeway system similar to the Los Angeles Metro, which has a relatively large number of public transport passengers and a relatively high price/performance ratio compared to the busway's BRT system. In order to analyze the stringency, two modeling methods are used in this paper: (1) negative binomial model of mixed effects and (2)



neural network model. Comparing these two methods, this is a secondary goal of the study.

## 2. Purpose of research

The purpose of this study is to understand the type and frequency of bus accidents without considering bus priority and considering bus priority. In addition, the project also aims to explore a grade model of traffic accidents developed in the Nanjing public transportation system.

## 3. Method

The first method uses the mixed-effect negative binomial regression method, which uses the negative binomial distribution hypothesis. This is a widely used road safety research method because it can handle non-negative and typically distributed accident count data. The second model was developed using neural network principles. In this study, a neural network based on the commonly used back propagation algorithm was selected and processed.

### 3.1 Mixed effect negative two term (MENB) bus accident model

In terms of road safety, the random-effects-binomial (RENB) modelling method was used in previous studies to address possible spatial inhomogeneities and time discontinuities in bus accident records. The study used a relatively recent computational statistical approach to model location and time-specific variables as cross-cutting, independent effects. Compared with RENB, the mixed-effect negative binomial (MENB) regression model provides the following key advantages:

- 1) It allows for random effects, not necessarily nested in traditional random effects models;
- 2) It is more flexible when dealing with missing data problems;
- 3) Due to repeated observations, the lack of statistical capacity was overcome.

With  $E(A_{ij})$  representing the predicted number of accidents along bus route segment  $i$  at time  $j$ , the structure of the MENB model is given as:  $E(A_{ij}) = \exp(X_{ij}\beta + L_i l_i + T_j t_j + \varepsilon_{ij})$  (1)

Where  $X_{ij}$  is the matrix representing factor contrasts and covariates;  $\beta$  is the vector of pooled coefficients (fixed effect);  $L_i$  is the matrix to account for location-specific effect;  $l_i$  is the vector of coefficients representing location-specific effects;  $T_j$  is the matrix to account for time-specific effect;  $t_j$  is the vector of coefficients representing time-specific effects;  $\varepsilon_{ij}$  is the vector of residual errors.

Following the combination of matrices  $L$  and  $T$  into to a single matrix  $Z$ , and random vector  $l$  and  $t$  into a single vector  $\gamma$ , the formulation can be re-written as  $E(A) = \exp(X\beta + Z\gamma + \varepsilon)$  (2)

The residual error ( $\varepsilon$ ) and random effects ( $\gamma$ ) terms are assumed to take on the normal distribution with means 0 and variances  $a$  and  $b$ , respectively. Table 1 provides a brief description and summary statistics of the covariates used in the MENB model.

Similar to the aggregate analysis, the  $R_\alpha^2$  as proposed by Miaou et al.(1996) is used to assess the model's goodness-of-fit:  $R_\alpha^2 = 1 - \frac{\alpha}{1 + \alpha_{\max}}$  (3)

where  $\alpha$  is the over-dispersion parameter for final MENB model; and  $\alpha_{\max}$  is the over-dispersion parameter for base model with only a constant term.

For the purpose of model comparison (MENB vs. neural net-work), the root-mean-square error (RMSE) is used and this is given by  $RMSE = \sqrt{\frac{\sum_{i=1}^m (Y_i - \hat{Y}_i)^2}{m}}$  (4)

Table 1- Summary statistics of variables used in the MENB model

Variable	Min	Max	Mean	S.D.
Accident frequency (collisions/year)	0	29	3.68	4.89
Year <sup>a</sup> (2009= 1; 2010 = 2; 2011= 3)	1	3	2	0.82
Location <sup>a</sup> (segment 1 = 1 to segment 99 = 99)	1	99	50	28.58
Length of bus route segment <sup>b</sup> (km)	2.5	55.0	15.94	10.11
Average annual daily traffic (AADT) of segment <sup>c</sup>	1,495	78,433	7,335	6,286
Number of bus services per week	6	314	111.43	87.63
Stop density (number of bus stops/km)	0.53	7.33	2.50	0.941
Presence of bus priority(with = 1; otherwise = 0)	0	1	0.15	0.36
Total observations, n = 297				
<sup>a</sup> Coded as string variable as required in R software.				
<sup>b</sup> Defined based on bus service route and presence of bus priority.				
<sup>c</sup> The weighted average method is applied to compute the AADT value for segments that comprise more than one road section.				

Here, and the frequency of the accident that is observed and predicted along the path, is the size of the dataset. RMSE statistics provide a method to measure the average error prediction of the model, which is close to 0, indicating that the model has predicted the observed data.

### 3.2 Neural network modeling

Neural networks are attractive in applications where nonlinear and complex functional forms exist between input and output. This is because unlike the statistical regression model, the neural network does not need to establish a functional form of linking dependent variables and independent variables. Another key advantage offered by these networks is that the general handling of the data is also accurate, that is, when the model presents incomplete data input, good results can be obtained. This may be an increasing reason for applying neural network modeling in the transportation field. In this study, a three-layer feed-forward neural network model (hereinafter referred to as BPNN) based on the back-propagation method using the Lavenberg-Marquardt algorithm was used.

The BPNN model structure is shown in Figure-1, where  $x_n$  are the input neurons that represent the accident related characteristics,  $z_k$  the hidden neurons and  $Y$ , the output neuron in the model.

The underlying concept in this technique is based on the popular back-propagation algorithm, which works by updating the weights in the model such that the error between the actual and desired outputs ( $E$ ) is minimised. This is essentially a four-step process that starts with a feed-forward computation where an input pattern  $x_i$  is presented to the network. The second step involves a back-propagation from the output layer, where the aim is to correct the weights  $w_{k,1}$  to minimise the error  $E$  :

$$\Delta w_{k,1} = -\eta \frac{\partial E_k}{\partial w_{k,1}} \quad (5)$$

In the Lavenberg-Marquardt algorithm, a similar parameter is used to adjust the training process. From the equation (6), it can be proved by using the rules of chain decomposition.

$$\Delta w_{k,1} = -\frac{\eta}{2} (y - o) f'(y_t^{inp}) z_k \quad (6)$$

Here,  $o$  represents the desired value for the output neuron based on the input pattern  $x_i$ , while  $y_t^{inp}$  is the summation of the Weighted outputs from the hidden neurons  $z_k$ . In a similar fashion, the weights for the hidden layer can be computed in the third step:

Here, it represents the expected value of the input neuron based on the input mode, but the sum of

the weighted output from the hidden neurons. In a similar way, the weight of the hidden layer can be calculated in the third step.

$$\Delta w_{n,k} = -\eta \cdot g'(z_t^{inp}) x_k (y - o) f'(y_t^{inp}) w_{k,1} \quad (7)$$

The last step of the backpropagation algorithm is to update the weight of each output and hidden neuron in the model:

$$w_{k,1}^{new} = w_{k,1}^{old} + \Delta w_{k,1} \quad (8)$$

$$w_{n,k}^{new} = w_{n,k}^{old} + \Delta w_{n,k} \quad (9)$$

In this study, the BPNN model was developed using MATLAB. To facilitate the comparison of model results, the same input variables were used in the development of the BPNN model. These variables are important in the final model. A separate neuron is set in the output layer to indicate the frequency of the accident. All transfer functions in the hidden and output layers are transition functions of the hyperbolic tangent function. In order to train and test the model, the data set was also randomly divided into two parts (with a ratio of 3:1). Another key step in developing a BPNN model is to determine the number of hidden neurons (k). In this study, a series of values were used, such as  $k = 1, 2, 3, \dots, 9, 10$ . In the development of the BPNN model, the MENB model was chosen to produce the minimum RMSE value. Assuming that the operation of each neural network model produces unique results, the BPNN model will run 10 times to obtain RMSE.

Finally, through the comparison of the model results and the understanding of the key accident risk factors, the sensitivity of the relationship between the accident frequency and each variable in the model is analyzed. This is achieved by perturbing the value of the variable while keeping other variables constant. The model generates a new network output for each simulation input, which records changes in results and determines the effect of a single variable of interest.

## 4. Results and discussion

### 4.1 Analysis on the type of bus accidents

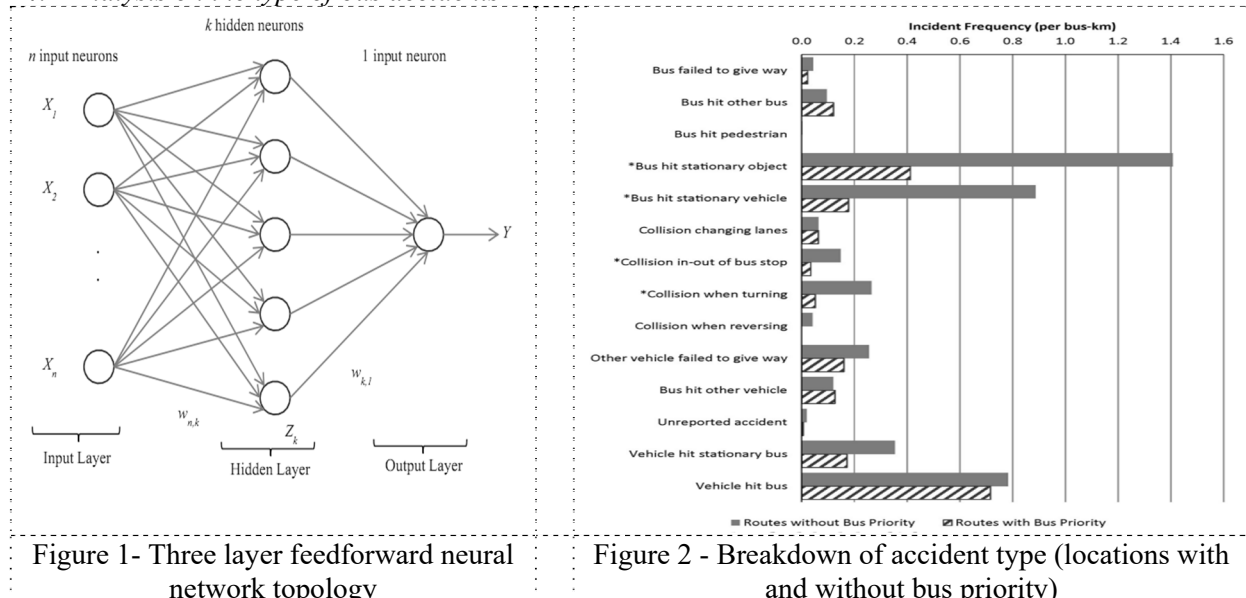


Figure 1- Three layer feedforward neural network topology

Figure 2 - Breakdown of accident type (locations with and without bus priority)

Table 2 -MENB model results for bus accident frequency

Variable	Estimate	P-value
Intercept	-6.640	0.000
Services per week	0.006	0.000
Ln(AADT)	0.431	0.001
Ln(route section length)	0.773	0.000
Stop density	0.389	0.000
Bus priority = yes	-0.766	0.002
Bus priority = no		0 (Reference)
Random effect	Variance	Standard deviation
Year	0.357	0.598
Location	0.195	0.441
Dispersion parameter, $\alpha$	0.242	
95% CI for $\alpha$	[0.169, 0.429]	
Log likelihood	-607.205	
AIC	1232.4	
$R_\alpha$	0.807	

Figure 2 shows the frequency of accidents that occur with priority and no priority for buses. Obviously, the most common accidents include collisions between buses and other vehicles or bus collisions with stationary objects. These findings reflect an earlier study that found that bus collisions with objects were the most common and accidents involving bus collision with pedestrians were rare.

When comparing routes considering bus priority and bus priority, the most obvious difference is the proportion of accidents that occur when buses hit stationary objects and vehicles. The former was reduced by about 70% compared with the control group, with significant significance ( $p < 0.05$ ). The latter had a large decrease (about 80%) and a significant decrease at the  $p < 0.05$  level. A similar decrease was also recorded in the number of collisions between bus records and the number of bus collisions with other vehicles ( $p$  less than 0.05). The change in these percentages is likely due to the fact that public transport priority measures have facilitated the operation of buses.

#### 4.2 MENB model

Table 2 shows the parameter estimation obtained by the maximum likelihood algorithm obtained in the glmmADMB package in the statistical software R, which is an open source language and an environment for statistical calculation. As a result, it was found that the dispersion coefficient is very different from zero, which indicates that the negative binomial error structure is more suitable than the Poisson structure. With the exception of the regional type variables, all other explanatory variables were found at the 5% level.

The model results show that the frequency of bus accidents on road sections increases with the increase in traffic volume (AADT), line length, and service frequency. These results are as expected, considering that these are related to open variables, for example, higher traffic volumes, longer route lengths, and higher service frequencies, which means buses are more likely to be in contact with other vehicles in the traffic stream. The model also shows that adding more bus stops per kilometer increases the risk of accidents ( $p = 0.000$ ), while bus priority can reduce the risk of accidents ( $p = 0.002$ ).

#### 4.3 BPNN model

From Figure 3, it is clear that AADT has a greater impact on the frequency of accidents than parking density. The results show that there is a parabolic relationship between the frequency of accidents and AADT. In general, the risk of accidents increases linearly with AADT, but when AADT is at a lower

range of AADT, the risk of accidents decreases. The impact of parking density is also obvious. The frequency of accidents increases with the number of stops per kilometer. The same observation can also be made when studying the influence of the route length (Figure 4). Similar to stop density, the risk of collision increases with time but is less pronounced than AADT. The results of the BPNN model also show that in the service frequency or stop density (Figure 5), the accident risk per unit increases.

#### 4.4 MENB and BPNN model

The results of the BPNN model are basically similar to the MENB model in explaining the relationship between the variable and the bus related accident frequency. The results show that the performance of the MENB and BPNN models can be compared, and the former has a slightly better performance (RMSE=2.59/2.75). Sensitivity analysis focuses on bus priority, because the result not only provides the impact on bus priority, but also provides a way to compare the performance of the model generated by the two variable state of the variables. To carry out this analysis, the initial data set is divided into two groups, depending on whether the bus priority exists. The first includes a route without bus priority (N=252), and the second includes a bus priority route (N=45). These two sets of data are submitted to the final MENB and BPNN models and predict the frequency of the bus accidents of the two models captured in Table 4.

According to the results of Table 4, the following key observations can be carried out. All data sets show the security of bus priority. The results of test show that all the data sets of the two models were statistically significant ( $P < 0.05$ ).

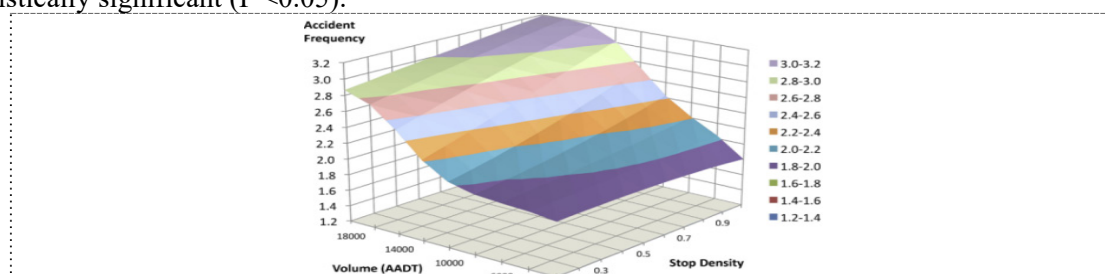


Figure 3. AADT and the effect of route stop density on the frequency of the accident (route -section25)

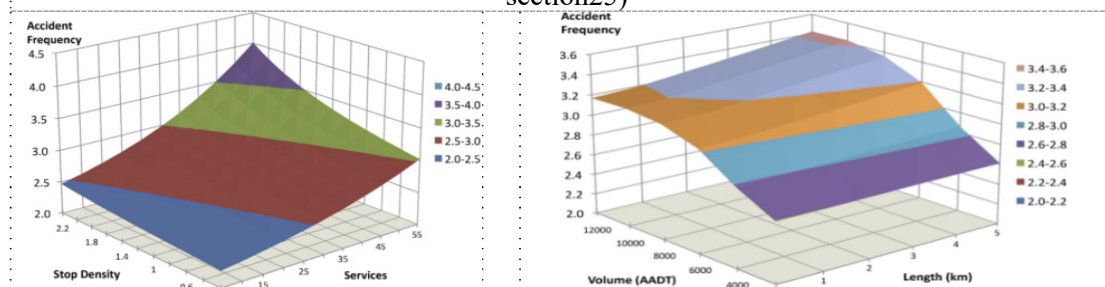


Figure 4. AADT and the effect of route length on the frequency of the accident (route -section25)

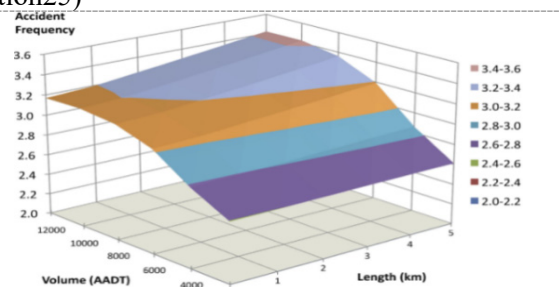


Figure 5. The effect of stop density and frequency on the frequency of the accident (route-section25)

The BPNN model shows that the frequency of accidents in the route segment decreased by 53.4% when there was a bus priority route. The MENB model results show that this effect is 53.5% (this result is equivalent when using the parameter estimates obtained from the MENB model in the previous section).

Overall, the results show that when modeling accident counts, it is advantageous to use mixed-effect negative binomial regression methods to consider the effects of time and location. In terms of RMSE, the MENB model outperforms the BPNN model in this study. In practice, the application of neural networks for prediction of accident counts is limited due to the complexity of the model. However, it may be useful for accident prediction models to consider developing equivalent



parameters and non-parametric models to assist in the development of neural network models because sensitivity analysis can be performed to provide useful insights.

## 5. Conclusions and suggestions

Through the analysis of bus-related accident data, the security impact of introducing bus priority measures in Nanjing was analyzed. Two models, the mixed-effect negative binomial (MENB) and the neural network based on the back-propagation method (BPNN), were developed to explore factors related to the frequency of bus accidents and to compare the two models.

The results of the accident analysis showed that after the implementation of the bus priority measures, the bus accidents hitting stationary objects and other vehicles were significantly less than when bus priority was not considered ( $p$  was less than 0.05), and the collision occurred at the bus stop. The accident is also decreasing. These reductions may be due to the priority of the bus to facilitate the operation of the bus. The results show that the frequency of bus accidents on road sections increases with the increase in traffic volume (AADT), line length, and service frequency. Although the impact on AADT, route length, service frequency and stop density is consistent with previous research results, it is found that the priority effect of public transport (positive safety benefit) is contrary to previous research results of. This may be due to the different design of Nanjing's traffic priority. In Nanjing, the bus priority lane is located in the slowest lane. Therefore, a more direct way can be used to isolate buses from mainstream traffic.

Table 3- Priority sensitivity analysis of public transport

Model	Route-section dataset	Predicted Accident frequency (per km)	
		With bus priority	Without bus priority
MENB (RMSE = 2.59)	Without bus priority (N= 252)	0.093(S.D=0.090)	0.201(S.D=0.194)
	With bus priority (N= 45)	0.499(S.D=0.293)	1.073(S.D=0.629)
	All route-sections (N= 297)	0.167(S.D=0.226)	0.359(S.D=0.486)
BPNN (RMSE = 2.75)	Without bus priority (N= 252)	0.173(S.D=0.216)	0.234(S.D=0.259)
	With bus priority (N= 45)	0.432(S.D=0.289)	1.682(S.D=1.421)
	All route-sections (N= 297)	0.213(S.D=0.247)	0.457(S.D=0.800)

Through the sensitivity analysis of the BPNN model, the intuitive perception of the relative influence of AADT, parking density, line length and service frequency on the accident frequency is given. It can be clearly noticed from these graphs that accident risk is more sensitive to AADT than other variables. The results show that compared with the estimated effect of the MENB model (53.5%), the incidence of bus-priority bus accidents was reduced by 53.4%. Compared with the BPNN model (RMSE=2.75), the MENB model (RMSE=2.59) records show that the MENB regression method can be used to better consider the specific impact of time and place when modeling accident statistics.

In summary, although the findings provide useful insights into bus accidents and may be a useful planning tool for transportation agencies, there is still much room for research in this area. In the future research, bus accident data can be further collected to improve the effectiveness of the model. These factors are of great significance in explaining the frequency of bus accidents. In addition, we will further study the different safety effects of different bus priority plans and identify key factors related to the severity of different accidents to further deepen our understanding of bus safety.

## Acknowledgment

Fund project: Jiangsu university brand professional construction project (PPZY2015A063)

## References

- [1] Goh, K., Currie, G., Sarvi, M., Logan, D., 2013a. Road safety benefits from bus priority—an empirical study. *Transportation Research Record—Journal of the Transportation Research Board*.
- [2] af Wåhlberg, A.E., 2002. Characteristics of low speed accidents with buses in public transport.

- Accident Analysis & Prevention 34 ,637–647.
- [3] Brenac, T., Clabaux, N., 2005. The indirect involvement of buses in traffic accident processes. *Safety Science* 43 , 835–843.
  - [4] Chin, H.C., Quddus, M.A., 2003. Applying the random effect negative binomial model to examine traffic accident occurrence at signalized intersections. *Accident Analysis & Prevention* 35 , 253–259.
  - [5] Mazloumi, E., Rose, G., Currie, G., Sarvi, M., 2011. An integrated framework to predict bus travel time and its variability using traffic flow data. *Journal of Intelligent Transportation Systems* 15 , 75–90.
  - [6] The Mathworks Inc., 2012a. Matlab. The Mathworks Inc., Natick, MA.
  - [7] Vlahogianni, E.I., Karlaftis, M.G., Orfanou, F.P., 2012. Modeling the effects of weather and traffic on the risk of secondary incidents. *Journal of Intelligent Transportation Systems* 16 , 109–117.
  - [8] af Wählberg, A.E., 2004a. Characteristics of low speed accidents with buses in public transport: Part II. *Accident Analysis & Prevention* 36 ,63–71.
  - [9] Yang, J., Peek-Asa, C., Cheng, G., Heiden, E., Falb, S., Ramirez, M., 2009. Incidence and characteristics of school bus crashes and injuries. *Accident Analysis & Prevention* 41 , 336–341.
  - [10] Abdel-Aty, M.A., Radwan, A.E., 2000. Modeling traffic accident occurrence and involvement. *Accident Analysis & Prevention* 32 , 633–642.
  - [11] Hagan, M.T., Menhaj, M.B., 1994. Training feedforward networks with the marquardt algorithm. *IEEE Transactions on Neural Networks* 5 , 989–993.
  - [12] Strathman, J.G., Wachana, P., Callas, S., 2010. Analysis of bus collision and non-collision incidents using transit its and other archived operations data. *Journal of Safety Research* 41 , 137–144.
  - [13] Cheung, C., Shalaby, A.S., Persaud, B.N., Hadayeghi, A., 2008. Models for safety analysis of road surface transit. *Transport Research Record* 2063, 168–175.
  - [14] Jovanis, P.P., Schofer, J.L., Prevedouros, P., Tsunokawa, K., 1991. Analysis of bus transit accident: empirical, methodological and policy issues. *Transportation Research Record* 1322, 17–28.
  - [15] Milton, J., 1998. The relationship among highway geometrics, traffic-related elements and motor-vehicle accident frequencies. *Transportation* 25 , 395–413.
  - [16] Chimba, D., Sando, T., Kwigizile, V., 2010. Effect of bus size and operation to crash occurrences. *Accident Analysis & Prevention* 42 , 2063–2067.