DEVELOPING ACCIDENT PREDICTION MODEL OF TWO-LANE HIGHWAY ON THE PLATEAU

Bingbing Feng*, Yulong He, Guixian Qu, Mingyang Jia

*Beijing University of Technology, Beijing Key Laboratory of traffic engineering, Beijing 100124;
 E-mail : bingbing_feng@163.com

This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited

ARTICLE DETAILS

ABSTRACT

Article History:

Received 02 october 2017
 Accepted 06 october 2017
 Available online 11 october 2017

Keywords:

traffic engineering; Qinghai-Tibet Plateau; two-lane highway; zero-inflation Poisson distribution; negative binomial distribution; accident prediction model

Based on the analysis of the traffic flow, in this research, roadside environment and traffic safety data of Qinghai-Tibet highway, the characteristics of road traffic and the influence of altitude, oxygen content, traffic flow and environment on the traffic accidents were studied. A non-rural road accident prediction model based on zero-inflation Poisson and an accident prediction model based on negative binomial distribution were established. In the paper, the effect of oxygen content, the proportion of large vehicles and the direction distribution coefficient of the traffic flow on the number of accidents in the rural roads were studied. The results indicate that there are significant differences between the characteristics of rural road accidents and non-rural road accidents. Roadside hazard level and the access density respectively have the most significant influence on the number of non-rural road and rural road accidents.

1. INTRODUCTION

The adverse weather and road environment of Qinghai-Tibet plateau are the main cause of frequent serious traffic accidents in Tibetan region. The mortality rate per 10,000 vehicles and mortality rate per 100,000 persons in Qinghai-Tibet highway are much higher than the national average mortality level. In 2012, the amount of traffic accidents in two areas of Qinghai Tibet is 1821, 871 people were killed and 2219 people were injured. The mortality rate is 8.33 per 10,000 vehicles and 10 per 100,000 persons, the loss is triple than the national average level^[1], thus the traffic safety conditions need to be improved urgently.

The secondary two-lane highway is the main road type in Qinghai and Tibet, in this paper, secondary two-lane highway in plateau area was investigated. Through collecting the traffic flow data, traffic accident data, road and roadside data, traffic environment data, the analysis of plateau characteristics was conducted, and the two-lane road accident prediction model was established to make quantitative analysis on the influence law of traffic factors.

There is few research on the highway traffic safety prediction model of high altitude area. Most of the research focuses on the traffic safety prediction model under the non-plateau condition.

There are many studies on two-lane road accident prediction. The American two-lane accident prediction method have been widely recognized in the United States, by collecting the highway data and accident data in Washington and Minnesota, the road accident frequency prediction model and accident frequency prediction model of three kinds of intersections were developed. The road accident frequency prediction model used the negative binomial distribution^{[2][3][4]}. Recently, the domestic researchers conduct the research on highway safety of the non-plateau area. Liande ZH et al.(2009) established a predication model based on negative binomial regression by collecting the accident data and traffic data of Jingjintang highway and the accident influence factors such as environmental variables, flow rate, velocity were analysed^[5]. Hailong G et al. (2007) established the traffic accident frequency prediction model for the discrete traffic accident data using the zero inflated probability

model^[6]. Chengcheng Tet al.(2008) established the accident prediction model of the two-lane highway^[7], the roadside accident prediction model of the two-lane highway^[8] and the two-lane highway in towns and villages located in mountainous area by using negative binomial distribution model and conducted the quantitative study to accident forming mechanism^[9].

In summary, there is few research on the traffic safety of high altitude highway on the plateau. The traffic accident prediction model of quantization the traffic influence factors mainly focuses on the expressway and two-lane highway in the plain area. Therefore, in this paper, based on a large number of field investigations and data collection of road conditions, traffic flow parameters, the roadside environment on two-lane highway in the Qinghai-Tibet plateau, the main influencing factors of traffic safety of were analysed, the accident prediction model of two-lane plateau highway was established, laying a good foundation for improving the safety level of the plateau highway.

2. Data Collection and Methodology

2.1 Data Collection and Processing

In this part, a lot of traffic safety data of two-lane highway in Qinghai was collected, including the traffic flow, accident data, roadside data and environment data, etc. The traffic flow data was collected through paving Metro Count on various road segments. Road information and roadside data were acquired through road design linear data, field survey and the video data during the field survey. Environment data mainly includes the altitude and oxygen content acquired by the handheld device equipment. The total length of the two-lane plateau highway is 793km. Traffic flow data is shown as table 1.

Table 1 Traffic flow data

Length	Sites	Sections	Altitudes	Vehicle sample number
--------	-------	----------	-----------	-----------------------

793 10 28 1800-3563 132917

Except for the roads through village, most of the sample roads are in mountainous areas, Gobi and other uninhabited areas with sparsely populated and little traffic volume. In order to make full use of the data, the roads were divided into non-rural roads and rural roads. Generally the length of road through the rural within 100 meters is considered to be the rural road [8]. In this research, the length of the rural road is 196km and non-rural road is 597km.

2.2 Accident prediction model of two-lane highway in the plateau

This study applied the ordered clustering of dividing sections [xi] [xii] in the establishment of accident prediction model of non-rural two-lane plateau road. In this way, the discrete accident data caused by going through large area of no-man-zone can be effectively solved. As for accident type prediction model, if a smaller variable statistical interval in the establishment of the model, the precision will be higher and the model will be more effective [xiii] [xiv]. The investigated road going through less villages, there are less rural roads rather than non-rural roads, with lots of accidents concentrated. Thus we use the fixed-length segmentation method with one kilometre statistical unit. The non-rural roads were divided into 218 segments and the rural roads were divided into 158 segments (for the reason of data acquisition, the collected accident segments are not all used to build the model). The purpose of the traffic accident impact factor analysis is to find out the factors related to the accident such as road alignment traffic flow and environment, so as to formulate prevention measures and reduce the frequency and severity of accident. At the same time, it is the foundation of the traffic accident prediction model. The relevant factors influencing the traffic accident are also the basis of the accident prediction model as the independent variables. In this chapter, some variables of road alignment, traffic flow and environment were selected (As shown in table 2) to analyse the correlation between variables and traffic accidents, and determine the independent variables used in the traffic accident frequency prediction model.

Through the Spearman's correlation coefficient analysis, the number of accidents in the non-rural road has a positive correlation with the traffic volume, the proportion of the cart, the roadside grade, the number of flat curves, and the number of access ports. And the correlation with the number of access ports is weak. Altitude is negatively correlated with oxygen content and correlation is strong, but both has a weak correlation with the accident number, so we choose one to enter the model. The positive correlation of the access number in rural road is higher than that of the non-rural road, and the influence of the roadside danger level is decreased and the proportion of the heavy vehicles shows negative correlation.

Table 2 independent variables

Variables	English abbreviations	Remarks
Number of accidents	NC	-
Exposure	Expo	One million vehicles-km/year
Daily volume	Q	-
Proportion of carts	Truck	-
Distribution coefficient of direction	Dir.	-
Roadside hazard level	Lars	Classified according to the roadside safety net area [xv]
Access density	Dace	Number of access ports in the study section
Flat curve density	DC	The number of flat curve with a radius of less than 600 meters [xvi]
Altitude	E	Average altitude in the study section
Oxygen content	O2	Ratio of oxygen content to the plain standard
Altitude drop	Edify	Absolute value of the elevation difference of one kilometre road

In this study, the independent variables were modelled by descriptive statistics and correlation analysis. The model is based on Poisson distribution regression model, negative binomial distribution regression model for accident prediction put forward by Miaou [xvii], and a zero-stacked Poisson/ negative binomial distribution-based regression model with good fit for large numbers of zero-data for the accident prediction [xviii]. The road section length and the traffic volume are the absolute variables that have effect on the number of accidents. In the past studies, they have been separated from other variables, written as the exposure product form, and not the same as other variables.

The model is as follows:

$$N_{ci} = E_{xpo} \times \exp(\beta_0 + \sum_{i=1}^n \beta_j x_{ij}) \quad (1)$$

In the formula, x_{ij} -the attribute value of road environment traffic for the i-th road segment; β_j -the parameter coefficients to be calibrated; E_{xpo} - traffic risk exposure index stands for road usage, the unit is one million vehicles-km/year.

3. Data Analysis and Results

3.1 Accident characteristics of two-lane plateau highway

3.1.1 General characteristics of two-lane plateau highway

Traffic accident characteristics of secondary two-lane roads in plateau area are shown in the following aspects:

- (1) Differences in accident sections: according to statistics, 67% of the accidents occurred on the rural roads, but only accounted for 25% of total mileage, so the average number of accidents per mileage is much higher than non-rural roads.
- (2) Differences in the distribution of accident patterns: according to statistics, accidents involving the motor vehicles are the main accident type on non-rural roads and rural roads. Single vehicle accidents in non-rural roads accounted for 30%, which is higher than that in the rural roads. Due to the complex and changeable terrain in the plateau area, long distance of depopulated roads and low oxygen content, it is more likely to lead to driver fatigue, which will result in the occurrence of abnormal conditions. The complex traffic composition of the rural road will inevitably lead to more accidents between vehicles and people.
- (3) Differences in the distribution of vehicle type in accidents: according to statistics, from the accident vehicle types, the passenger car is the main accident vehicle type on the rural roads and the non-rural roads, accounting for more than 40%. The main difference between the two roads is that the proportion of heavy trucks in non-rural road is higher, while the proportion of other vehicles in rural roads is far higher than that in non-rural roads. This is related to the functions of different roads, while the non-rural roads are served for long-distance transport and the tourism of Qinghai area, and the rural roads are served for resident travel.

3.1.2 Accident characteristics in the plateau

Under the special geographical conditions of the plateau, the effects of altitude, oxygen content and geographical features on traffic accidents were analysed respectively.

- (1) The relationship between Altitude and HMVK

Figure 1 is the scatter plots of HMVK and the altitude. (The interrupt data in the figure is part of Xi-Ning urban roads and Xi-Huang highway.). There is no significant distribution difference with altitude in the HMVK of rural roads.

In 0-50, the distribution is more uniform, the change of altitude has little effect on the accident on rural roads. On the non-rural roads, the HMVK increases with the altitude rising, becomes more dispersed and the number of roads with higher accident rate increases. In spite of randomness of the accidents, the discreteness of accident rate is high, the fitted decision coefficient is low, but the distribution trend is more obvious.

(2) The aggregation of traffic accidents with different oxygen contents

The level of the oxygen content is mainly related to the local altitude, climate and vegetation cover, so regional characters is obvious. Though the oxygen content is not the main influencing factor of the accident, it has a great influence on the driver's state and the performance of the vehicle [xix] [xx] [xxi], increasing the possibility of traffic accident. At the same time, the oxygen content also has an impact on the traffic environment. In the lower oxygen content, human living environment is poor, so there is few traffic volume in non-human area. Therefore, the different oxygen content area should have a certain correlation with the distribution of the accident characteristics. Figure 2 shows the changes of traffic accidents in different oxygen content areas (The oxygen content in the figure is the ratio of the measured value to the standard oxygen content of the plain). Clearly in the figure, the proportion of single motor vehicle accidents decreased with the increase of oxygen content; accidents between vehicles, vehicle and person show the overall growth.

Figure 1 Scatter plots of altitude and HMVK

Figure 2 Distribution of accident types of different oxygen contents

(2) HMVK by crash types on different geographical features

As shown in Fig.6, the HMVK of rural roads is much higher than the other three areas, consistent with the previous analysis. Among them, the accident rate among motor vehicles is triple or more than the other three areas, and the accident rate of motor vehicles and pedestrians is more than 6 times, and the rate of single motor accident is relatively low. The HMVK of the remaining three non-rural areas of the road is similar. However, the roadside environment in mountainous non-human areas is poor with more cliffs and longitudinal sections, causing more rollover and collision accidents. In the figure, single vehicle accidents and multi-vehicle accidents are more prominent. The number of multi-vehicle accidents in tourist areas is slightly higher than that in mountainous areas, and the accidents of vehicles and people are more significant than that of no-human-zone in mountainous areas. Accident rate of no-human-zone in Gobi is low because of the simple lateral roadside environment, less

interference and low traffic volume.

Figure 3 Distribution of HMVK by crash types on different geographical features

From the analysis of the traffic accidents of the investigated roads because of the differences in traffic composition and roadside environment between rural roads and non-rural road, there are large differences in the number of accidents distribution, morphological characteristics, causing the different accidents and proportion of vehicles. At the same time, we also analyse the accident distribution under different geographical characteristics such as altitude, oxygen, etc. In the following, suitable quantitative indicators for various impact factors were selected, and the accident prediction models for further quantitative analysis were used.

3.2 Estimation of the accident prediction model

In our study, regression analysis of backward culling is applying to the four regression models of Poisson regression, negative binomial regression, zero-stack Poisson and zero-negative quadratic regression model. First, the whole model was established. Based on the determination of the relevant indexes in the output, each time a variable that can't fit into the model was (For example, the p value is the smallest) until the regression equation no longer contains an independent variable that can't meet the criteria. As a result, the non-conforming independent variables were removed in turn and regressed. After several steps until the variables in the model contributed to the model, each coefficient was considered to be significant.

The parameters of the four models are estimated respectively, and the results of each parameter estimation are not enumerated here. Table 4 compares the statistical indexes of the four distribution models. In Poisson regression parameter estimation, P value of chi-square test of the model is 0.0000, and that means the model is statistically significant. In parameter estimation of negative binomial distribution model, alpha 95% confidence interval is (0.50, 1.25) and chibar² = 49.21, thus it can be at the significance level of 5% to reject the excessive dispersion parameter alpha = 0 null hypothesis (corresponding to Poisson regression). Table 3 shows that the negative binomial regression Log Likelihood, AIC and BIC are slightly better than Poisson regression; the Voong test statistic of the zero-stack Poisson regression parameter estimation is 3.04, which is more than 1.96, indicating that the model is superior to Poisson regression. And the Log Likelihood, AIC and BIC test statistic are all superior to negative binomial regression. The mean value of non-rural road accident is 1.13, which is close to the variance of 1.33, and a large number of zero-accident sample sections are more suitable for the establishment of the zero-inflation Poisson regression model. Based on the above results, zero-inflation Poisson regression model was used to predict the non-rural road accident.

The non-zero partial expression of the last zero-inflation Poisson regression model is as follows:

$$N_{ci} = E_{xpo} \times \exp(-2.7020 + 4.8616P_{truck} + 0.4869L_{rs} + 0.3126D_{hc} - 0.0154E_{diff}) \quad (2)$$

The modelling analysis was carried out on the rural road sections in the same way. Finally, the negative binomial regression model was used to predict the accident (formula 3), the cumulative standard residuals for model prediction ranged from -7.68 to 2.26 for exp. The sum of the

cumulative standard residuals is 1.7237 and far below 12.57 ($\sqrt{158}$), so the model is stable.

$$N_{eti} = E_{spo} \times \exp(-18.4388 - 6.3571P_{truck} + 0.1957D_{dc} - 10.0928O_2 - 10.3742D_{tr}) \quad (3)$$

Table 3 Index comparison between 4 regression models

Model	Log likelihood	AIC	BIC	Vuong	chibar2
Poisson	-350.63	3.27	-750.53		
NB	-328.97	3.07	-938.97		49.21
ZIP	-320.40	3.01	-1003.57	3.04	
ZINB	-319.84	3.01	-1005.33	1.84	1.12

3.3 Analysis on safety influencing factors

According to the accident prediction model, when an independent variable increases by one unit while other independent variables remain unchanged, the reduction percentage of accident rate is the accident reduction factor. The negative value of the accident reduction factor stands for the increase of accident when the independent variable increases by one unit and vice versa. Based on the modelling results, the accident reduction factor can be calculated, reflecting the effect of influencing factors on rural roads and non-rural roads in plateau regions. The results are shown in Table 4 (the unit of the oxygen content, the proportion of the large vehicles and other forms of variables is a percentage point). The proportion of the large vehicles, the roadside hazard level and the flat curve density all have a positive effect on the occurrence of the accidents on the non-rural roads. The percentage of large vehicles increased by 1% and the percentage of accident will increase by 4.97%. Hazard level of roadside raises a level, the accident will increase by 62.73%.

The flat curve density in each section increases by one, the percentage of accident will increase by 36.70%, which is less significant than the altitude. On the rural roads, only the access density has a strong positive impact on the percentage of accident. The number of accidents increases by 21.62% for each additional access port. The rest of the factors have a negative impact, the influence of oxygen content has been greatly influenced by the size of villages and towns in the sample section, and the influence is limited in micro accident analysis (The effect of oxygen content is not significant in non-rural roads). The percentage of each car increases by 1%, the corresponding reduction in accidents decreases by 6.16%. As the investigated sections are the state road with the distributing function, and the number of the freight vehicles is stable. The proportion of the large vehicles increases resulting from the decreased mixed rate of the passenger car, so it has a negative influence on the accident. Qinghai 109 state road is mainly two-way two-lane design, and the increase of direction coefficient is interpreted as the increase of the vehicles at the same direction and the decrease of vehicle at the opposite direction, which can help reduce the risk of overtaking.

Table 4 Accident decrease factors data

	Non-rural roads	Rural roads
O2		9.60%
Edif	1.53%	
Ptruck	-4.97%	6.16%
Dir		9.85%
Lrs	-62.73%	
Dhc	-36.70%	
Dac		-21.62%

4 Conclusion

Plateau highway is mainly distributed in Qinghai and Tibet provinces. With the accelerated development of the western region, rapid economic construction in Qinghai-Tibet region and the vigorous development of tourism, the demand for transport has increased year by year. The special climatic and environmental conditions in the plateau have caused the road

traffic safety problem to be solved urgently. The improvements needs the theoretical guidance of correlation analysis urgently. In this paper, investigation and collection of data were carried out on the two-lane highway with the main transport function in the plateau. The traffic accidents in this area were analysed and studied in detail, through the establishment of accident prediction model, the influence factors and the laws of the accident were analysed. The conclusions are as follows:

- (1) There are significant differences in the accident characteristics between non-rural roads and rural roads in the plateau, the 25% of mileage in the rural roads accounted for 67% of the traffic accidents, so the accident rate of per unit mileage was far higher than that of non-rural roads. Roadside accidents in non-rural roads are 30 percent higher than those in rural roads, while the proportion of heavy trucks in the accident model is higher than non-rural roads.
- (2) In the quantitative analysis of single factor, there is a certain positive correlation trend between the altitude and the HVMK of non-rural roads, and it is not significant for the rural roads. With the decrease of oxygen content, the proportion of single motor accidents increases, and the HVMK accident rate is the highest in the non-human-zone mountainous area.
- (3) Through modelling and analysis of indexes, for the non-rural roads, the zero-inflation Poisson distribution model is better than the Poisson, negative binomial distribution and zero-inflation negative binomial distribution. The negative binomial distribution is more suitable for the rural roads according to the accident characteristics.
- (4) In the case of exposed variable for million vehicle kilometres per year, the effect of oxygen content, large vehicles proportion and traffic flow direction coefficient on the number of accidents on rural roadster was negative. The access density had a positive effect, and the effect was the most significant, accidents increased by 21.62% with an access added; the influence of proportion, roadside hazard level, the number of accidents on the flat curve density is positively correlated with the non-rural sections. Moreover, the roadside hazard level has the most significant impact, however it is negative compared to altitude although the effect is small. Due to the limitations of data types and data accuracy, the environment influence factors are not considered in the research. In the future work, more intensive and extensively data will be collected.

4. Acknowledgements

We appreciate the support from the National Science & Technology Pillar Program during the 12th Five-year Plan Period (2014BAG05B02) for this work.

- [1] Traffic Management Bureau of Ministry of public security. Chinese annual report on road traffic accident statistics (2012). Traffic Management Science Research Institute Press of Ministry of public security
- [2] Andrew, V., Joe, B.: Accident Models for Two-Lane Roads: Segments and Intersections, Report No. FHWA-RD-98-133. Federal Highway Administration. 1998: 170-181
- [3] FHWA. Turner-Fairbank Highway Research Center, Geometric Design Laboratory, ITT Industries Inc., Systems Division. IHSDM (Public Test) Version 2.05b. Mar 7, 2003
- [4] Transportation Research Board. NCHRP web Document 62 (project 17-18 (4)): Contractor's Final Report: Development of a Highway Safety Manual. Washington, D.C., 2003
- [5] Liande, ZH., Xiaoduan, S., Yongsheng, C., et al.: Accident Prediction Model of Freeway. Journal of Beijing University of Technology, 2009,07:966-971
- [6] Hailong, G., Weisheng, K., Changcheng, L., et al.: Roadside Accident Prediction Model for Two-lane Highway. Journal of Highway and Transportation Research and Development, 2007, 07: 127-131
- [7] Chengcheng, T., Tiejun, Z., Yong, H.: Basic Model of Safety Performance Prediction for Two-Lane Highways. Journal of Tongji University (Natural Science), 2008,09:1199-1204
- [8] Tiejun, Z., Chengcheng, T.: Research of Roadside Accident Prediction Models for Common Segments of Two-lane Roads in China. Journal of

Highway and Transportation Research and Development, 2008, 04:147-150+158

[9] Zhang, T., Tang, C., Kang, Y.: Safety Characteristics of Two-lane Highway Sections Passing through Towns/villages in Mountainous Area Based on Negative Binomial Prediction Model. Journal of Highway & Transportation Research & Development, 2012, 06: 110-115+132

[10] Michael, G.H.B.: Future Directions in Traffic Signal Control. Transportation Research Part A: Policy and Practice, 1992, 26 (4) : 303-313

[11] Liande, Z .Research on traffic accident prediction model for freeway. PhD thesis, Beijing University of Technology, 2008

[12] Kaitai, F. Some clustering methods for the order sample. Acta Mathematicae Applicatae Sinica, 1982, 5(1)

[13] Abdel-Aty, M., and Pande, A. Classification of real-time traffic speed patterns to predict crashes on the freeways. Presented at the 83rd Annual Meeting of the Transportation Research Board (TRB), Washington D.C., 2004

[14] Abdel-Aty, M., Uddin, N., Abdalla, F., et al.: Predicting freeway crashes based on loop detector data using matched case-control logistic

regression. Presented at the 83rd Annual Meeting of the Transportation Research Board (TRB), Washington D.C., 2004

[15] Hailong, G., Changcheng, L.: Roadside Safety Design Guide. Beijing: China Communication Press, 2008:38-41

[16] Chelbi Engineering Consultants. Guidelines for Safety Audit of Highway. China Communication Press, 2004

[17] Miaou, S. and Lum, H.: Modeling vehicle accident and highway geometric design relationships, Accidents Analysis and Prevention, Vol. 25, No. 6, 1993: 42-51

[18] Venkataraman, S. Milton, C.J., Fred, L.: Manner modeling accident frequencies as zero-altered probability processes: An empirical inquiry, Accident Analysis and Prevention, 1997, 29(10):829-837

[19] Raichle, M.E.: Food for thought: altitude versus normal brain function. Advances in Experimental Medicine & Biology, 1999, 474:171-183

[20] Bonnon, M., Noël, J., Marie, C., et al.: Psychological changes during altitude hypoxia. Aviation Space & Environmental Medicine. 1995, 66(4):330-335

[21] Baoxin, L., Shulin, F.: The characteristics of high altitude traffic and its requirement for cars. Automobile Research & Development, 1995, 06: 56-57+47