Study of Rural Highway Safety Factors in Mountain Area Based on Negative Binomial Model

WENHENG LIU

ABSTRACT

In order to study the influence factors of rural highway safety in mountain area, a field survey was carried out on a rural road in a mountainous area of Guizhou. The negative binomial model was used to analyze the relationship between traffic accidents and road safety. The results show that several significant factors including pavement conditions, delineation, sight distance and the roadside hazards for the effect of rural highway traffic safety.

INTRODUCTION

The rural roads in mountain areas usually have two lanes, no central separation facilities with 20km/h of its design speed limit. Its main function is to connect the administrative area at county level and township level administrative region, as contact Township street village where the branch highway. According to statistics, the road traffic accident at the end of 2015, China's total highway mileage of 457.73 million kilometers, of which the 342.63 million kilometers are rural roads, accounting for 74.85%, the total length of the rural highway in China accounted for a large, but because of China's rural highway construction funds is generally limited, although the highway traffic volume is small, but every year many accidents still occur, only in 2015, the number of accidents occurred in China reached 18547 on the country road, accounting for the total number of accidents of 18.13%[1], combined with the domestic country road traffic safety level, the situation is more severe.

Wenheng Liu, Beijing Jiaotong University, Beijing, China, Tel: 1881032219, Email: 15120946@bjtu.edu.cn
In recent years, with the promotion of new rural construction, rural road pavement and the profile is improved. However, the safety awareness of rural residents, bad motor vehicle conditions and backward security measures have not kept pace with the pace of rural road construction. Therefore, the frequency of accidents caused by rural roads is high and the severity of accidents is serious. To focus on the safety problem of the rural highway in mountainous area, Scholars have obtained many research results, Xu Jinliang, Zhou Yuming, Yang Hongzhi proposes the influence factors of mountainous rural highway roadside safety: the influence are traffic volume and compositions, roadside features and road alignment. Chen Lesheng study on the division of general highway roadside hazard in mountainous area, and there are many effective attempts to deal with the relationship between road accident rate and road safety influence factors. It includes the analysis of the causes of road traffic accidents based on simple statistics and empirical analysis, grey relational analysis, model of highway traffic prediction technology of two lane highway in our ordinary road traffic accidents, and the analysis of the influence factors of freeway traffic accidents based on the negative binomial model. However, the above literatures do little research on the quantitative analysis of the factors affecting the rural road safety in mountain areas. Therefore, this paper takes the rural highway in Guizhou mountain area as an example, and analyzes the safety factors that influence the safety of rural highway in the mountain area based on the collected data.

SAFETY FACTORS OF RURAL ROAD IN MOUNTAIN AREA

To collect the basic data for analysis of mountainous rural highway combined with the investigation results and the characteristics of mountainous rural highway, concluding that the main factors of mountainous rural highway safety are the following aspects: pavement condition, delineation, speed management, sight distance, radius of horizontal curve, and the roadside hazards.

Pavement condition

Affected by the natural conditions and the level of economic development, the pavement condition of the rural highway in mountainous area shows the characteristics of low pavement paving rate, low maintenance rate and serious pavement damage. As shown in Figure 1, the ratio of hardened pavement, cement pavement and the asphalt pavement is 12%, 32%, 56% separately on the studied rural road.
The skid resistance of no hardening of pavement and cement pavement is poor, which is prone to occur roadside accident particularly in the extreme unfavorable weather conditions. Good pavement condition helps the vehicle to run smoothly on the road surface. The hardened asphalt pavement has high skid resistance and helps to reduce vehicle skidding and rear end collision. Good pavement condition also contributes to the painting of all kinds of pavement markings, such as the central line, roadside edge line, deceleration markings, rumble markings etc.

**Delineation**

The delineations of rural highway in mountain areas mainly have the problems of simple setting up of traffic signs, improper setting position and serious lack of traffic markings, to some extent, these problems has laid a hidden danger for the occurrence of rural road traffic accidents. Delineations can provide road information to the driver, allowing the driver to adjust driving state according to the road information; delineations will play an important role in road safety under the condition of poor distance on roads. The rural highway guidance signs can be divided into the central line, the road edge line, the curve guidance sign and the warning sign, etc.

**Speed management**

Speed is an important factor affecting traffic safety, especially in the pavement with good condition; the driver is easy to speeding, so in some key dangerous sections, it is necessary to add speed management measures. As for the speed management measures for rural roads, it is feasible to set the stone pavement, or deceleration rumble marking at the sections of school, village, sharp slope etc. Using bump and method of increase road friction reduce the driving speed, prompting the driver driving gear.

**Sight distance**

Insufficient sight distance and limited forward visibility can adversely affect safety and increases the risk of a collision by reducing reaction times and stopping distances. Adequate sight distance provides drivers with sufficient time to identify and appropriately react to all elements of the road environment, including other road users and hazards. Where visibility is restricted, sight lines should be cleared through the...
removal of obstructions or through road realignment. For example, the removal of vegetation or the cutting of an embankment on the inside of horizontal curves will improve sight distances. Consideration must also be given to the location and height of fences and barriers to ensure they do not obstruct visibility. If sight distance cannot be significantly improved, speed management or restrictions to vehicle movements such as right turns and overtaking may be considered along with appropriate advance warning signs.

Radius of flat curve

The research data at home and abroad show that when the radius of flat curve is less than 400 m, the accident rate increases obviously\(^8\). The mountainous rural highway especially Township Road, Village Road, the curve radius is less than 100 m accounted for more than 3/4, is mainly caused by the vehicle ran off the road traffic accident.

Roadside hazards

The main hazards of mountainous rural highway roadside distribution are shown in Figure 2. From the figure, the mountainous rural highway roadside hazards are metal safety barrier, concrete safety barrier, aggressive vertical face, upwards slope, downwards slope, the tree, the cliff and the sign, post or pole etc. Among them, the risk coefficient is the highest in the roadside cliffs and deep valleys, rivers and lakes, deep, water precipice sections; an effective way to eliminate the sections of water precipice is to install the road safety barrier.

The rural road vehicles are mainly central passenger cars, freight cars and motorcycles, so the safety barrier anti-collision grade is A and B, and the safety barrier type can include class B corrugated beam safety barrier and class A concrete safety barrier.

![Figure 3. Percentage of different roadside hazards on the studied rural road.](image-url)
If the road section is wide enough, the end of the safety barrier shall be in the form of abduction or be hidden in the mountain, if not, warning marks of reflective film with black and yellow meets should be set at the end of the safety barrier, corrugated beam safety barrier and concrete safety barrier shall be connected by transition.

ANALYSIS OF INFLUENCING FACTORS OF RURAL ROAD SAFETY IN MOUNTAIN AREA BASED ON NEGATIVE BINOMIAL MODEL

On the base of analyzing the influence factors of rural highway, conducting depth analysis of the specific impact of these influence factors (pavement condition, delineation, speed management, sight distance, curve radius and roadside hazards) and the safety of the rural highway in mountainous area. The theoretical basis is the accident prediction model based on the negative binomial distribution, and the content of the analysis is the analysis of traffic accidents.

Analysis theory and method

(1) Model theoretical basis
The number of road traffic accidents belongs to discrete counting data, and Poisson regression model is usually used to deal with discrete counting data. Poisson
regression takes the event number \( Y \) as the object of study, and uses \( y \) as the observation value of the event number, assuming that the event follows Poisson distribution, and the random variable \( Y \) is equal to the probability of \( Y \), and the distribution density function is:

\[
P(Y = y_i) = \frac{e^{-\mu} \cdot \mu^{y_i}}{y_i!} \quad (y_i = 1, 2, 3, \ldots)
\]

In Poisson regression, \( E(y_i) = \text{Var}(y_i) = \mu \), That is, the variance of the explanatory variable is equal to the expected value and this value is equal to the parameter of Poisson distribution \( \mu \). Assuming that \( X = (x_1, x_2, \ldots, x_n) \) is the independent variable of \( n \) that affects \( Y \), and the basic form of Poisson regression is as follows:

\[
E(y_i) = \mu = e^{\alpha + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n}
\]

After linearizing the equation, the expression is:

\[
\log[E(y_i)] = \log \mu = \alpha + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n
\]

\( \beta \) is the parameter to be estimated, and it can be estimated by iterative nonlinear weighted least squares or maximum likelihood. If the expected value of the explanatory variable \( Y \) is equal to the variance, the Poisson regression model is suitable and valid, however, in practical applications, the number of traffic accidents often has the characteristics of over discrete, and the variance of the explanatory variable \( Y \) is greater than the mean. If the Poisson regression model is still used in this case, it is possible to underestimate the standard error of the parameter, overestimate its significance level, and eventually lead to unreasonable results. The variance of the number of road accidents is usually far greater than the average, so there will be excessive dispersion. At this point, the negative binomial regression model is more suitable for data fitting than Poisson regression model. The Poisson distribution model improves the negative binomial regression model, and its distribution density function is in the form of:

\[
P(Y = y_i) = \frac{\Gamma(y_i + \alpha^{-1})}{\Gamma(\alpha^{-1})y_i!} \left( \frac{\alpha^{-1}}{\alpha^{-1} + \mu} \right)^{\alpha^{-1}} \left( \frac{\alpha^{-1} + \mu}{\alpha^{-1} + \mu} \right)^{y_i}
\]

In the formula, \( Y \) is a random variable of traffic accident for a given point and a given period of time; \( y \) is the predicted number of traffic accidents occur for given point and a given period of time; \( \mu \) is the average number of accidents predicted for a given place and a given period of time; \( K \) is dispersion coefficient, the bigger the \( K \) value, the greater the discreteness. The parameter estimation of the negative binomial model can be carried out by the maximum likelihood function (MLE), and the form of the likelihood function is:

(2) Modeling verification

At present, there are several methods to assess the advantages and disadvantages of regression models: ① Akaike’s Information Criteria AIC criterion is used to evaluate the quality of a model. Generally, the smaller the AIC value, the better. ② Pseudo \( R^2 \) statistic is used to test the goodness of fit of the model, and the \( R^2 \) value is larger, which shows that the fitting is better. ③ T estimates the significance of the parameter at the 5% level.
Model fitting analysis

Based on the negative binomial model, the safety influence factors of rural highway in mountain area are analyzed, the first is the selection of model form, $E$ is defined as exposure level, which is a combination of traffic volume, section length and time span to characterize the exposure level in a safe environment. The formula of the exposure level is:

$$E = \frac{Q_{ij}L_iT}{10^6}$$

Among them, $Q_{ij}$ is the traffic volume of number i section and number j years, $T$ is statistical time, $T = 365d$, then the expression of time and space analysis model of road traffic accidents is:

$$\lambda_{ij} = \frac{Q_{ij}L_iT}{10^6} \exp \left[ \sum_{k=1}^{n} \alpha_k x_k(i,j) \right]$$

Among them, $\lambda_{ij}$ is the number of traffic accidents of number i section and number j years; $Q_{ij}$ is the traffic volume of number i section and number j year; is the number $k$ independent variable of number i section and number j year. $\alpha_0$ and are $\alpha_k$ the model parameters.

In this paper, the method of fixed length is used to divide the rural road in mountainous area into several sections, each length is 100m, and 1066 basic data are obtained. In this study, the total mileage of rural highway in mountain area is 106.6 km, and the time span of traffic accident data is 3 years, and the total number of accidents is 121.

The safety influence factors listed in the previous chapter are taken as independent variables, and the number of traffic accidents in each section is taken as the dependent variable. The selected independent variable is defined as follows:

Pavement condition. Classification variable, in this paper, the pavement condition is three types of cement concrete, asphalt concrete and unhardened pavement, denoted as RC.

Delineation. Classification variable indicates whether or not there is an induced marking line on the road, denoted as D.

Speed management. Classification variable, in this paper, Speed management indicates whether or not there is speed control measures (such as speed hump, speed reduction markings) on the road surface, denoted as SM.

Sight distance. Classification variable indicates that weather the sight distance is good or poor, denoted as SD.

Roadside hazards. Classification variable indicates that the type of roadside hazard that can cause injury to road users, denoted as RH.

Radius of flat curve. Classification variable indicates the degree of road curvature, denoted as R.

The assignment meaning of an independent variable is shown in Table I.
### TABLE I. MEANING OF VARIABLES AND ASSIGNMENT STATISTICS.

| Variable name | Coding meaning                          | Statistical results |
|---------------|----------------------------------------|---------------------|
| Pavement condition (RC) | 1—Unhardened pavement | 128                 |
|                | 2—Cement pavement                     | 342                 |
|                | 3—Asphalt pavement                    | 597                 |
| Delineation (D) | 1—Poor (no or badly damaged)          | 855                 |
|                | 2—Only markings or signs              | 212                 |
| Speed management (SM) | 1—Not present                      | 939                 |
|                | 2—Present                             | 128                 |
| Sight distance (SD) | 1—Poor                               | 354                 |
|                | 2—Adequate                            | 713                 |
| Roadside hazards (RH) | 1—Safety barrier                   | 84                  |
|                | 2—Aggressive vertical face            | 144                 |
|                | 3—Upwards slope (>15°)                | 126                 |
|                | 4—Downwards slope (>15°)              | 151                 |
|                | 5—Cliff                               | 246                 |
|                | 6—Tree                                | 150                 |
|                | 7—Sign, post or pole                  | 166                 |
| Radius of flat curve (R) | 1—>=1500m                            | 344                 |
|                | 2—1500~700m                           | 381                 |
|                | 3—700m~400m                           | 299                 |
|                | 4—400m~200m                           | 37                  |
|                | 5—20~100m                             | 3                   |

### Model fitting results

The negative binomial model is used to fit the data; all the independent variables are substituted into the model. The regression results show that some variables are not statistically significant and cannot be rejected by the assumption that the coefficient is 0. The regression coefficients of some variables are contrary to common sense, while multicollinearity is found due to too many qualitative indexes. Stepwise regression was used to eliminate multicollinearity, fitting the explanatory variable with the one variable regression of each explanatory variable, and ranking the goodness of fit $R^2$ of each regression equation according to the size order. Then, $R^2$ large explanatory variables are added to the model to estimate, and the t test of the parameter estimation is carried out according to the model estimation results. If the t test is significant, it is retained, otherwise the variable is excluded. Repeat the process until all significant variables are added. The E is used as the exposure degree, and the total number of accidents is taken as the dependent variable, EViews software is used to fit the data, and the results are shown in table II and table III:
TABLE II. MODELING RESULTS OF ALL ACCIDENTS IN NEGATIVE BINOMIAL MODEL.

| Parameter | Coefficient | Standard deviation | The value of Z | The value of P |
|-----------|-------------|--------------------|----------------|---------------|
| Intercept | -2.1750     | 0.5065             | -4.294         | <0.000 1      |
| RC2       | -1.48356    | 0.5304             | -4.025         | <0.000 1      |
| RC3       | -1.6632     | 0.6858             | -4.718         | <0.000 1      |
| D2        | -1.2024     | 0.6489             | -2.856         | 0.005 3       |
| SM2       | -0.6849     | 0.7855             | -1.987         | 0.008 9       |
| SD2       | -0.7433     | 0.4544             | -2.021         | <0.000 1      |
| RH2       | 2.5633      | 0.2443             | 8.934          | <0.000 1      |
| RH3       | 2.2652      | 0.3408             | 7.331          | <0.000 1      |
| RH4       | 1.8856      | 0.5359             | 5.868          | 0.00167       |
| RH5       | 3.5576      | 0.4352             | 12.877         | <0.000 1      |
| RH7       | 1.3247      | 0.7543             | 4.983          | 0.009 3       |
| R²        |             | 0.397909           |                |               |
| AIC       |             | 24.38231           |                |               |
| LL        |             | -1561.188          |                |               |

The advantages and disadvantages between model selection based on AIC statistics and log likelihood as the criterion, by comparing three regression indexes of the distribution of model in table 4. It can be seen that the prediction model of the negative binomial distribution model is better. The fitting of the two models shows that the goodness of fit of the negative binomial regression model is better than that of the Poisson regression model.

The results of negative binomial regression model show that RC, D, SM, SD and RH were the significant influencing factors in the model. According to the parameter estimation results, the regression equation is as follows:

$$\lambda ij = \exp(-1.3927 - 1.2164RC2 - 1.8144RC3 - 0.9032D2 - 0.6465SD2 + 2.6978RH2 + 2.0945RH3 + 2.1432RH4 + 3.8437RH5 + 1.4525RH7)$$
The estimation results show that in the property of pavement condition, cement pavement (Coefficient=-1.2164) and asphalt pavement (Coefficient=-1.8144) show less traffic accidents than unhardened pavement. In the attributes of delineation, the adequate delineation (Coefficient=-0.9032) show less traffic accidents than poor delineation. In sight distance attributes, good sight distance (Coefficient=-0.6465) show less traffic accidents than poor delineation. In the attribute of the roadside hazards, aggressive vertical face (Coefficient=2.6978), Upwards slope (>15°) (Coefficient=2.0945), Downwards slope (>15°) (Coefficient=2.1432), Cliff (Coefficient=3.8437), Sign, post or pole (Coefficient=1.4525) show more traffic accidents than safety barriers. Among them, the traffic accidents in the cliffs sections increased most obviously.

The above findings are consistent with safety knowledge. Therefore, in the process of improving the safety of mountainous rural highway, we should focus on the roadside hazards especially the cliff sections for improvement, followed by the improvement of pavement conditions.

CONCLUSION

This paper established the forecast model of mountainous rural highway accident based on negative binomial model. Select the number of accidents as the dependent variable, select parameters affecting accident (pavement condition, delineation, speed management, sight distance, radius of horizontal curve and the roadside hazards) as independent variables. The analysis found that the pavement condition, delineation, sight distance, roadside hazards are significant variables of the model.

Based on the accident prediction model, this paper analyzes the relationship between traffic accidents and safety factors in rural highway in mountainous area. However, due to the limited data collection, weather conditions, population, land use and other macro factors have not been taken into consideration, therefore, a comprehensive analysis of rural road traffic accidents in the mountain areas is still to be further explored.

REFERENCES

1. Ministry of Public Security. (2015). Statistical Report of Road Accident of China.
2. Zhu Hanrong, Liu Tangzhi. Analysis and improvement of traffic safety problems in rural highway [J]. Highway traffic technology, 2009 (4): 139-144.
3. Xu Jinliang, Zhou Yuming, Yang Hongzhi. Roadside safeguard technique of rural highway in mountainous area [J]. Journal of Chang'an University (Natural Science Edition), 2011, (03): 1-6.
4. Chen Lesheng, Li Yongjiang, Yu Hong, Bai Shufeng. The mountain highway roadside hazard degree classification method of highway [J], 2005, (11): 164-168.
5. Pan Zhaoyu, Guo Xiucheng, Sheng Yugang, Pan Minrong, Lu Guangming. Cause analysis on the formation of the highway traffic black-spot with the gray association degree method [J]. Journal of traffic and transportation engineering and information, 2008, (03): 96-101.
6. Zhang Tiejun, Tang Chengcheng. Model of Highway traffic prediction technology of two lane highway in our ordinary road traffic accidents [J], 2008 (4): 147-151.
7. Xia Pingping, Wei Liying, Ma Zhuanglin. Negative binomial regression model based analysis for influential factors of expressway traffic accidents [J]. Shandong science, 2013, (04): 79-84.
8. Jiang JinLong. Study on the safety design of rural road [D]. Xi'an: Chang'an University, 2009.