Periodica Polytechnica
Transportation Engineering

47(4), pp. 318-328, 2019
https://doi.org/10.3311/PPtr.12060

RESEARCH ARTICLE

Relationship between Speed Characteristics and Traffic Safety on Freeways in Mainland China

Ziyue Tang[1*]

Received 07 February 2018; accepted 25 April 2018

Abstract

Based on the traffic accidents statistical data of 10 typical freeways in mainland China, by using of some kinds of regression model, the influences of the average vehicle speed and the speed standard deviation on the traffic safety are studied. According to the regression results, the accidents show an increasing trend with the increase of the vehicle average speed and the speed standard deviation. On this basis, in view of the regression results, the strategy is put forward for controlling the vehicle average speed and the speed standard deviation, which has important theoretical and practical significance for improving highway safety. After a comprehensive comparison among these regression methods, it is found that the nonlinear regression method of user-defined model expression has the best fitting effect, and it can also more accurately describe the objective reality. It has high practicality and popularized value.

Keywords

multivariate nonlinear regression model, average vehicle speed, speed standard deviation, traffic safety, control strategy

[1] School of Resources and Safety Engineering,
Central South University,
Changsha Hunan, 410083, China
[*] Corresponding author: Ziyue Tang, e-mail: 1074106649@qq.com

1 Introduction

In recent years, the average annual growth rate of China's highways has been nearly 20 %, which has covered all provinces, autonomous regions and municipalities of China (Integrated planning division in Ministry of Transport of China, 2017). The total operation length reaches 131,000 kilometers, ranking the first in the world (Integrated planning division in Ministry of Transport of China, 2017)), which plays an extreme important role in China's transportation system. However, according to rmjtxw.com (Tian, 2017), the number of people killed in road traffic accidents in China 2016 was about 40,824, an increase of 4,646 from 2015, and the number of traffic fatalities is also the highest in the world (State Statistics Bureau of China, 2017). The overall level of China's highway traffic safety has obvious gap with the developed country. Many security problems are worth greatly concerning, such as the massive highway deaths, high traffic growth, and the accidents keep increasing year by year.

There are many factors affecting traffic safety, such as vehicle structure and performance, traffic environment, road conditions, driver situation, natural conditions and so on. However, the improper speed is the key in traffic accidents (Jomaa et al., 2016). The most familiar parameter of a distribution is the average (or mean). However, it is not the only parameter in application and research. In road design, the practice over the last two decades has been for engineers to use what is known as 85th percentile speed (v_{85}) of traffic. The 85th percentile is the speed at or below which 85 percent of drivers drive, and is thus a measure of the higher speed end of the distribution on a particular road. The amount by which the 85th percentile exceeds the mean speed will depend on the spread of the speed distribution. The most common measure used to characterize the spread or variability of speeds found on any road is the standard deviation of the speed distribution. The coefficient of variation is then the ratio of the standard deviation to the mean (Taylor et al., 2000). In April 2017, the all-level highway management departments of Shandong province had investigated a total of 519,000 illegal acts, of which 2.75 million belonged to over speed limit, accounting for 52.987 percent (Chen, 2016). Transport Research

Laboratory (TRL) has undertaken a major program of research for the Department of Environment, Transport and the Regions (DETR) to investigate the impact of traffic speed on the frequency of road accidents. This work found a positive relationship between speed and injury accidents-the higher the speed, the more accidents-indicating that a 5 percent change in accidents was associated with a 1.609344 km/h change in average speed (Taylor et al., 2000). According to Solomon, the severity of accidents will increase rapidly as speed over 96 km/h. And the likelihood of fatal casualties will increase more rapidly at speeds above 112 km/h (Solomon, 1964).

In view of the complex relationship between the speed characteristics and accident rate, many studies at home and abroad have different views, which can be divided into three types:

1. The traffic accident rate is related to the vehicle speed, such as average speed and v_{85}. Australian authorities believed that accident rate was up 2 times as a 5 km/h was added to the speed, at speeds above 60 km/h, and also the severity of the accident was exponentially growing (Federal Highway Administration, 1998). According to Canada Liu and Popoff's (Liu and Popoff, 1997) long-term analysis on the Saskatchewan expressway, a 1 km/h reduction in average speed should reduce traffic accident casualty rate by 7 percent. By means of survey and statistics analysis on highways in China, Chongqing Jiaotong University drew a conclusion that the accident rate will increase by 2.221 percent when the speed is increased from 100 to 101 km/h (Hu, 2008).
2. The accident rate is related to dispersion of speed distribution (such as variance, $v_{85} - v_{15}$ (v_{15} is 15th percentile speed), the speed difference between large vehicle and car, the coefficient of variation, and speed gradient - the difference between the speed of the section and the average speed, etc.). High speed dispersion will cause frequent vehicle overtaking and overtaken phenomenon, which can lead accidents. Garber and Gadiraju (1989) believe that the accident rate will increase with the increase of speed variance, and it is not necessarily associated with the average speed. Solomon (1964) attributed the accident incentive to the speed gradient. Based on the traffic observation of four-lane highway, the U - shaped curve relationship between traffic accident rate and speed gradient was obtained. When the speed gradient was 5.544 km/h, the accident rate was at the bottom of the curve, 162.742 times per 100,000 km. And with the change of speed gradient (whether increase or decrease), the accident rate was increasing. The National Crash Severity Study (NCSS) argues that traffic accident death rate is proportional to the fourth power of the vehicle speed gradient (Du, 2002), and when the speed gradient reaches up to 114.24 km/h, it has a 100 percent chance of death accident.
3. Both the vehicle speed and the dispersion of vehicle speed are related to accident rate. The EURO model, developed by A. Buruya in the British TRL, showed that the ratio of speed to average speed affected accident rate directly. At the average speed of 60 km/h, a 1 km/h reduction in the speed gradient should reduce traffic accident rate by 2.56 percent (Baruya et al., 1999). According to Chinese scholars, the average speed plus speed standard deviation is approximately equal to v_{85}, while v_{85} is in the speed range of the lowest the accident rate. If the vehicle is traveling at a speed of 2 times the average speed, the accident rate will be significantly increased (Department of Transport China, 2004). Existing literature research indicates that there is obvious nonlinear correlation between the accident rate and vehicle speed characteristics (Lidbe et al., 2017). Owing to the speed of the large vehicle far below that of the car on the freeway, the vehicle speed and its dispersion have distinct uniqueness. With the addition of the freeway traffic environment, the performance of vehicles, vehicle, the driver's driving habits, as well as the method of accident statistics in China are quite different from other countries, it is necessary to study the traffic safety of freeways in China.

According to the geographical differences, China can be divided into 8 regions, including east China, south China, central China, north China, northwest China, southwest China, northeast China and the region of Taiwan, Hong Kong, and Macao. In this paper, 10 typical and representative freeways were selected which distribute in 7 regions in mainland China. Taking the 10 freeways data as the sample, the influence of the average vehicle speed and the speed standard deviation on the traffic safety were researched. The research main feature is that some kinds of regression models are established, including the nonlinear regression of user-defined model expression, which is an important method of solving complex multivariable problem. However, in practice, single independent variable parameter expression is often used, and few researches on the multiple independent variable parameter expression have been made in published literatures.

In this paper, a numerical analysis and comparison of scatter plots are used to pre-estimate the nearest mathematical model, and by using of many times trial regression, the regression and coefficient fitting of mathematical models are finished. The method can effectively build relation model among vehicle average speed, speed standard deviation and the accident rate. It can reach high fitting precision and can provide a powerful reference for exploring the relationship between the characteristics of speed and traffic safety on freeways in China.

2 The relationship between vehicle speed characteristics and traffic safety

2.1 The influence of the vehicle speed characteristics on traffic safety

The vehicle speed represents the actual traffic running situation on a certain road. It is related to the driver's technical level, driver's expectations, climate conditions and terrain, freeway alignment, traffic flow, traffic composition, traffic management facilities and so on. The traffic accidents occurrence is the process of energy conversion. The higher the speed, the greater the energy conversion is, the more serious the traffic accident consequence is, and the greater the likelihood of casualties is. 9 in 10 accidents are due to the high speed. About 30 percent of fatalities are related to over-speeding. When the speed is more than 80.4672 km/h, a 16.0934 km/h increasing in the speed should cause a double collision force, which can dramatically increases fatal traffic accident rate (Chengdu Business Daily, 2014). When the speed reaches 160, no matter which car is driving and the seat belt is fastened or not, the death rate of the crew member was 100 percent. According to Kloeden et al. (2001) study on rural road data in Adelaide city, Australia, it showed that lowering the average speed was more helpful to reduce the risk of accidents than reducing the speed gradient. The vehicle speed characteristics directly impact on traffic safety. Therefore, it is necessary to study the relationship between speed characteristics and traffic safety, as well as to establish the relationship model between them.

2.2 The nonlinear regression analysis method and its principle

Regression analysis is a statistical method to process interdependent relationship between two or more variables. The mathematical expression (regression function, or called "regression equation") among variables can describe quantitatively their relationships. According to regression function and the number of independent variables, the regression can be divided into linear regression and nonlinear regression, simple (univariate) regression and multivariate regression. Due to particular significance of the regression function, the regression analysis mainly studies the problem of regression function statistic inference by using of observed value of the independent variable and dependent variable, including estimate and test of hypothesis. For complex things, a phenomenon is often associated with multiple factors, so multivariate nonlinear regression usually is more effective and more practical to predict or deduce tend.

Due to the complexity of natural law, the vast majority of research objects in the real world usually are nonlinear systems. So it makes the nonlinear regression model has become more and more important in the statistics analysis. The nonlinear model is the generalization of linear model.

The general form of nonlinear regression model as follows:

$$Y=\varphi(x_1,x_2,\cdots,x_m,\beta_1,\beta_2,\cdots,\beta_r)+\varepsilon. \tag{1}$$

For a given set of observations (x_i, y_i), i = 1, 2, ..., n, the Eq. (1) can be rewritten as

$$y_i=g(x_i,\theta)+\varepsilon_i. \tag{2}$$

In Eq. (1), the non-random vector $x_i = (x_{i1}, x_{i2}, ..., x_{ik})^T$ is the independent variable, y_i is the dependent variable. g is a regression equation, and its structure is usually a quadratic or cubic polynomial form. $\theta = (\theta_0, \theta_1, ..., \theta_p)^T$ is the undetermined parameter vector. ε_i is the random error and it satisfies the independent identically distributed assumption, that is

$$\begin{cases} E(\varepsilon_i)=0 \\ \operatorname{cov}(\varepsilon_i,\varepsilon_j)=\begin{cases}\sigma^2, i=j \\ 0, i\neq j\end{cases} \end{cases} \quad i,j=1,2,\cdots,n\ . \tag{3}$$

In Eq. (3), $\varepsilon_i \sim N(0, \sigma^2)$, the mean of ε_i is 0 and its variance is σ^2.

For Eq. (2), according to the known data pairs, based on least square method, the estimated value of undetermined parameter vector θ can be solved and determined by making the residual sum of squares up to the minimum. θ's estimation is $\hat{\theta}$. The residual sum of squares is:

$$f(\theta)=\sum_{i=1}^{n}(y_i-g(x_i,\theta))^2. \tag{4}$$

The $\hat{\theta}$ is called as nonlinear least squares estimation of θ, who can minimize the total residual sum of square. When the function $f(\theta)$ is assumed to be continuous and differentiable for θ, the partial derivative of function $f(\theta)$ of parameter θ_j (j = 0, 1, ..., p) in the vector θ can be set and its value is set as zero. And the equations are established, and $\hat{\theta}$ of the equations can be solved. The p + 1 equations f solve $\hat{\theta}$ as the follows:

$$\frac{\partial f}{\partial\theta_j}\Big|_{\theta_j=\hat{\theta}_j}=-2\sum_{i=1}^{n}\left(y_i-g\left(x_i,\hat{\theta}\right)\right)\frac{\partial g}{\partial\theta_j}\Big|_{\substack{\theta_j=\hat{\theta}_j \\ (j=0,1,2,\cdots,p)}}=0. \tag{5}$$

The solution of Eq. (5) is the values of θ's the least square estimation $\hat{\theta}$.

In addition, the gradient descent method can be used to solve the problem as the follows:

$$\theta^{(i+1)}=\theta^{(i)}-\alpha\frac{\partial f}{\partial\theta},\ (\alpha>0). \tag{6}$$

In Eq. (6), $\frac{\partial f}{\partial\theta}=-2\sum_{i=1}^{n}\left(y^{(i)}-g_\theta\left(x^{(i)}\right)\right)x^{(i)}$, to iterate and solve, the solution of θ is optimal.

Using software IBM SPSS Statistics to solve the nonlinear algebraic equations, the initial values of parameters must be given in advance. If the initial value is improperly selected, it can cause the non-convergence issue, and the nonlinear regression comes to a stop. The nonlinear regression module in software IBM SPSS Statistics can also obtain the approximate 95 percent confidence interval and approximate standard deviation of regression parameters.

2.3 Basic data

In this paper, it is purposeful to choose the 10 freeways as sample, including the Shanghai-Nanjing freeway in the east China, the Guangzhou-Foshan freeway in south China, Badaling freeway, Shijiazhuang-Taiyuan freeway and Beijing-Tianjin-Tanggu freeway in north China, Xihan freeway and Xi'an-Xianyang airport dedicated freeway in northwest China, Chengdu-Chongqing freeway and Suining-Chongqing freeway in southwest China, Shenyang-Dalian freeway in northeast China. The 10 freeways are distributed in 7 large regions of China. They are typical and representative.

In Table 1, the data comes from those literatures (Pei and Cheng, 2004; He et al., 2010; Chen et al., 2009; Zou et al., 2011; Li et al., 2010; Shao, 2004; Yang et al., 2009). Chengdu-Chongqing freeway is the traffic artery between Chengdu and Chongqing. The freeway passes through the heart of Sichuan basin, with a total length of 337.5 km. Shijiazhuang-Taiyuan (Hebei section) is an important channel to transport the coal from Shanxi to the east China, there is more than 40 km road located in curved steep slope of mountains, the safety vision of driver is restricted at night, and it is easy to occur traffic accidents. Guangzhou-Foshan freeway east starts from suburb Hengsha of Guangzhou, where is connecting north section of Guangzhou ring expressway, and west ends in Xiebian of Foshan, where is connecting the Fokai expressway. The Shanghai-Nanjing freeway starts from the Zhenru of Shanghai, and ends in the eastern suburb Maqun of Nanjing, with a total length of 274 km. Shenyang-Dalian freeway is the transportation artery of northeast China. It is located in the Liaodong peninsula, and spans Shenyang, Liaoyang, Anshan, Haicheng, Yingkou and Dalian. The Beijing-Tianjin-Tanggu freeway is a way connecting the three places, with a total length of 142.69 km. According to the administrative areas, it includes 35 km in Beijing city, 6.84 km in Hebei province and 100.85 km in Tianjin city. Chongqing section of Suining-Chongqing freeway, stars from the main city Shapingba district, ends in town Shuangjiang of county Tongnan. Xihan freeway is an important part of the main skeleton road with "*" type in Shanxi province. It north starts from Laoyukou in the county Huxian, and south ends to Yuandun in the county Mianxian. The Badaling freeway, south starts from Madianqiao of Haidian district in Beijing, north ends to the Great Wall Badaling of county Yanqing in Beijing, with the total length of 69.98 km. Xi'an-Xianyang airport special highway starts from Zhuhonglu where located in the intersection of urban expressway and expressway around the Xi'an city, ends in the east approach road of Xi'an airport. The traffic accident statistics of the 10 typical freeways are shown in Table 1.

According to the existing literature, the traffic accident rate is significantly associated with vehicle speed characteristics. In this paper, the statistical data of the 10 typical freeways will be taken as the foundation. Considering the accident rate per 100 million vehicle kilometers less affected by other factors, which has good comparability to different sections, the accident rate per 100 million vehicle kilometers was selected as evaluation index of traffic safety. The correlation and regression analysis on the accidents rate and speed were done. And the influence of the average speed and the speed standard deviation on the traffic safety was researched.

Table 1 Statistical data of freeways speed and accident

Name of Freeways	Average speed (km/h)	Speed standard deviation (km/h)	Number of accidents (times per year)	Traffic volume (vehicle per year)	The mile-age (km)	The accident rate (times per 100 million vehicle kilometers)
Chengdu-Chongqing (Chongqing section)	87.61	17.16	206	7708800	114	23.441
Shijiazhuang-Taiyuan	71	20.32	244	3972470	213.4	28.783
Guangzhou-Foshan	58.13	13.01	145	42223200	16	21.463
Shanghai-Nanjing (Shanghai section)	79.86	14.22	194	12511608	74.08	20.931
Shenyang - Dalian	79.5	12.73	887	12334480	375	19.177
Beijing- Tianjin-Tanggu (Beijing section)	88.7	22.57	140	12859680	35	31.105
Suining-Chongqing	89.12	16.31	561	16790000	111.8	29.886
Xihan	72.8	19.3	227	932675	360	32.626
Badaling freeway (located in 50-55km of Beijing direction)	68	12.9	37	57195500	5	12.938
Xi'an-Xianyang airport freeway	108	18.8	55	6651750	20.58	40.1774

3 Regression model and Curve estimation

In the real world, there are some phenomena in which the relationship types among variables can be determined based on experience and expertise. However, there are also some unknown relationship phenomena. Only the data scatterplot can indicate that some mathematical model may be close to the relationship. In this case, generally, by using of numerical analysis and scatterplot comparative method, the most closed mathematical model can be pre-estimated. Then by means of statistical analysis method, the regression and coefficient fitting of mathematical model can be done. It is complex that the relationship among speed characteristics and traffic safety of freeway in China. And the type of relationship among variables is not enough clear. In this paper, by using of the custom secondary development function provided by software IBM SPSS Statistics, through the numerical analysis and scatterplot comparative method, forecasting the user-defined fitting function, combining multiple times trial regression algorithm and least squares fitting tools, the relationship model can be determined.

According to the data in Table 1, the scatterplot is gotten with the accident rate per 100 million vehicle kilometres versus the average travel speed, as shown in Fig.1.

Seeing the scatter trend from Fig. 1, the accident rate per 100 million vehicle kilometres basically shows a trend of increase with the increase of average speed. Curves were estimated by using of the relation models, such as linear model, logarithmic model, inverse model, quadratic model, cubic model, compound model, power model, S model, growth model, exponential model and logistic model. Model summary and parameter estimates are shown in Table 2.

All specific relation models are shown as follows.

Linear model:

$$Accident\ Rate\ Per\ 100\ Million\ Vehicle\ Kilometers = 0.367*(speed) - 3.420. \quad (7)$$

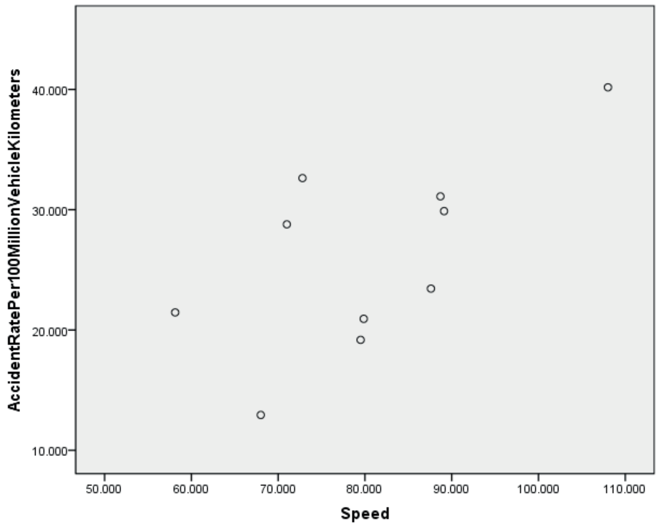

Fig. 1 Scatterplot

Logarithmic model:

$$Accident\ Rate\ Per\ 100\ Million\ Vehicle\ Kilometers = 28.175*\ \ln(speed) - 97.124. \quad (8)$$

Inverse model:

$$Accident\ Rate\ Per\ 100\ Million\ Vehicle\ Kilometers = -2063.428/(speed) + 52.464. \quad (9)$$

Quadratic model:

$$Accident\ Rate\ Per\ 100\ Million\ Vehicle\ Kilometers = 0.008*(speed)^2 - 1.030*(speed) + 52.917. \quad (10)$$

Cubic model:

$$Accident\ Rate\ Per\ 100\ Million\ Vehicle\ Kilometers = (5.177E-5)*(speed)^3 - 0.004*(speed)^2 + 25.970. \quad (11)$$

Table 2 Model summary and parameter estimates

Dependent Variable: Accident Rate Per 100 Million Vehicle Kilometers									
Eq.	Model Summary					Parameter Estimates			
	R Square	F	df1	df2	Sig.	Constant	b1	b2	b3
Linear	0.420	5.787	1	8	0.043	-3.420	0.367		
Logarithmic	0.382	4.941	1	8	0.057	-97.124	28.175		
Inverse	0.339	4.111	1	8	0.077	52.464	-2063.428		
Quadratic	0.486	3.306	2	7	0.098	52.917	-1.030	0.008	
Cubic	0.488	3.332	2	7	0.096	25.970	0.000	-0.004	5.177E-5
Compound	0.358	4.464	1	8	0.068	8.105	1.014		
Power	0.332	3.977	1	8	0.081	0.218	1.083		
S	0.300	3.427	1	8	0.101	4.238	-79.969		
Growth	0.358	4.464	1	8	0.068	2.092	0.014		
Exponential	0.358	4.464	1	8	0.068	8.105	0.014		
Logistic	0.392	5.151	1	8	0.053	0.147	0.979		
The independent variable is Speed.									

Compound model:

$$Accident\ Rate\ Per\ 100\ Million\ Vehicle\ Kilometers = 8.105*(speed)^{1.014}. \quad (12)$$

Power model:

$$Accident\ Rate\ Per\ 100\ Million\ Vehicle\ Kilometers = 0.218*(speed)^{1.083}. \quad (13)$$

S model:

$$Accident\ Rate\ Per\ 100\ Million\ Vehicle\ Kilometers = \exp\left(-79.969/(speed)+4.238\right). \quad (14)$$

Growth model:

$$Accident\ Rate\ Per\ 100\ Million\ Vehicle\ Kilometers = \exp\left(0.014*(speed)+2.092\right). \quad (15)$$

Exponential model:

$$Accident\ Rate\ Per\ 100\ Million\ Vehicle\ Kilometers = 8.105*\exp\left(0.014*(speed)\right). \quad (16)$$

Logistic model:

$$Accident\ Rate\ Per\ 100\ Million\ Vehicle\ Kilometers = 1/\left(1/80+0.147*(speed)^{0.979}\right). \quad (17)$$

The results are analyzed as follows:

1. From the perspective of goodness-of-fit, seeing the parameter evaluation result in the Table 2, the coefficient of determination R square in each model is between 0.300 and 0.488, which is too small (under 0.5). It indicates that the regression curve is not enough proximate to the scattered points. The models can not reflect the relationship between the accident rate and the vehicle speed well. It is necessary to optimize the models or adopt other regression models as well as consider the interaction between input variables.
2. From the perspective of F-test, referring to the critical value table of F-test, they can be known that $F_{0.05(1,8)} = 5.32$ and $F_{0.05(2,7)} = 4.74$. And the values F of all model except for the linear model are too small (3.306 ~ 5.787). In those models of $df_1 = 1$ and $df_2 = 8$, except that the linear model can satisfy $F = 5.787 > F_{0.05(1,8)} = 5.32$, while the values of F in the other models are all less than the critical value ($F_{0.05(1,8)} = 5.32$). And in those models of $df_1 = 2$ and $df_2 = 7$, the values F of the quadratic model and cubic model are all less than the critical value ($F_{0.05(2,7)} = 4.74$). It shows that all models except for the linear model believe that the data have not significant difference at the given significance level $\alpha = 0.05$. I. e., the average speed, used as an independent variable, has no significant explanatory power on the accident rate per 100 million vehicle kilometres. The overall regressions equations are meaningless.
3. From the view of the statistical significance of F-test, the P value of each model is between 0.043 and 0.101. Except for in the linear model $P = 0.043 < 0.05$, the P values in other models are between 0.053 and 0.101, which all are bigger than α ($\alpha = 0.05$). It illustrates that there is no statistical support (no significant difference) between the speed and the accident rate in other models. The models should be changed for regression.
4. From the point of the fitting effect, as shown in Fig. 2, the fitting curve is far from the observed value.

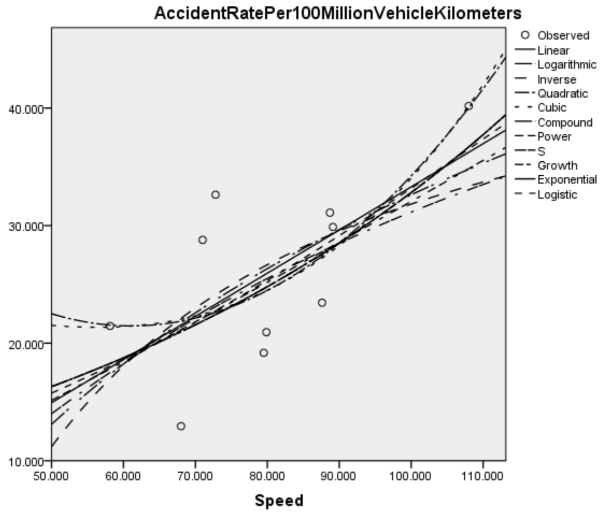

Fig. 2 Curve estimation of each model

In summary, only the regression result of the linear model is not too bad, and those of other models are not ideal.

4 The nonlinear regression of user-defined model expression

4.1 Nonlinear regression analysis on speed - accident rate

The above regression can't meet the needs of the fitting and regression effects by the built-in models of software IBM SPSS Statistics. In order to further improve the precision of fitting and regression effect, in this article, the user-defined model expression was used which reserved by the IBM SPSS Statistics, and the nonlinear regression with special function was implemented.

There is complex relationship among the speed characteristics and traffic safety of highway in China. User is difficult to define the expression of relationship. In this paper, through the some known scattered data of actual measurement, according to the shape and trend of the graph, combining the numerical analysis and experience, the user-defined model expression can be pre-estimated. Through multiple trial regression, the closest relation model can be built.

This process of curve regression can be finished in software IBM SPSS Statistics. In order to achieve a better fitting effect, it usually needs to be repeated and amended many times. In the user-defined model expression of nonlinear regression, the speed was selected as the independent variables which had a big influence on traffic safety.

According to the observation and analysis in Fig. 2, it shows that the rising tendency of dependent variable is slightly faster than the quadratic model with the increase of independent variable. Therefore, based on the quadratic model, adding logarithm of the cubic, via multiple trial regression, the user-defined model expression was constructed, and shown as follow:

$$\begin{aligned} &Accident\ Rate\ Per\ 100\ Million\ Vehicle\ Kilometers \\ &= b1*\ln(b2*(speed)*(speed)*(speed) \\ &+b3*(speed)*(speed)+b4*(speed)) \\ &+b5*(speed)*(speed)+b6*(speed)+b7. \end{aligned} \tag{18}$$

The initial value of the parameter is given as: $b1 = 0.0269$, $b2 = 0.008$, $b3 = -0.761$, $b4 = 4.787$, $b5 = 0.0034$, $b6 = 0.0007$, $b7 = 2.2877$. The nonlinear regression option is set as: Levenberg-Marquardt. The iteration history of nonlinear regression is shown in Table 3. The parameter estimates and ANOVA are shown in Table 4 and Table 5.

According to Table 3, the iteration was stopped after the 34th model evaluation and the 12th derivative evaluation. The residual sum of squares was 8.064 at the 1st iteration. In the subsequent iteration, the residual sum of squares increases or decreases. Parameters b1 to b7 also constantly changed, the last iteration terminated when stopping criterion for iteration is satisfied. The parameter values in Table 4 and Table 5 came from the iteration number 12.0 in Table 3, where the residual sum of squares reached the minimum 7.915. According to Table 4, the user-defined model expression is as follows:

$$\begin{aligned} &Accident\ Rate\ Per\ 100\ Million\ Vehicle\ Kilometers \\ &= 0.025*\ln(0.108*(speed)^3 \\ &-0.759*(speed)^2+4.950*(speed)) \\ &+0.003*(speed)^2+0.001*(speed)+2.289. \end{aligned} \tag{19}$$

The calculation results are analyzed as follows:

1. The standard error of each parameter estimate is between 0 and 0.471. The standard error is small. It shows that the parameters estimations values have high confidence in the 95 percent confidence interval. The obtained confidence interval is narrower, indicating that the parameter estimation has a better sensitivity, and the fitting effect is ideal.
2. According to Table 5, $R^2 = 0.944$ in the model. It shows that the regression equation can explain the 94.4 percent variation of the dependent variable. It has a good fitting degree.
3. As Fig. 3, the observation value and regression curve were drawn by using software Matlab. It can be seen that the regression equation has a high goodness-of-fit.

According to Fig. 3, as the speed increases, the accident rate has a growing trend. Based on Eq. (19), when vehicle speed is 100 km/h, accident rate per 100 million vehicle kilometers is 32.5592. If vehicle speed reaches 120 km/h, the accident rate will rise to 45.8129, i.e., it will increase by 40.71 percent. It is thus clear that a lower average speed has a lower accident rate.

Table 3 Iteration history [b]

Iteration Number [a]	Residual Sum of Squares	Parameter						
		b1	b2	b3	b4	b5	b6	b7
1.0	8.064	0.027	0.008	-0.761	4.787	0.003	0.001	2.288
1.1	315.847	0.027	0.011	-0.761	4.787	0.003	0.041	2.288
…	…	…	…	…	…	…	…	…
12.0	7.915	0.025	0.008	-0.759	4.950	0.003	0.001	2.289
12.1	31.406	0.025	0.008	-0.759	4.950	0.003	0.001	2.289

Derivatives are calculated numerically.

a. Major iteration number is displayed to the left of the decimal, and minor iteration number is to the right of the decimal.

Table 4 Parameter estimates

Parameter	Estimate	Std. Error	95 % Confidence Interval	
			Lower Bound	Upper Bound
b1	0.025	0.000	0.025	0.025
b2	0.008	0.031	-0.092	0.108
b3	-0.759	0.000	-0.759	-0.759
b4	4.950	0.000	4.950	4.950
b5	0.003	0.004	-0.010	0.017
b6	0.001	0.471	-1.498	1.499
b7	2.289	0.000	2.289	2.289

Table 5 ANOVA[a]

Source	Sum of Squares	df	Mean Squares
Regression	4016.483	7	573.783
Residual	7.915	3	2.638
Uncorrected Total	4024.398	10	
Corrected Total	142.522	9	

Dependent variable: Accident Rate Per 100 Million Vehicle Kilometers

a. R squared = 1 - (Residual Sum of Squares) / (Corrected Sum of Squares) = 0.944.

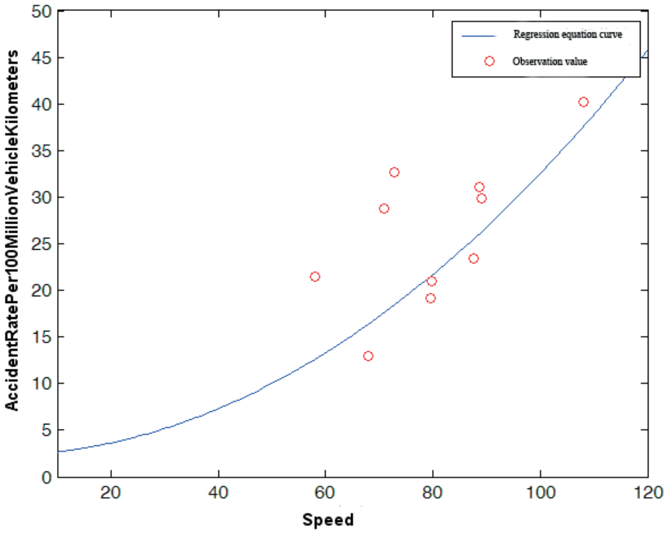

Fig. 3 The vehicle speed-accident rate regression curve of user-defined model expression

4.2 Nonlinear regression analysis on speed standard deviation-accident rate

In the same way, to analyse the nonlinear regression on speed standard deviation-accident rate, the user-defined model expression as follows:

$$\begin{aligned}&Accident\ Rate\ Per\ 100\ Million\ Vehicle\ Kilometers\\&= b1 * \ln(b2 * (Speed\ Standard\ Deviation)\\&* (Speed\ Standard\ Deviation) * (Speed\ Standard\ Deviation)\\&+ b3 * (Speed\ Standard\ Deviation)\\&* (Speed\ Standard\ Deviation)\\&+ b4 * (Speed\ Standard\ Deviation))\\&+ b5 * (Speed\ Standard\ Deviation)\\&* (Speed\ Standard\ Deviation)\\&+ b6 * (Speed\ Standard\ Deviation) + b7.\end{aligned} \tag{20}$$

The parameters initial values are given as follows: $b1 = 0.1069$, $b2 = 0.05$, $b3 = -0.05$, $b4 = 4.5$, $b5 = 0.0654$, $b6 = 0.005$, $b7 = 7.3$. Constraints are defined as follows: $0 \le b4 \le 4$, $b3 \ge 0.2$, $b7 \ge 3$, and $0.1 \le b1 \le 0.2$. The option of nonlinear regression is set as sequential quadratic programming. The parameter estimates are shown in Table 6 and ANOVA are shown in Table 7.

According to Table 6, the user-defined model expression is as follows:

$$\begin{aligned}&Accident\ Rate\ Per\ 100\ Million\ Vehicle\ Kilometers\\&= 0.200 * \ln(-3.184 * (Speed\ Standard\ Deviation)^3\\&+ 71.858 * (Speed\ Standard\ Deviation)^2\\&+ 0.317 * (Speed\ Standard\ Deviation))\\&+ 0.029 * (Speed\ Standard\ Deviation)^2\\&+ 0.801 * (Speed\ Standard\ Deviation) + 3.000.\end{aligned} \tag{21}$$

Table 6 Parameter estimates

Parameter	Estimate	Std. Error	95 % Confidence Interval	
			Lower Bound	Upper Bound
b1	0.200	2312.292	-7358.545	7358.945
b2	-3.184	92566049.425	-294586485.170	294586478.801
b3	71.858	2089208857.803	-6648794938.623	6648795082.338
b4	0.317	35586.596	-113252.113	113252.747
b5	0.029	25.715	-81.807	81.866
b6	0.801	855.884	-2723.005	2724.608
b7	3.000	5807593.876	-18482352.673	18482358.674

Table 7 ANOVAa

Source	Sum of Squares	df	Mean Squares
Regression	7141.769	7	1020.253
Residual	209.462	3	69.821
Uncorrected Total	7351.231	10	
Corrected Total	563.779	9	

Dependent variable: Accident Rate Per 100 Million Vehicle Kilometers

a. R squared = 1 - (Residual Sum of Squares) / (Corrected Sum of Squares) = 0.628.

The calculation results are analyzed as follows:

1. According to Table 6, the parameter estimates values are all in 95 percent confidence interval. It shows that the parameters estimations values have high confidence and the fitting effect is good. Relatively, the confidence interval of the parameters $b2$, $b3$ and $b7$ are wide. The main reason is that there are only a few samples. The confidence interval can be narrowed effectively by increasing the sample size.
2. According to Table 7, $R^2 = 0.628$ in the model. It shows that the regression equation can explain the 62.8 percent variation of the dependent variable. It has a good fitting degree.
3. According the observation value and regression curve in Fig. 4, it can be seen that the regression equation has a high goodness-of-fit.

According to Fig. 4, as the speed standard deviation increases, the accident rate has a growing trend. Based on Eq. (21), when the speed standard deviation is 20 km/h, accident rate per 100 million vehicle kilometres is 32.2390. If the speed standard deviation reaches 30 km/h, the accident rate will rise to 55.1232, which increases by 42.11 percent. It can be seen that a lower speed standard deviation has a lower accident rate.

4.3 Multivariate nonlinear regression

In the same way, by using of numerical analysis and scatterplot comparative method, user-defined model expression is constructed, and multivariate nonlinear regression is carried out. By nonlinear regression model of average speed-accident rate and the speed standard deviation-accident rate, the million vehicle

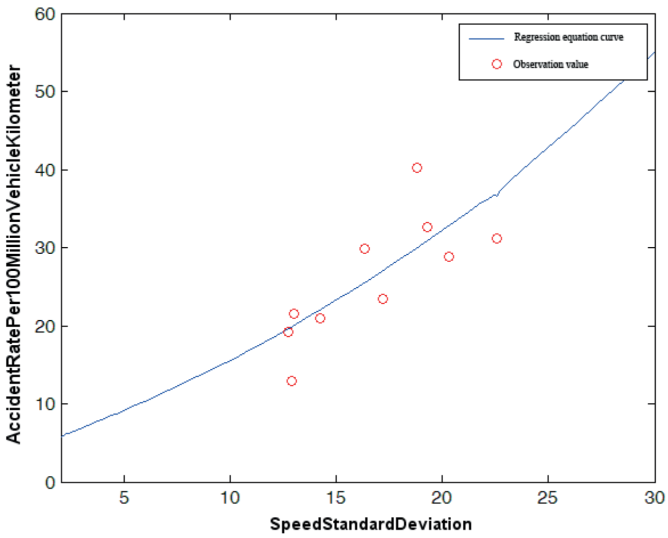

Fig. 4 The speed standard deviation - accident rate regression curve of user-defined model expression

kilometres accident rate increases with the increase of average speed and the speed standard deviation. Based on the principle of the 3 sigma (3 σ), the effect of the average speed plus 3 times the standard deviation on the accident rate was studied.

For the nonlinear regression of average speed and speed standard deviation-accident rate is carried out. The user-defined model expression as follows:

$$\begin{aligned}
&Accident\ Rate\ Per\ 100\ Million\ Vehicle\ Kilometers\\
&= b1*\ln(b2*(Speed+3*Speed\ Standard\ Deviation)\\
&*(Speed+3*Speed\ Standard\ Deviation)\\
&*(Speed+3*Speed\ Standard\ Deviation)\\
&+b3*(Speed+3*Speed\ Standard\ Deviation)\\
&*(Speed+3*Speed\ Standard\ Deviation)\\
&+b4*(Speed+3*Speed\ Standard\ Deviation))\\
&+b5*(Speed+3*Speed\ Standard\ Deviation)\\
&*(Speed+3*Speed\ Standard\ Deviation)\\
&+b6*(Speed+3*Speed\ Standard\ Deviation)+b7.
\end{aligned} \tag{22}$$

Similarly, the parameters initial values are given as follows: $b1$ = 0.22, $b2$ = 0.005, $b3$ = -0.509, $b4$ = 3.40, $b5$ = 0.0015, $b6$ = 0.001, $b7$ = 0.009. Constraints are defined as follows: $b7 \geq 0.05$, $b3 \geq -0.6$ and $b4 \geq 0$. The parameter estimates are shown in Table 8 and ANOVA are shown in Table 9. According to Table 8, the regression model can be expressed as below:

$$\begin{aligned}
&Accident\ Rate\ Per\ 100\ Million\ Vehicle\ Kilometers\\
&= -0.401*\ln(0.001*(Speed+3*Speed\ Standard\ Deviation)^3\\
&-0.114*(Speed+3*Speed\ Standard\ Deviation)^2+3.381\\
&*(Speed+3*Speed\ Standard\ Deviation))+0.001\\
&*(Speed+3*Speed\ Standard\ Deviation)^2+0.099\\
&*(Speed+3*Speed\ Standard\ Deviation)+0.05.
\end{aligned} \tag{23}$$

Table 8 Parameter estimates

Parameter	Estimate	Std. Error	95 % Confidence Interval	
			Lower Bound	Upper Bound
b1	-0.401	9023.290	-28716.537	28715.735
b2	0.001	28.441	-90.512	90.514
b3	-0.114	1644.652	-5234.131	5233.903
b4	3.381	413596.625	-1316245.669	1316252.432
b5	0.001	0.323	-1.027	1.029
b6	0.099	63.884	-203.209	203.407
b7	0.050	95376.239	-303529.709	303529.809

Table 9 ANOVAa

Source	Sum of Squares	df	Mean Squares
Regression	7190.518	7	1027.217
Residual	160.713	3	53.571
Uncorrected Total	7351.231	10	
Corrected Total	563.779	9	

Dependent variable: Accident Rate Per 100 Million Vehicle Kilometers
a. R squared = 1 - (Residual Sum of Squares) / (Corrected Sum of Squares) = 0.715.

The calculation results are analyzed as follows:

1. According to Table 8, the parameter estimates values are all in 95 percent confidence interval. It shows that the parameters estimations values have high confidence and the fitting effect is good. Relatively, the confidence interval of the parameters $b4$ and $b7$ are wide, and it can become narrow by increasing the sample size.
2. According to Table 9, R^2 = 0.715 in the model. It shows that the regression equation can explain the 71.5 percent variation of the dependent variable. It has a good fitting degree.
3. According the observation value and regression curve in Fig. 5, it also can be seen that the regression equation has a high goodness-of-fit.

According to Fig. 5, as the average speed plus 3 times the standard deviation increases, the accident rate has a growing trend. Based on Eq. (23), when the speed standard deviation is respectively 20 km/h and 30 km/h, the accident rate per 100 million vehicle kilometres is respectively 30.7469 and 42.6951. At the speed standard deviation of 20 km/h, if the speed is respectively 100 km/h and 120 km/h, the accident rate is respectively 38.5029 and 47.0948, which increase by 22.32 percent as a 20 km/h speed increase. It can be obviously seen that a lower speed and a lower speed standard deviation have a lower accident rate.

5 Conclusion

The average speed and dispersion of speed distribution is directly related to the accident rate. The unitary nonlinear regression models of average speed-accident rate, and the speed

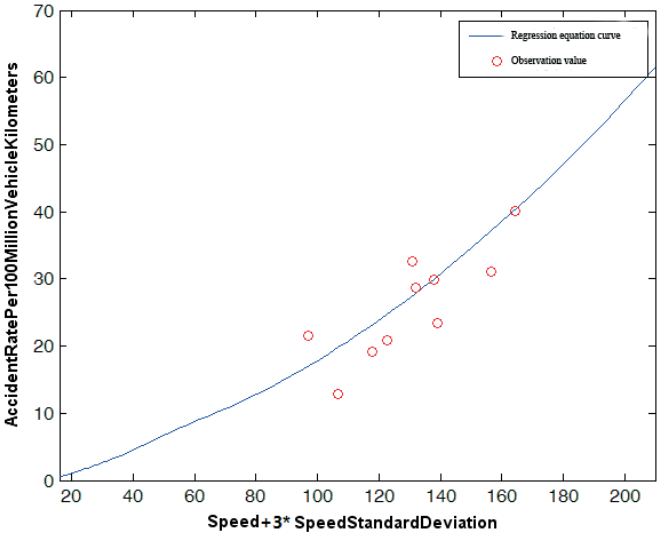

Fig. 5 The average speed plus 3 times the standard deviation - accident rate regression curve

standard deviation-accident rate, as well as multivariate linear regression model of the average speed plus 3 times the standard deviation-accident rate were built. The fitting effects of three kinds of models are all ideal. Compared to the univariate regression, the multivariate nonlinear regression model can explain better to the dependency relationship among the speed, speed standard deviation and accident rate, also more effectively predict or estimate the accident rate according to the observation data. Its practical significance is better. The difficulty of nonlinear regression lies in the establishment of the user-defined model expression and the setting of initial parameter. It needs the ability to grasp the general layout of data scatter and the shape of equation, good experience and patient debugging.

Through the analysis on the quantitative relationship among the average speed, the dispersion of speed distribution, and the accident rate, it can be known that the accident rate will increase rapidly with the increase of the vehicle speed and its dispersion of distribution, i.e., that a lower speed and a lower speed standard deviation has a lower accident rate. Although considering efficiency, to improve the speed will cut travel time. However, because the higher speed could lead more and more destructive accidents, based on the viewpoint of safety, highway management department should be more effective to control the vehicle speed standard deviation, and give a limit between the high speed and low speed, decrease the difference of the average speed between large vehicle and car as possible, reduce the dispersion of the speed distribution, as well as guide reasonably the average speed and its distribution, so as to guarantee a lower accident rate.

References

Baruya, A., Finch, D. J., Wells, P. A. (1999). A Speed-accident Relationship for European Single-carriageway Roads. *Traffic Engineering & Control.* 40(3), pp. 135-139. [Online]. Available from: https://trid.trb.org/view/505191 [Accessed: 16th January 1999]

Chengdu Business Daily. (2014). The traffic accident was half due to the speed of the car. *Chengdu Business Daily digital subscriptions.* (in Chinese) [Online]. Available from: http://e.chengdu.cn/html/2014-11/ 17/content_497746.htm [Accessed: 17th November 2014]

Chen, H., Sun, X., He, Y., Wang, R. (2009). Rationality of speed limit for Anyang-Xianxiang highway and speed control. *Transport Research.* 19, pp. 202-207. (in Chinese)

Chen, T. (2016). There are 275,000 investigated illegal speeding cases on the Shandong province's freeway a month. *JiNan Times.* (in Chinese) [Online]. Available from: http://news.163.com/16/0511/19/BMQESBLD00014AEE.html [Accessed: 11th June 2016]

Department of Transport China. (2004). China construction standardization association highway engineering committee. Technical standards of highway engineering. *China Communications Press.* (in Chinese)

Du, B.-Y. (2002). Expressway Traffic Accident and Matrix of Vehicle Speed. *Journal of Highway and Transportation Research and Development.* 19, pp. 116-118. [Online]. Available from: https://www.researchgate.net/publication/296042768_Expressway_traffic_accident_and_matrix_of_vehicle_speed [Accessed: 20th December 2002]

Federal Highway Administration. (1998). Synthesis of Safety Research Related to Speed and Speed Limits. Federal Highway Administration Research and Technology, Washington, USA, Publication Number: FHWA-RD-98-154.

Garber, N. J., Gadiraju, R. (1989). Factors Affecting Speed Variance and Its Influence on Accidents. *Transportation Research Record.* 1213, pp. 64-71. [Online]. Available from: http://onlinepubs.trb.org/Onlinepubs/trr/1989/1213/1213-009.pdf [Accessed: 13th December 1989]

He, Y., Wang, S., Sun, X., Zhang, J., Hou, S. (2010). Operating speed and highway safety in China and the United States. *China Journal of Highway and Transport.* (Special issue) 1, pp. 73-78. (in Chinese)

Hu, G. (2008). Research on Safety Evaluation and Improvement Strategies of Freeway Traffic Flow. Chongqing Jiaotong University, Chongqing, China. (in Chinese)

Integrated planning division in Ministry of Transport of China. (2017). The Statistical bulletin of transport industry development. (in Chinese) [Online]. Available from: http://was.mot.gov.cn:8080/was5/web/search?channelid=225629 [Accessed: 17th April 2017]

Jomaa, D., Dougherty, M., Yella, S., Edvardsson, K. (2016). Effectiveness of trigger speed of vehicle-activated signs on mean and standard deviation of speed. *Journal of Transportation Safety & Security.* 8(4), pp. 293-309. https://doi.org/10.1080/19439962.2014.976690

Kloeden, C., Ponte, G., McLean, J. (2001). Travelling speed and the risk of crash involvement: The South Australian experience. In: National Speed and Road Safety Conference, Adelaide, Australia, Aug. 23-24, 2001. [Online]. Available from: http://hdl.handle.net/2440/48576 [Accessed: 24th August 2001]

Lidbe, A., Tedla, E., Hainen, A., Sullivan, A., Jones Jr, S. (2017). Comparative assessment of arterial operations under conventional time-of-day and adaptive traffic signal control. *Advances in Transportation Studies.* 42, pp. 5-22.

Li, G., Li, D., Gou, X. (2010). Analysis on traffic accidents and safety prevention countermeasures of expressway in Xihan expressway. *China ITS Journal.* 2, pp. 101-103. (in Chinese)

Liu, G., Popoff, A. (1997). Provincial-Wide Travel Speed and Traffic Safety Study in Saskatchewan. *Transportation Research Record: Journal of the Transportation Research Board.* 1595, pp. 8-13. https://doi.org/10.3141/1595-02

Pei, Y.-L., Cheng, G.-Z. (2004). Research on the relationship between discrete character of speed and traffic accident and speed management of freeway. *China Journal of Highway and Transport.* 17, pp. 74-78. (in Chinese)

Shao, J. (2004). For The Accident Of Badaling Expressway Of The Cause Of Analysis And Countermeasure Research. Thesis, Beijing University of Technology, Beijing, China. (in Chinese)

Solomon, D. (1964). Accidents on Main Rural Highways Related to Speed, Driver, and Vehicle. Bureau of Public Roads, Washington, USA. [Online]. Available from: https://trid.trb.org/view/1186988 [Accessed: 24th August 2012]

State Statistics Bureau of China. (2017). National economic and social development statistics bulletin of the People's Republic of China. (in Chinese) [Online]. Available from: http://www.stats.gov.cn/tjsj/zxfb/201702/t20170228_1467424.html [Accessed: 28th February 2017]

Statistics department of integrated planning division in Ministry of Transport of China. (2017).What are the new features of highway transportation? Look at the statistics. (in Chinese) [Online]. Available from: http://www.zgjtb.com/2017-06/09/content_117400.htm [Accessed: 9th June 2017]

Taylor, M. C., Lynam, D. A., Baruya, A. (2000). The Effects of Drivers' Speed on the Frequency of Road Accidents. In: TRL Report No. 421. Transport Research Laboratory TRL, Workingham, United Kingdom. [Online]. Available from: http://20splentyforus.org.uk/UsefulReports/TRLREports/trl421SpeedAccidents.pdf [Accessed: 9th September 2000]

Tian, X. (2017). People's transportation: let civilized driving become a habit. [Online]. Available from: http://www.rmjtxw.com/news/sd/13474.html [Accessed: 1st September 2017]

Zou, Y., Niu, G., Luo, Y., Zhong, Y. (2011). Suggestions to widen university town section inner freeway to 8 lanes. *Chongqing Architecture.* 10(2), pp. 5-9. (in Chinese)

Yang, Y., Dang, Y., Chen, C., Zhang, Z., Zhang, C. (2009). Overall design of accommodation expressway project of Xianyang international airport in Xi'an city. *Highway.* 6, pp. 1-4. (in Chinese)