Eurasia Journal of Science and Technology	ISSN: 2663-1024	Volume 2, Issue 1, pp 34-42	Feb 2020

ROAD TRAFFIC ACCIDENTS PREDICTION MODEL IN CHINA

Lingxiang Wei[1]*, Yuxuan Li[1], Mingjun Liao[1,2,]*

[1] *School of Materials Science and Engineering, Yancheng Institute of Technology, Yancheng 224051, China;*
[2] *Key Laboratory of Transport Industry of Big Data Application Technologies for Comprehensive Transport, Ministry of Transport, Beijing Jiaotong University, Beijing 710064, China*
Corresponding email: weilx@ycit.edu.cn; mjliao@126.com

Abstract
In China, as in other countries, road traffic deaths are a burden for society. More than 58,000 people die in road crashes, and approximately 213,000 are injured annually in China. In 2018, there are 244,924 road traffic accidents all over China. To prevent traffic incidents, it is crucial to understand where and how they take place, the change trend in recent years and so on. The aim of this study is to make a change trend analysis in road traffic accidents concerning time and locations within macroscopic traffic accidents data from Yearbook of Road Traffic Accidents in China and National Bureau of Statistics of China using Smeed's Law. The result of this study shows that Smeed's Law clearly represents road traffic accidents prediction model in China. In addition, further studies will turn to analyze microcosmic causes of the traffic accidents.
Keywords: traffic engineering; road traffic accidents; Smeed's law; prediction model

1. INTRODUCTION

From the perspective of the whole world, at present, due to the increase of the number of vehicles and the speed of vehicle operation, the number of road traffic accidents is also increasing, which has become a worldwide social hazard. According to the World Health Organization, road traffic injuries were the 10th leading cause of death worldwide in 2018 and, unless immediate action is taken, they will become the fifth leading cause of death by 2030 [1-3]. Since the mid-1990s, as the largest developing country in the world, China's social and economic development is rapid, the process of automobile is rapid, and the road traffic volume is increasing rapidly, which brings great pressure to our road traffic safety work. In China, as in other countries, road traffic deaths are a burden for society. In 2018, there are 244,924 traffic accidents all over China, which results in a direct economic loss of 1,384.56 million yuan. Around 64,377 people died, and 258,532 people got injured. In addition, there were 520 accidents with more than 3 deaths, 6accidents with more than 5 death, and 5 accidents with more than 10 deaths. Nowadays, China has also witnessed a substantial rise in the number of traffic crashes, injuries and fatalities, especially since 1998 [4,5]. As a result, it is important to study road traffic accidents change trend and prediction model in China.
According to the data from the database of the National Bureau of statistics, in 2018, China's car ownership reached 452416400, and the number of drivers reached 1395380000. Compared with 2001, the number of motor vehicles has increased about six times, which to a large extent causes the shortage of traffic resources, and aggravates the situation of traffic congestion and the number of traffic accidents. According to the data provided by the Traffic Management Bureau of the Ministry of Public Security Ministry, the number of traffic accidents in 2003 is 667,507. With the improvement of traffic road norms, as of 2018, the data still reached 244,924, still at a high value. The number of deaths in 2003 was 104,372, down to 64,377 in 2018. It is foreseeable that, in the context of increasing private car ownership, in addition to the unsound road norms and the black spots of traffic accidents, the fact that the road traffic accident prediction model is difficult to establish leads to the social and economic losses in China, and leads to the material and spiritual loss of traffic accident personnel. Therefore, the establishment of an effective traffic accident prediction model can accurately predict traffic accident prone areas and drivers from the source, so as to achieve the goal of reducing road traffic accidents.
In response to these issues, this study follows an earlier work on the analysis of traffic accidents prediction model using Smeed's Law [6-9] and aims to identify the traffic accidents change trend in China. By identifying the changing trend of

road traffic accidents in China, the prevention and control of road traffic accidents can be realized. The road traffic accidents features are including frequency in district, general trends and analysis of the causes. Frequency in district includes the accumulated total number of accidents or fatalities of China's provinces or cities in 2018, and general trends contain the general trends of car ownership, freeway fatalities, the number of accidents with more than 10 deaths.

2. LITERATURE REVIEW

With the development of traffic accident analysis, the main purpose of these studies is to establish a new model based on statistical methods to determine the area of traffic accident concentration. Under such circumstances, it is necessary to accurately and accurately determine the measures to be taken to reduce accidents. Therefore, the first step is to choose the appropriate prediction model. Most of the modeling methods are constructed in the discrete response model. These models include binary response model, ordered discrete response model, and disordered multiple discrete response models using other uncommonly used methods, including artificial neural networks and some data mining techniques, such as classification and regression tree (CART) analysis.

Since 2000, the spatial statistical study of traffic accidents has been ongoing. Rolison has collected expert opinions of the police, opinions of pedestrians and official road traffic accident records by using a variety of investigation methods. Through the statistical analysis of these data, the causes of traffic accidents have been analyzed from the aspects of cognition, experience and action ability [10]. Potoglou also studied the causes of non-fatal traffic accidents in southern Europe by establishing a mixed impact logistic regression model in 2018 [11]. On the basis of fully analyzing the causes of road traffic accidents, it is necessary to build a model. At present, deep learning algorithm is widely used to build accident prediction model. Zhang uses Deep Belief Network (DBN) and Long Short-Term Memory (LSTM) to realize traffic accident detection [12]. Yuan, Z realized the prediction of road traffic accidents by building a new place, the Convolutional Long Short-Term Memory (ConvLSTM) neural network model [13]. Based on the spatiotemporal correlation model, Ren constructed a high-precision deep learning model for traffic accident risk prediction, which was used to achieve traffic accident prediction [14]. Shi, X. established a traffic accident prediction model based on feature recognition, i.e. eXtreme Gradient Boosting（XGBoost） realized road traffic accident prediction [15]. In addition, some prediction models do not use deep learning algorithm. Some people choose to use ARIMA and ARIMAX to build models [4]. Gianfranco, on the other hand, uses Poisson and negative binomial algorithm to build a road traffic accident prediction model [16]. Traffic accidents are divided into two types: single vehicle collision and multi vehicle collision, and relevant accident prediction models are established respectively [17]. Artificial neural network is a well-researched modeling tool, which can map complex nonlinear models and non-intuitive relations among variables. They have proven their effectiveness in various engineering fields, such as estimating the cost of construction [9] or the number of building materials [18].

In 1949, the fatality prediction model was developed by Smeed in which he used both time-series and cross-sectional analysis of fatality data from developed countries. Smeed's Law linked the number of fatalities to population with the number of vehicles [19]. Reference [20] used Using the 1991–2009 panel data from all states of India to illustrates the robustness of Smeed's law, and provides evidence for time-invariance but state-specificity. Reference [21] analyzed the traffic accidents in the Kingdom of Saudi Arabia (KSA). This study used Smeed's formula based on KSA data. Reference [22] built a fatality predictive model using Smeed's law and general linear regression model for motorcycle accidents data in Malaysia from 1996 to 2010. So Smeed's law was widely used to analyze accident data about traffic fatalities.

3. DATA COLLECTION

In China, the Traffic Management Bureau of the Ministry of Public Security Ministry is a state organ for organizing, guiding and supervising local public security organs to investigate and deal with road traffic illegal acts and traffic accidents according to law. The Yearbook of Road Traffic Accidents in China in 2013 is written by the Traffic Management Bureau of the Ministry of Public Security Ministry. It records the types of road traffic accidents, the severity of the accidents, the type of vehicles, the number of vehicles, the number of injuries, the number of deaths, the cause of death, the date, time and place of the accident. For this study, the Traffic Management Bureau of the Ministry of Public Security Ministry provided the 2013 annual statistical report on road traffic accidents in People's

Republic of China. The book consists of data from a sample of police-reported crashes that include the number of road traffic fatalities, road traffic accidents, vehicles and population size about 31 provinces or cities in China from 2003 to 2013 and also includes fatality rates per 100,000 population and fatality rates per 100,000 vehicles in China from 1983 to 2013. Another source of data is National Bureau of Statistics of China. Details include annual average daily traffic fatalities and accidents, traffic fatalities and accidents per provinces or cities, freeway mileages and the number of freeway fatalities, road traffic accidents with over 10 deaths and so on from 2003 to 2018 in China.

4. ROAD TRAFFIC ACCIDENTS FEATURES

The number of vehicle ownership has increased along with the growth of China's economy. Let us look at the road traffic accidents features in China: frequency in district, general trends and analysis of the causes.

5. FREQUENCY IN DISTRICT

In 2018, more than 64,377 people died in road crashes, and approximately 258,532 are injured in China. Figure 1 displayed the accumulated total number of accidents or fatalities of China's provinces or cities in 2018. The data shows that a large number of accidents happened in developed regions and forelands such as Guangdong province, Zhejiang province, Anhui province, Jiangsu province while the number of accidents of those economy developed regions and forelands were not so high as well as number of fatalities.

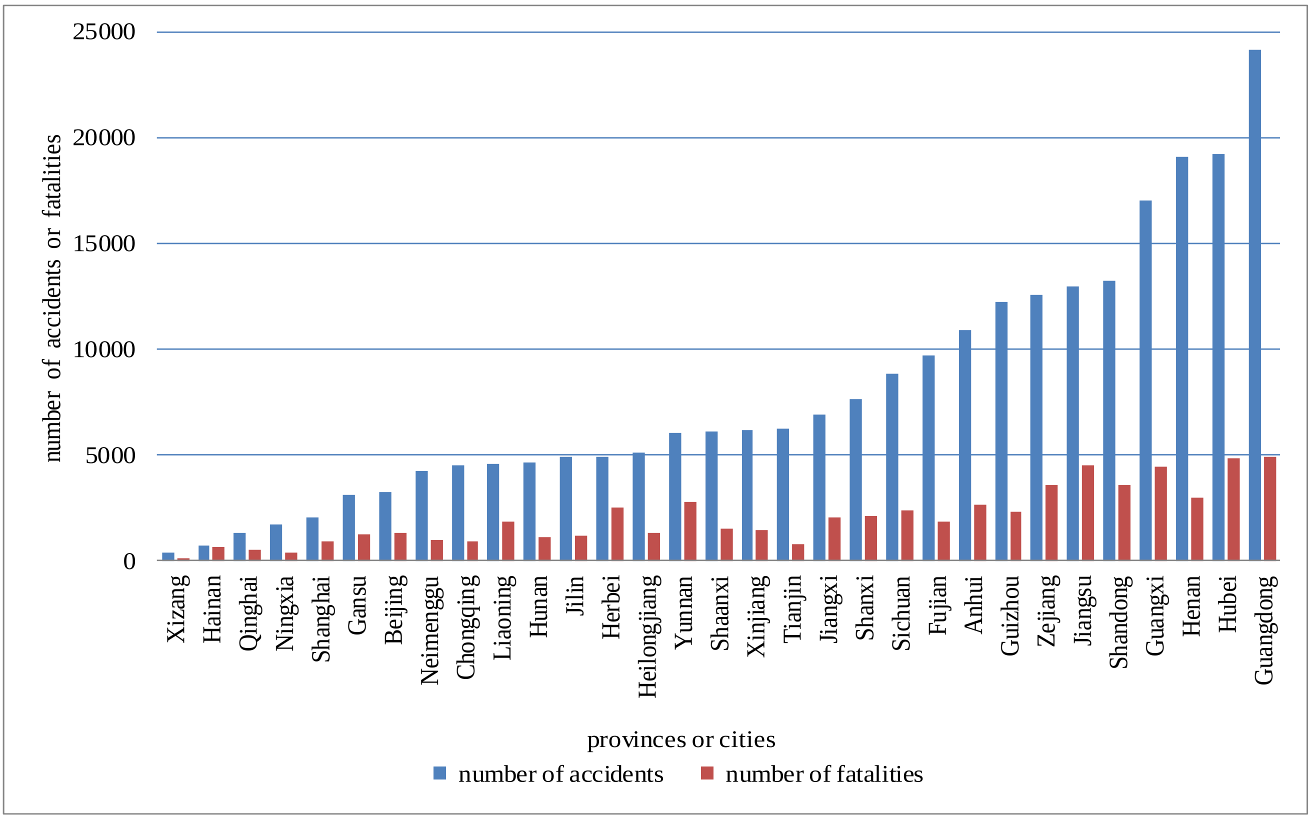

Figure 1 Number of accidents or fatalities by provinces or cities, China, 2018

The total number of vehicles ownership and drivers have increased along with the growth of China economy. In 2018, the total numbers of vehicles were more than 382 million and the total numbers of vehicle drivers were more than 279 million in China. Figure 2 displays the total number of vehicles and vehicle drivers in different provinces or cities of China in 2018.

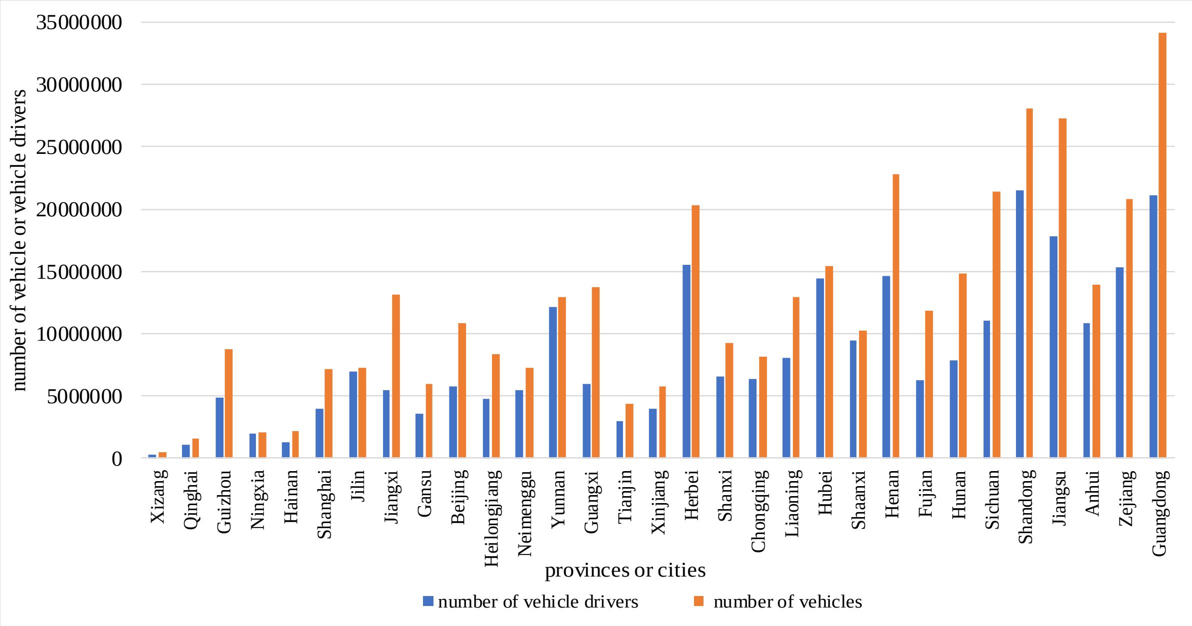

Figure 2 Number of vehicles or vehicle drivers by provinces or cities, China, 2018

6. GENERAL TRENDS

Analyses were performed on the data collected from statistical yearbook of road traffic accidents in China from 2003 to 2018. The general trends included the car ownership, freeway fatalities, and the number of accidents with more than 10 deaths.

6.1 General trends of car ownership

The number of newly vehicles increased in multi-folds along with the increase in the level of vehicles ownership from 2001 to 2011 (lack of corresponding data after nearly two years) as shown in Figure 3. The data shows that the number of vehicles per 100 populations increased linearly with the increasing of number of vehicles from 2001 to 2011, and we deduced that it will continue to increase in recent years.

6.2 General trends of freeway fatalities

Figure 4 illustrates general trends of the freeway mileage and number of freeway fatalities in China between 2003 and 2018. The statistics show that the freeway mileages were increasing rapidly while the numbers of freeway fatalities were falling in the last few years, but the levels of fatalities were still too high. The cause of freeway fatalities included the illegal act and fault of motor vehicles (89.73%), the illegal act of non-motorized vehicle (0.4%), the illegal act of pedestrians and (9.24%), and other reasons (0.63%) in 2013. The occurrence of the object of freeway fatalities included vehicles & vehicles accident (62.33%), vehicles & pedestrians accident (14.30%), single vehicle accident (23.37%) in 2013.

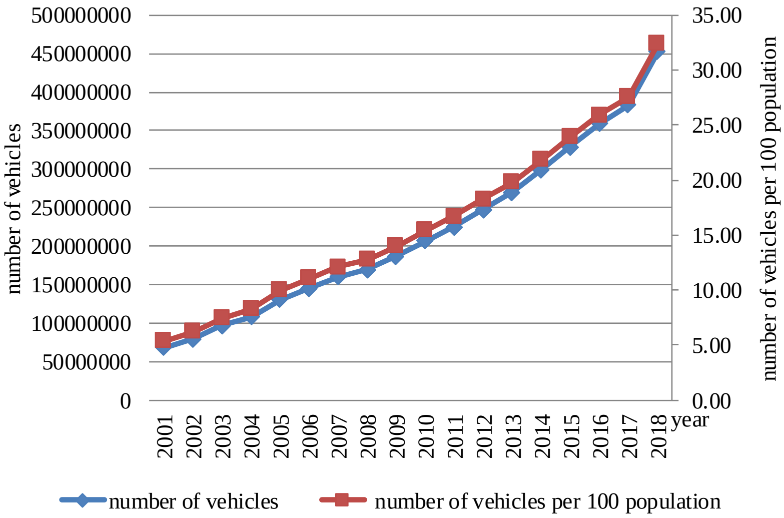

Figure 3 Number of vehicles and number of vehicles per 100 population, China, 2001-2018

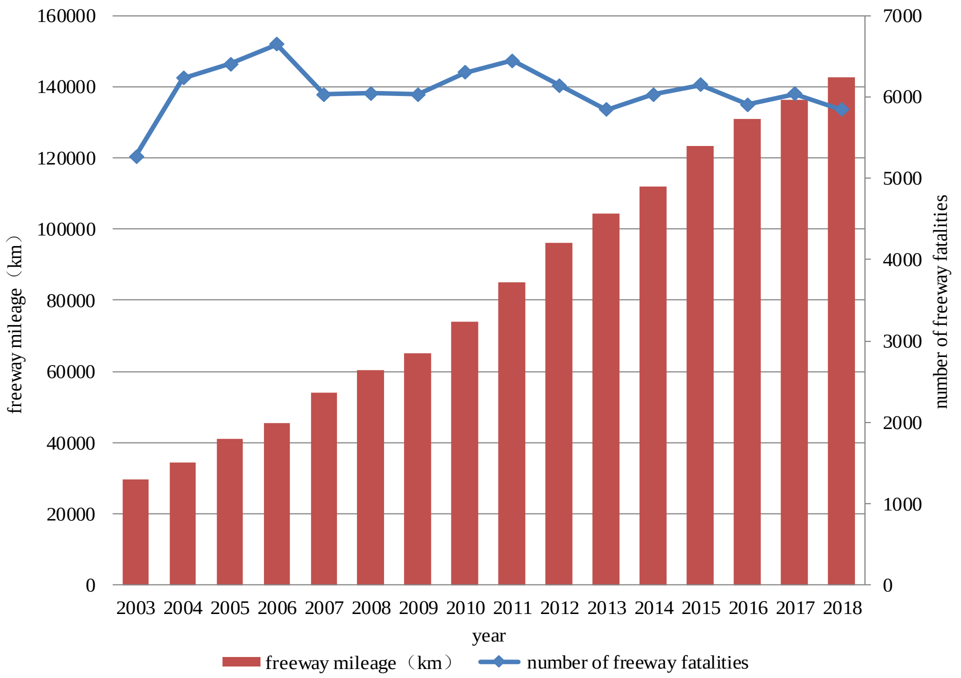

Figure 4 Freeway mileages and number of freeway fatalities, China, 2003-2018

6.3 Accidents and fatalities with more than 10 deaths.

Figure 5 illustrates the number of accidents and fatalities of a death over ten people in one traffic accident in China between 2003 and 2015. The statistics show that although the numbers of newly vehicles were increasing rapidly as seen from Figure 3, number of accidents and fatalities of a death over ten people in one traffic accident were falling in the last few years.

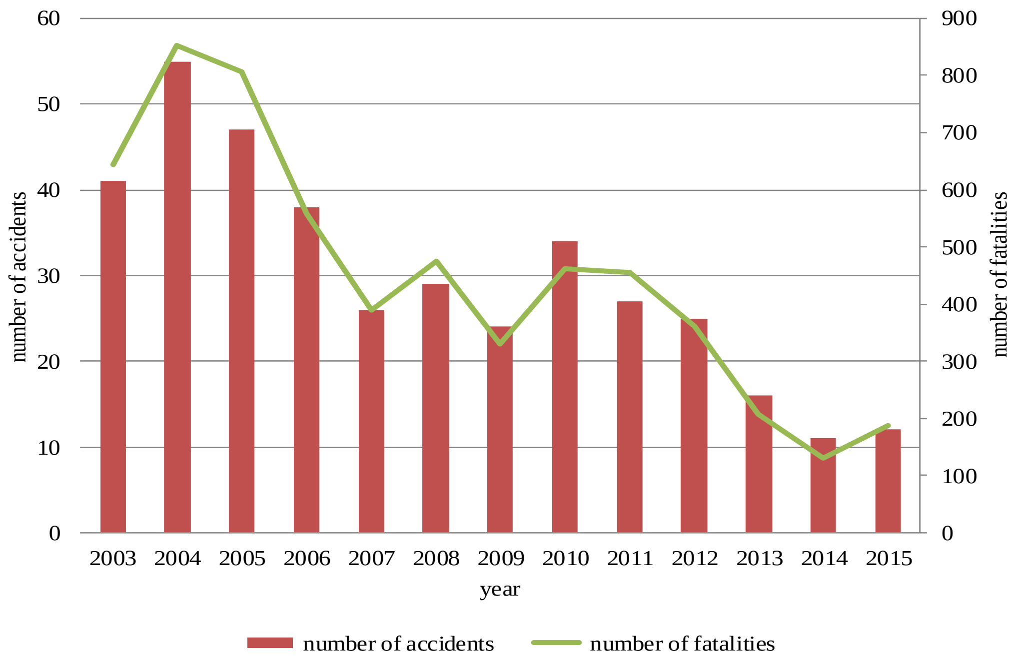

Figure 5 Number of accidents and fatalities of a death over ten people in one traffic accident per year, China, 2003-2015

7. ANALYSIS OF THE CAUSES

The causes of road traffic accidents steadily according to road users during the years 2006 to 2011, and 2013, which are shown in Table 1. However, law-breaking in motor vehicles remains

at a high percentage from 85.46% to 91.08%, and mistake but not law-breaking in motor vehicles and law-breaking in non-motorized vehicles also contributes to the causes of road traffic accidents.
In order to reduce the accident rate in the next few years, law-breaking in motor vehicles which included over speed driving, driving after drink, fatigue driving must be given heavier penalties and better managed. In addition, sufficient traffic safety education is not only necessary for drivers but also for voluntary participants (i.e., pedestrians, bicyclists, et al).

Table 1 Percent of the main causes of road traffic accidents in China, 2006 to 2011, and 2013

year	2006	2007	2008	2009	2010	2011	2013
Law-breaking in motor vehicles	85.46 %	89.35 %	90.68 %	90.93 %	91.08 %	90.86 %	88.93 %
Mistake but not law-breaking in motor vehicles	8.80%	4.97%	4.22%	3.79%	3.74%	3.50%	4.19%
Law-breaking in non-motorized vehicles	3.52%	3.81%	3.58%	4.01%	3.98%	4.51%	5.77%
Law-breaking in pedestrians	1.85%	1.66%	1.37%	1.13%	1.08%	1.04%	0.99%
Causes in roadways	0.01%	0.01%	0.00%	0.01%	0.01%	0.01%	0.01%
Fortuitous event	0.36%	0.20%	0.15%	0.13%	0.11%	0.08%	0.11%

8. PREDICTION MODEL

The Smeed's model to represent the relationship between road traffic fatalities, road traffic accidents, vehicles and population is in the following formula (5, 6 and 8).

$$\frac{F}{V} = \alpha_1 \left(\frac{V}{P}\right)^{-\beta_1} \quad (1)$$

$$\frac{TAN}{V} = \alpha_2 \left(\frac{V}{P}\right)^{-\beta_2} \quad (2)$$

Where, F=number of road traffic fatalities, TAN = number of road traffic accidents, V = number of vehicles, P=population size.
Smeed's Law is usually criticized for having the number of vehicles on both sides of the equation and that there is a considerable deviation between the expected and actual number of road fatalities(8). Therefore, this study tries to fit another predictive model that can produce better estimates than Smeed's equation. The model as suggested by reference [23]is as follows:

$$F = z_1 V^{x_1} P^{y_1} \quad (3)$$

$$TAN = z_2 V^{x_2} P^{y_2} \quad (4)$$

The parameters x1, x2, y1, y2, z1 and z2 are estimated using regression analysis by first converting (3) and (4) into a linear form:

$$\ln(F) = \ln z_1 + x_1 \ln M + y_1 \ln P \quad (5)$$

$$\ln(TAN) = \ln z_2 + x_2 \ln M + y_2 \ln P \quad (6)$$

Then, the parameters x1, x2, y1, y2, z1 and z2 can be determined by multiple linear regression theory [24]. After taking the antilog of the above model, the estimated model is written as (7) and (8).

$$F = e^{\ln z_1} V^{x_1} P^{y_1} \quad (7)$$

$$TAN = e^{lmz_2} V^{x_2} P^{y_2} \quad (8)$$

In order to measure the strength and the direction of a linear relationship among the variables, the coefficient of determination (R2) will be counted in the Smeed's model. The coefficient of determination (R2) is useful because it gives the proportion of the fluctuation of one variable that is predictable from the other variable [25]. The coefficient of determination is such that 0 < R2 < 1, and denotes the strength of the linear association. If there is linear correlation or a good linear correlation, R2 is close to 1. A correlation greater than 0.7 is generally described as strong, whereas a correlation less than 0.5 is generally described as weak.

9. PARAMETERS CALIBRATION

Table 2 Road traffic fatalities and road traffic accidents predictive model from 2013 to 2018.

Predictive model	Year	Sample size	x1	y1	lnz1	R2
road traffic fatalities predictive model	2013	31	0.279	0.428	-4.439	0.8113
	2014	31	0.281	0.450	-4.902	0.8174
	2015	31	0.426	0.325	-5.093	0.8390
	2016	31	0.514	0.273	-5.552	0.8577
	2017	31	0.493	0.280	-5.353	0.8047
	2018	31	0.543	0.231	-5.062	0.7464
	2013-2018	186	0.408	0.000	-0.346	0.9142
road traffic accidents predictive model	2013	31	0.459	0.286	-3.710	0.6711
	2014	31	0.439	0.400	-5.453	0.6992
	2015	31	0.493	0.330	-5.161	0.6817
	2016	31	0.401	0.465	-5.943	0.7053
	2017	31	0.409	0.434	-5.594	0.6972
	2018	31	0.574	0.303	-5.541	0.6892
	2013-2018	186	0.462	0.000	-0.212	0.7670

Table 1 shows a moderately fit with a coefficient correlation, $R^2 \in (0.68,1)$ with the residuals normal and acceptable. Hence, the predictive models on the above clearly represent road traffic accidents prediction model in China. The work on data from developed countries suggests that fatality rates per vehicle to decrease over time while fatality rates per person to increase over time (4, and 6). However, Figure 6 shows that since 2003, the fatalities per 100,000 populations and 10,000 vehicles shows a decreasing trend due to the enhancement of the awareness of traffic safety, improvement of traffic environment, the sound of traffic facilities progressively.

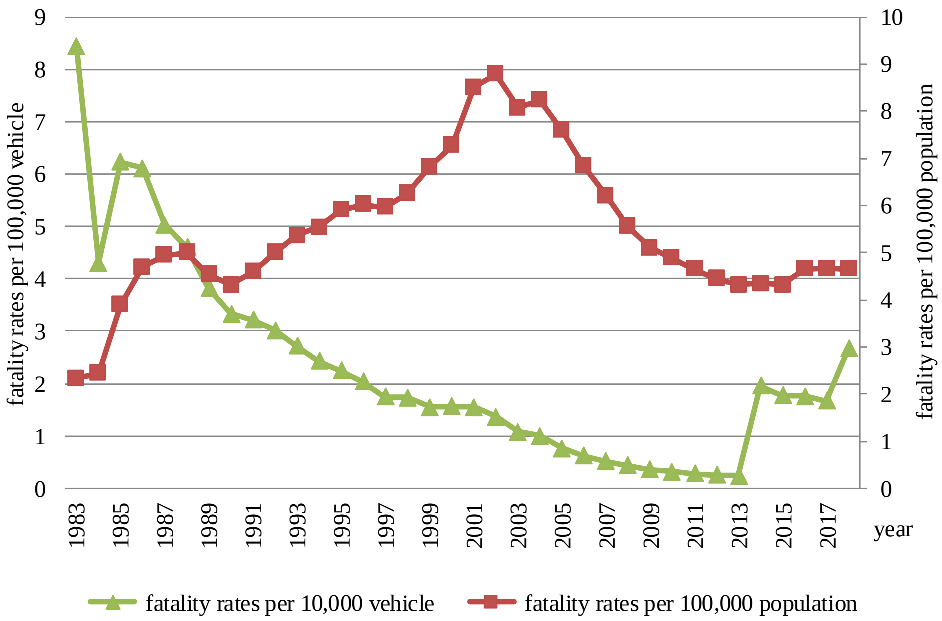

Figure 6 Fatalities per 100,000 populations and 10,000 vehicles

10. CONCLUSIONS

Several significant conclusions can be drawn:

(1) In contrast with less developing area such as

Tibet autonomous region, Qinghai province, Guizhou province, Ningxia Hui autonomous region., the total number of traffic accidents in developed area (such as Guangdong province, Zhejiang province, Anhui province, Jiangsu province etc.) is larger but number of fatalities is smaller.

(2) The mortality of road traffic fatalities in national roads, provincial roads and high-grade roads is much higher than that in other kind of roads [26, 27]. In addition, the freeway mileages were increasing rapidly while the numbers of freeway fatalities were falling in the last few years, but the levels of fatalities were still too high.

(3) In order to reduce the accident rate in the next few years, law-breaking in motor vehicles which included overspeed driving, driving after drink, fatigue driving and so on must be given heavier penalties and better managed efforts. In addition, sufficient traffic safety education is also necessary not only for drivers but for voluntary participants (i.e., pedestrians, bicyclists, et al).

(4) This study shows a moderately fit with a coefficient correlation, $R^2 \in (0.68,1)$ with the residuals normal and acceptable. Hence, Smeed's Law on the above clearly represent road traffic accidents prediction model in China. In addition, since 2003, the fatalities per 100,000 population and 10,000 vehicle shows a decreasing trend due to the enhancement of the awareness of traffic safety, improvement of traffic environment, the updating of traffic facilities.

AUTHOR CONTRIBUTIONS

Y.L. contributed to data collection; L.W. and M.L. proposed the research framework, analyzed the data and wrote the article; L.W. and Y.L. contributed to revising article.

FUNDING

This study was supported by the National Natural Science Foundation of China (Grant No.51338008), Key Laboratory of Transport Industry of Big Data Application Technologies for Comprehensive Transport (Ministry of Transport, Beijing Jiaotong University).

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

REFERENCES

[1] Patel A, Krebs E, Andrade L, et al. The epidemiology of road traffic injury hotspots in Kigali, Rwanda from police data. *BMC Public Health*, 2016, *16(1),* 697.

[2] Anjuman, Tahera, et al. Road traffic accident: A leading cause of the global burden of public health injuries and fatalities. *Proc. Int. Conf. Mech. Eng*, 2020,29-31

[3] Castillo-Manzano, J. I., Castro-Nuño, M., López-Valpuesta, L., & Vassallo, F. V. An assessment of road traffic accidents in Spain: the role of tourism. *Current Issues in Tourism,* 2020, *23(6)*, 654-658.

[4] Ihueze, C. C., and U. O. Onwurah. Road traffic accidents prediction modelling: An analysis of Anambra State, Nigeria. *Accid Anal Prev*, 2018, *112,* 21-29.

[5] Liang, Ci, et al. Developing accident prediction model for railway level crossings. *Safety Science*. 2018,*101*,48-59.

[6] Deublein, M, et al. Prediction of road accidents: A Bayesian hierarchical approach. *Accident Analysis & Prevention.*2013, *51 (4),* 274-291.

[7] Deublein, Markus, M. Schubert, and B. T. Adey. Prediction of road accidents: comparison of two Bayesian methods. *Structure & Infrastructure Engineering.*2014, *10(11),* 1394-1416.

[8] Yueh-Tzu Lu,Mototsugu Fukushige. Smeed's law and the role of hospitals in modeling traffic accidents and fatalities in Japan. *Asia-Pacific Journal of Regional Science.*2019, *3(2),* 319-332.

[9] Luca Persia, Roberto Gigli, Davide Shingo Usami. Smeed's law and expected road fatality reduction: An assessment of the Italian case. *Journal of Safety Research.*2015, *55,* 121-133.

[10] Rolison J J, Regev S, Moutari S, et al. What are the factors that contribute to road accidents? An assessment of law enforcement views, ordinary drivers' opinions, and road accident records. *Accident Analysis & Prevention*, 2018, *115,* 11-24.

[11] Potoglou, D., Carlucci, F., Cirà, A., & Restaino, M. Factors associated with urban non-fatal road-accident severity. *International journal of injury control and safety promotion*, 2018, *25(3)*, 303-310.

[12] Zhang, Z., He, Q., Gao, J., & Ni, M. A deep learning approach for detecting traffic accidents from social media data. *Transportation research part C: emerging technologies,* 2018, *86,* 580-596.

[13] Yuan, Z., Zhou, X., & Yang, T. Hetero-convlstm: A deep learning approach to

traffic accident prediction on heterogeneous spatio-temporal data. *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining,* 2018, 984-992.

[14] Ren, H., Song, Y., Wang, J., Hu, Y., & Lei, J. A deep learning approach to the citywide traffic accident risk prediction. *2018 21st International Conference on Intelligent Transportation Systems (ITSC) IEEE,* 2018, 3346-3351.

[15] Shi, X., Wong, Y. D., Li, M. Z. F., & Chai, C. Accident risk prediction based on driving behavior feature learning using CART and XGBoost. *No. 18-06270*, 2018.

[16] Gianfranco, F., Soddu, S., & Fadda, P. An accident prediction model for urban road networks. *Journal of Transportation Safety & Security,* 2018, 10*(4),* 387-405.

[17] La Torre, F., Meocci, M., Domenichini, L., Branzi, V., & Paliotto, A. Development of an accident prediction model for Italian freeways. *Accident Analysis & Prevention,* 2019*, 124,* 1-11.

[18] Soto, Borja Garcíã-A De, B. T. Adey, and D. Fernando. A Process for the Development and Evaluation of Preliminary Construction Material Quantity Estimation Models Using Backward Elimination Regression and Neural Networks. *Journal of Cost Analysis & Parametrics*, 2014, 7*(3)*, 180-218.

[19] R. J. Smeed, Some statistical aspects of road safety research. *Journal of the Royal Statistical Society A. P.* 1949, *112(1),* 1-34.

[20] Raj V. Ponnaluri. Modeling road traffic fatalities in India: Smeed's law, time invariance and regional specificity. *IATSS Research*, 2012, 36*(1),* 75-82.

[21] Alghamdi A S. Road accidents in Saudi Arabia: a comparative and analytical study. *Urban Transport & the Enviroment for Century II,* 1996, 26, 23.

[22] Nasaruddin, Norashikin, et al. Fatality prediction model for motorcycle accidents in Malaysia. *Statistics in Science, Business, and Engineering (ICSSBE), 2012 International Conference on. IEEE*, 2012.

[23] Taamneh, Madhar, S. Alkheder, and S. Taamneh. Data-mining techniques for traffic accident modeling and prediction in the United Arab Emirates. *Journal of Transportation Safety & Security,* 2017, 9*(2),* 146-166.

[24] Seber G A F, Lee A J. Linear Regression Analysis, *Second Edition,* 2012.

[25] Abbondati, F., F. S. Capaldo, and R. Lamberti. Predicting driver speed behavior on tangent sections of low-volume roads. *International Journal of Civil Engineering & Technology,* 2017, 8*(4),* 1047-1060.

[26] Shen, Kun, and S. O. Management. Casualty Toll Prediction and Management Countermeasures for Road Traffic Accident in China. *Safety & Environmental Engineering,* 2017.

[27] Jiang, Bing Jiao, J. W. Shi, and Z. L. Chen. The Application of Cubic Exponential Smooth Model in Road Traffic Accidents Prediction. *Value Engineering*, 2017.