

## **Assessment of Expressway Traffic Safety Using Gaussian Mixture Model based on Time to Collision**

**Sheng Jin**

*College of Civil Engineering and Architecture, Zhejiang University  
Hangzhou 310058, China*

**Xiaobo Qu**

*Department of Civil and Environmental Engineering, National University of Singapore  
Singapore, 117576*

**Dianhai Wang**

*College of Civil Engineering and Architecture, Zhejiang University  
Hangzhou 310058, China*

*Corresponding Author's E-mail: wangdianhai@zju.edu.cn*

Received 6 July 2011

Accepted 25 November 2011

### **Abstract**

Traffic safety is of great significance, especially in urban expressway where traffic volume is large and traffic conflicts are highlighted. It is thus important to develop a methodology that is able to assess traffic safety. In this paper, we first analyze the time to collision (TTC) samples from traffic videos collected from Beijing expressway with different locations, lanes, and traffic conditions. Accordingly, some basic descriptive statistics of 5 locations' TTC samples are shown, and it is concluded that Gaussian mixture model (GMM) distribution is the best-fitted distribution to TTC samples based on K-S goodness of fit tests. Using GMM distribution, TTC samples can be divided into three categories: dangerous situations, relative safe situations, and absolute safe situations, respectively. We then proceed to introduce a novel concept of the percentage of serious traffic conflicts as the percentage of TTC samples below a predetermined threshold value in dangerous situation. After that, assessment results of expressway traffic safety are presented using the proposed traffic safety indicator. The results imply that traffic safety on the weaving segment is lower than that on mainlines and the percentage of serious traffic conflicts on median lane is larger than that on middle lane and shoulder lane.

*Keywords:* Time to collision, Gaussian mixture model, Expressway traffic safety.

### **1. Introduction**

Urban expressway is the highest urban road level in Chinese city and is a very important vehicular passageways for motorists and commuters. For example, Beijing urban expressway systems consist of 5 ring lines and 15 connecting lines, with a total length of about 380 km. With the increasing traffic volume, traffic safety is more severe. Traffic safety of expressway is most commonly measured in terms of the number of traffic accidents and the consequences of accidents in terms of

fatalities and injuries of differing severity. Whereas, traffic safety is a particularly difficult phenomenon to study, given the fact that accidents occur randomly in time and space thereby making measurement, assessment and comparison of this concept particularly difficult<sup>1</sup>. Thus, the underlying principle for a more effective safety evaluation strategy is to develop models based on proximal safety indicators that represent the temporal and spatial proximity characteristics of unsafe interactions and near-accidents. Research has shown

that proximal safety indicators have an established statistical relationship with accidents<sup>2-7</sup>. For expressway traffic safety studies that are aimed at assessing relative changes in levels of safety for different road environments or traffic conditions. It is necessary to use proximal indirect indicators of safety. These “safety indicators” are usually defined as traffic measures that are statistically correlated with the numbers of road traffic accidents at a particular location. These values are based on the temporal and spatial proximity between road-users during safety critical events. Svensson<sup>8</sup> states that for proxy measures or indicators of safety to be useful they must: (a) Complement accident data and be more frequent than accidents; (b) Have a statistical and causal relationship to accidents; (c) Have the characteristics of “near-accidents” in a hierarchical continuum that describes all severity levels of road-user interactions with accidents at the highest level and very safe passages with a minimum of interaction at the lowest level.

In the literature, there were many safety indicators applied for safety analysis such as time headway (TH)<sup>9</sup>, traffic conflict number (TCN)<sup>10</sup>, time-to-collision (TTC)<sup>11-13</sup>, extended time-to-collision (TET, TIT)<sup>14</sup>, deceleration rate (DR)<sup>15</sup>, proportion of stopping distance (PSD)<sup>16</sup>, and standard deviation of lateral position (SDLP)<sup>17</sup>. In these indicators, the TTC value is widely accepted as a highly useful and valid safety indicator for traffic conflicts on highways<sup>8, 13, 17</sup>. TTC between two consecutive vehicles is a common traffic parameter applied for safety estimation, obstacles avoidance, and collision warning system design. TTC is defined as “*the time that remains until a collision between two vehicles would have occurred if the collision course and speed difference are maintained*” by Hayward<sup>11</sup>, and it is discussed extensively in Hydén<sup>18</sup>. According to Svensson<sup>8</sup>, TTC is inversely related to accident risk (smaller TTC values indicate higher accident risks and vice versa). TTC has often been used as a safety indicator for certain maneuvers by determining the minimum TTC measured during the maneuver<sup>19-21</sup>. Furthermore, drivers are assumed to be motivated for a safety-related reason to exhibit an accelerating or decelerating response to TTC, and TTC was applied for modeling driving behavior by Van Winsum<sup>22</sup>, Jin et al.<sup>23, 24</sup>, and Bubb<sup>25</sup>. Another important research of TTC is to measure TTC threshold value for distinguish dangerous

situation and safe situation. Different opinions can be found in the literature as to which value should be used as safety limit. TTC threshold value suggests range from 1.5 s in urban areas<sup>8</sup> to 5s<sup>26</sup>, and 2.6 seconds for supported drivers and 3.5 seconds for non-supported drivers<sup>27</sup>. Thus, TTC threshold value is related to TTC distributions, and it is difficult for researcher to measure in different traffic conditions. Mainly for this reason, there is a need to further the concept of traffic safety indicator and assessment method that can indirectly be used to measure expressway traffic safety.

The objective of this paper is to develop a novel method for traffic safety assessment in urban expressway based on TTC distribution. The TTC data was collected from traffic videos in Beijing urban expressway, and a Gaussian mixture model (GMM) based TTC distribution was proposed for estimating TTC parameters value with a given traffic flow condition. A novel traffic safety indicator was presented considering the percentage below TTC threshold in dangerous condition. Finally, a valid and reliable assessment method of expressway traffic safety was established and TTC indicators in different scenarios were analyzed and compared.

The paper is organized as five sections, of which this is the first. The next section briefly describes collection of TTC data, the definition of TTC, and the basic statistical parameters of TTC. In Section 3, the GMM is built to establish the distribution of TTC. The Section 4 describes in detail the development and discussion of the novel assessment method of traffic safety using GMM based TTC distribution. The last section concludes with a summary of the findings of this study.

## 2. Data Collection

### 2.1. Study sites

The field data analyzed in this paper were collected through video survey between Si-tong Bridge and Lian-xiang Bridge in Beijing North Ring III expressway. Ring III expressway built overhead is a vital infrastructure in Beijing’s road system, and has a total length of 48 kilometers. The data were collected with different traffic conditions from 7:00 am to 10:00 am (including morning peak hour) and from 16:00 pm to 19:00 pm (including evening peak hour) on Jun. 20th

(Tuesday), 2006. Five locations are chosen as the studies sites. Location 1-5 are located on mainline at upstream of on-ramp (direction W→E), weaving segment at downstream of on-ramp (direction W→E), mainline at downstream of off-ramp (direction W→E), weaving segment at downstream of on-ramp (direction E→W), and mainline at downstream of off-ramp (direction E→W), respectively. Each location including three lanes named median lane, middle lane, and

shoulder lane are chosen for this paper. The locations of these stations are shown in the red rectangles on Figure 1. Using video image processing technology, it is easy to get traffic data including speed, traffic volume, and time occupancy. The video was Similarly, the dataset for calculating TTC is obtained including speed, headway, vehicle length, and classification of vehicle types.

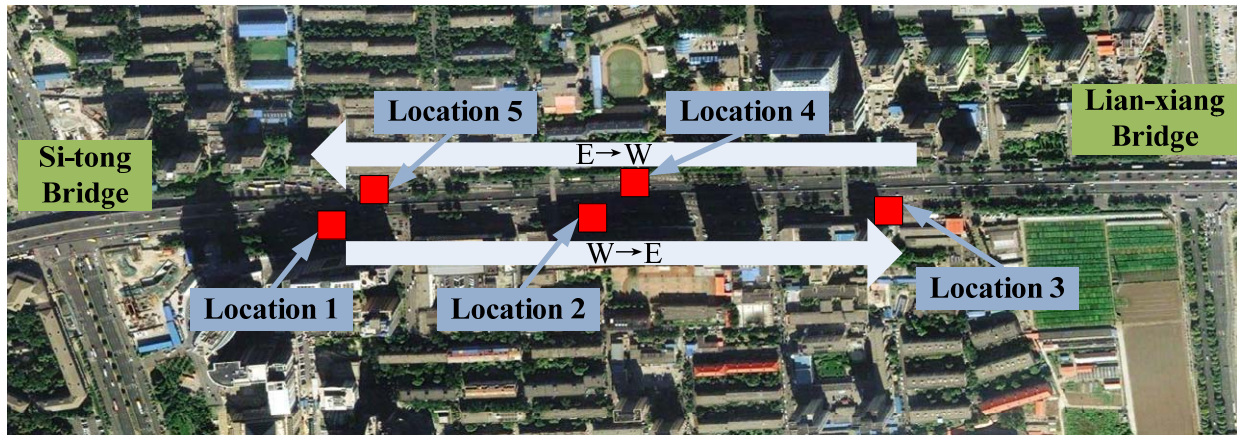


Fig. 1. Data collection sites from Si-tong Bridge to Lian-xiang Bridge in Beijing expressway.

## 2.2. Basic statistical parameters of TTC

TTC can be defined as the distance between a following and a leading vehicle divided by the relative velocity between two consecutive vehicles at a particular time. Thus, the formulation is shown as follows:

$$TTC_i(t) = \begin{cases} \frac{x_{i-1}(t) - x_i(t) - VL_{i-1}}{v_i(t) - v_{i-1}(t)} & \text{if } v_i(t) > v_{i-1}(t) \\ \infty & \text{otherwise} \end{cases} \quad (1)$$

where  $TTC_i(t)$  is the TTC of following vehicle  $i$  at time  $t$ ,  $x_{i-1}(t)$  and  $x_i(t)$  is the position of leading vehicle  $i-1$  and following vehicle  $i$  at time  $t$ , respectively,  $v_{i-1}(t)$  and  $v_i(t)$  is the speed of leading vehicle  $i-1$  and following vehicle  $i$  at time  $t$ , respectively, and  $VL_{i-1}$  is the vehicle length of leader  $i-1$ .

It is difficult for video to capture the positions and speeds of following vehicle and leading vehicle at a particular time, especially, the distance between two consecutive vehicles is large. Therefore, TTC should be calculated through fixed station traffic parameters. Assume that vehicles have a consistent travel speed through a fixed station in a short time interval, the

distance headway  $x_{i-1}(t) - x_i(t)$  while following vehicle through the detector can be estimated by follower speed multiplying time headway<sup>13</sup>. Then, the expression of TTC in car-following scenario will be rewrote as:

$$TTC_i = \frac{TH_i \cdot v_i - VL_{i-1}}{v_i - v_{i-1}} \quad \forall v_i > v_{i-1} \quad (2)$$

where  $TTC_i$  is the TTC of following vehicle  $i$  at the particular station,  $v_{i-1}$  and  $v_i$  is the speed of leading vehicle  $i-1$  and following vehicle  $i$  through the particular station, respectively, and  $TH_i$  is the time headway between two vehicles.

Some basic statistical results of five locations' TTC for different position, lane, and traffic flow level are shown in Table 1.

From Table 1, 11389 TTC samples with respect to different traffic volume and conditions were collected. Each lane has nearly 800 TTC samples for analysis. According to the results of Tables 1, we can find that: (a) The means of TTC samples on different lanes are within a relatively large range from 13.19 to 23.14 and TTC samples have great variances at different locations. (b) The means of TTC on different lanes are with large

differences. (c) The means of TTC at location 2 are also smaller than these at location 1 and location 3. In other words, traffic flow on weaving segment has more collision risks and leads to more dangerous situations. All of above results imply that means of TTC samples have strong correlations with locations, lanes, and

traffic conditions. Consequently, it is unrealistic to use an uniform TTC threshold value for assessment of short term traffic safety with different road environment and traffic conditions. The distribution characteristics of TTC have great effect on TTC threshold value and assessment of traffic safety.

Table 1. Descriptive statistics of TTC.

Study sites		Number of samples	Traffic Volume (veh./h/lane)	Mean (s)	SD (s)	Min (s)	Max (s)
Location 1 (upstream of on-ramp)	Median Lane	713	995	16.79	14.38	0.02	76.36
	Middle Lane	780	1245	17.20	12.73	0.24	65.46
	Shoulder Lane	835	1410	15.81	10.52	0.01	53.74
Location 2 (weaving segment)	Median Lane	951	1179	13.19	11.52	0.01	79.60
	Middle Lane	937	1294	15.75	11.93	0.03	64.44
	Shoulder Lane	885	1181	14.51	8.62	0.04	44.24
Location 3 (downstream of off-ramp)	Median Lane	660	1393	21.66	16.70	0.04	81.36
	Middle Lane	749	1467	14.96	9.74	0.04	47.64
	Shoulder Lane	799	1736	14.82	8.76	0.08	46.76
Location 4 (weaving segment)	Median Lane	753	821	15.78	11.90	0.10	63.64
	Middle Lane	736	1105	15.55	11.67	0.08	60.92
	Shoulder Lane	762	1205	16.24	10.55	0.01	54.96
Location 5 (downstream of off-ramp)	Median Lane	496	1546	23.14	17.68	0.09	89.56
	Middle Lane	692	1321	16.88	11.58	0.19	66.40
	Shoulder Lane	641	1722	17.34	12.21	0.05	66.60
All		11389	1308	16.34	11.79	0.01	89.56

### 3. Gaussian Mixture Model Distribution for TTC

Gaussian mixture model is a parametric probability density function represented as a weighted sum of Gaussian component densities. GMMs are commonly used as a parametric model of the probability distribution of continuous measurements. GMMs have been successfully used in a wide variety of fields, such as speaker recognition systems<sup>28, 29</sup>, video image processing<sup>30</sup>, and pattern classification<sup>31, 32</sup>.

#### 3.1. Gaussian mixture model

A Gaussian mixture model for TTC distribution is a weighted sum of  $M$  component Gaussian densities as given by the equation<sup>33</sup>,

$$P\{TTC_k | (\omega_i, \mu_i, \sigma_i^2)\} = \sum_{i=1}^M \omega_i g(TTC_k | \mu_i, \sigma_i^2) \quad (3)$$

where  $TTC_k$  is the TTC of the  $k^{\text{th}}$  vehicle,  $\omega_i$ ,  $i=1, \dots, M$ , are the mixture weights, and  $g(TTC_k | \mu_i, \sigma_i^2)$ ,  $i=$

$1, \dots, M$ , are the component Gaussian densities. Each component density is a one-variate Gaussian function of the form,

$$g(TTC_k | \mu_i, \sigma_i^2) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left\{-\frac{1}{2\sigma_i^2}(TTC_k - \mu_i)^2\right\} \quad (4)$$

with mean  $\mu_i$  and variance  $\sigma_i^2$ . The mixture weights satisfy the constraint that  $\sum \omega_i = 1$  and indicate the percentage of TTC belonging to category  $i$ . The values of the mixture weights are associated with safety situations of traffic flow. The complete Gaussian mixture model is parameterized by the mean, variance and mixture weights from all component densities. These parameters are collectively represented by the notation  $\Theta = (\omega_i, \mu_i, \sigma_i^2)$ .

There are several variants on the GMM shown in Eq. (1). The choice of model configuration (number of components and model parameters) is often determined by the amount of data available for estimating the GMM parameters and how the GMM is used in a particular application.

One of the powerful attributes of the GMM is its ability to form smooth approximations to arbitrarily shaped densities. Due to their capability of representing a large class of sample distributions, we applied GMM to analyze TTC distribution data and capture the component Gaussian distribution patterns for different traffic safety situations. The use of a GMM for representing feature distributions of TTC may also be motivated by the intuitive notion that the individual component densities may model some underlying set of hidden classes. For example, it is reasonable to assume the TTC values corresponding to different safety situations, such as dangerous situations, relative safe situations or absolute safe situations. These safety situations classes reflect some general accident risks that are useful for characterizing safety situation identity. The  $i^{\text{th}}$  TTC class can be represented by the mean  $\mu_i$  of the  $i^{\text{th}}$  component density and the variation  $\sigma_i^2$ . Thus, GMM can be used to group the TTC samples and describe the parameters of different classes.

### 3.2. Maximum likelihood parameter estimation

Given training samples and a GMM configuration, we wish to estimate the parameters of the GMM,  $\Theta$ , which in some sense best matches the distribution of the training samples. There are several techniques available for estimating the parameters of a GMM<sup>34</sup>. By far the most popular and well-established method is maximum likelihood (ML) estimation. The aim of ML estimation is to find the model parameters which maximize the likelihood of the GMM given the training data. For a sequence of  $N$  training data  $TTC_i$ , the GMM likelihood, assuming independence between the training samples  $TTC_i$ , can be written as,

$$p(TTC|\Theta) = \prod_{i=1}^N p(TTC_i|\Theta) = L(\Theta|TTC). \quad (5)$$

This function  $L(\Theta|TTC)$  is called the likelihood of the parameters given the data, or just the likelihood function. The likelihood is thought of as a function of the parameters  $\Theta$  where the data  $TTC$  is fixed. In the

maximum likelihood problem, the goal is to find the  $\Theta$  that maximizes  $L$ . That is, we wish to find  $\Theta^*$  where,

$$\Theta^* = \arg \max_{\Theta} L(\Theta|TTC). \quad (6)$$

Often we maximize  $\log[L(\Theta|TTC)]$  instead because it is analytically easier. Unfortunately, this expression is a non-linear function of the parameters  $\Theta$  and direct maximization is not possible. However, ML parameter estimates can be obtained iteratively using a special case of the expectation-maximization (EM) algorithm. The EM algorithm<sup>35</sup> is a general method of finding the maximum-likelihood estimate of the parameters of an underlying distribution from a given data set when the data is incomplete or has missing values. The basic idea of the EM algorithm is, beginning with an initial parameters  $\Theta$ , to estimate a new parameters  $\bar{\Theta}$ , such that  $p(TTC|\bar{\Theta}) > p(TTC|\Theta)$ . The new parameters then becomes the initial parameters for the next iteration and the process is repeated until some convergence threshold or iteration number is reached.

The EM algorithm first finds the expected value of the complete-data log-likelihood. The evaluation of this expectation is called the E-step of the algorithm, and the second step (the M-step) of the EM algorithm is to maximize the expectation computed in the first step. These two steps are repeated as necessary. Each iteration is guaranteed to increase the log-likelihood and the algorithm is guaranteed to converge to a local maximum of the likelihood function. The detailed descriptions of EM algorithm refer to Wu<sup>36</sup>, Redner and Walker<sup>37</sup>, Jordan and Jacobs<sup>38</sup>, and Jordan and Xu<sup>39</sup>.

GMM parameters are estimated from training data using the iterative EM algorithm. In this study, the number of component  $M$  is set as 3 on empirical observation and analysis. Three categories of TTC data represent dangerous situations, relative safe situations, and absolute safe situations, respectively. Multiple normal distributions are applied to fit TCC data and mixture weight of each distribution reflects the percentage of different safe situations of traffic flow.

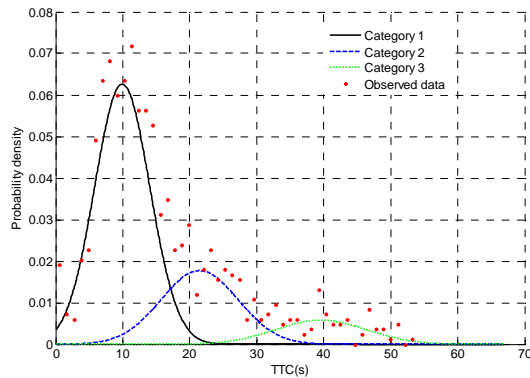


Fig. 2. GMM fits with TTC data on shoulder lane, location 1.

Figure 2 depicts three component Gaussian densities function and the observed data collected on shoulder lane, location 1. It is shown that the GMM has ability to fit two-peaks distribution of TTC data. Figure 3 shows the empirical cumulative distribution function (CDF) for TTC data samples with the fitted GMM distribution. The GMM fits to empirical data very well and has small error. To further verify the fitting results in statistically, Kolmogorov-Smirnov (K-S) test was adopted to perform the goodness-of-fit test<sup>40</sup>. In statistics, the K-S test is a nonparametric test for the equality of continuous one-dimensional probability distributions that can be used to compare a sample with a reference probability distribution (one-sample K-S test), or to compare two samples (two-sample K-S test). The K-S statistic quantifies a distance between the empirical

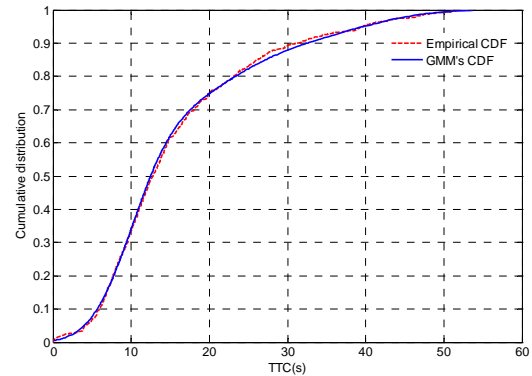


Fig. 3. Empirical CDF and GMM's CDF on shoulder lane, location 1.

distribution function of the sample and the CDF of the reference distribution. The null distribution of this statistic is calculated under the null hypothesis that the sample is drawn from the reference distribution.

Table 2 shows the estimated Gaussian distribution parameters and results of K-S test for sample data with all lanes of five location. As can be seen from Table 2, the K-S goodness of fit tests suggest the GMM performs well. All of samples from 15 stations are demonstrated that the samples are drawn from the GMM distribution statistically at significance level  $\alpha = 0.05$ . The estimated component weighs of different Gaussian distributions also indicate that percentages of TTC data in different traffic situations are related to sampling stations and traffic conditions. TTC data in dangerous situation is nearly range from 50% to 60%.

Table 2. Results of GMM distribution parameters and K-S test.

Study sites		Weights $\omega_i$	Means $\mu_i$	Variances $\sigma_i^2$	K-S value	Critical value	T-test at significance level $\alpha = 0.05$
Location 1	Median Lane	(0.545, 0.287, 0.168)	(8.2, 19.0, 40.7)	(21.0, 56.5, 248.2)	0.022	0.051	0
	Middle Lane	(0.614, 0.256, 0.130)	(9.7, 23.4, 40.6)	(23.5, 51.6, 140.4)	0.021	0.048	0
	Shoulder Lane	(0.646, 0.256, 0.099)	(9.9, 21.5, 39.6)	(16.9, 33.0, 46.1)	0.015	0.047	0
Location 2	Median Lane	(0.651, 0.272, 0.078)	(7.4, 19.0, 40.9)	(16.2, 50.5, 227.0)	0.025	0.044	0
	Middle Lane	(0.544, 0.289, 0.167)	(8.0, 19.4, 34.6)	(18.5, 41.0, 142.6)	0.025	0.044	0
	Shoulder Lane	(0.442, 0.326, 0.232)	(8.0, 14.9, 26.2)	(15.0, 20.1, 45.9)	0.032	0.045	0
Location 3	Median Lane	(0.599, 0.247, 0.154)	(11.7, 26.4, 53.0)	(29.6, 61.8, 190.5)	0.040	0.053	0
	Middle Lane	(0.561, 0.316, 0.123)	(8.8, 18.6, 33.5)	(14.8, 36.4, 58.9)	0.030	0.049	0
	Shoulder Lane	(0.510, 0.335, 0.155)	(8.7, 17.1, 30.2)	(9.2, 17.7, 55.4)	0.021	0.048	0
Location 4	Median Lane	(0.560, 0.319, 0.121)	(8.4, 19.5, 40.0)	(17.1, 41.1, 111.0)	0.031	0.049	0
	Middle Lane	(0.559, 0.335, 0.107)	(8.4, 19.6, 40.2)	(17.3, 45.6, 119.6)	0.024	0.050	0
	Shoulder Lane	(0.556, 0.355, 0.089)	(9.7, 20.8, 39.3)	(19.0, 40.9, 81.2)	0.028	0.049	0
Location 5	Median Lane	(0.612, 0.295, 0.093)	(12.6, 31.9, 64.8)	(38.5, 76.2, 153.2)	0.023	0.061	0
	Middle Lane	(0.558, 0.284, 0.158)	(10.3, 19.3, 35.8)	(22.4, 38.6, 176.8)	0.022	0.051	0
	Shoulder Lane	(0.522, 0.340, 0.138)	(9.3, 20.6, 39.6)	(16.5, 38.8, 154.2)	0.029	0.053	0

#### 4. Assessment of Traffic Safety using a Novel Safety Indicator

##### 4.1. Percentage of serious traffic conflicts

In Section 3, we use GMM based TTC distribution to divided TTC samples into three categories. Each TTC category  $i$  (e.g. dangerous situation) belongs to a normal distribution with parameters  $(\mu_i, \sigma_i^2)$  and the percentage of category  $i$  equals to weigh  $\omega_i$ .

We define percentage of serious traffic conflicts as the percentage of TTC samples below a predetermined threshold value in dangerous situation. Thus, the percentage of serious traffic conflicts can be quantified as follows:

$$L_{sc} = \omega_{ds} \cdot p_{ds}[TTC(x) \leq \tau] \quad (7)$$

where  $L_{sc}$  is the percentage of serious traffic conflicts,  $\omega_{ds}$  and  $p_{ds}()$  are the weigh factor and CDF of TTC samples in dangerous situations, respectively, and  $\tau$  is a predetermined TTC threshold value.

This indicator can be as a novel safety indicator for assessment of expressway safety. Having had the TTC distributions of expressway sections with different environments, the percentage of serious traffic conflicts can be calculated for safety assessment.

##### 4.2. Analysis of traffic safety

According to Eq. (7), the percentage of serious traffic conflicts in dangerous situations could be calculated as a safety indicator for assessment of expressway traffic safety in a specific time period. Using the field TTC data mentioned in Section 2, the percentages of serious traffic conflicts are reported in Table 3 where 2 s, 3 s, 4 s, and 5 s are considered as the TTC threshold values.

Table 3. Assessment results of expressway safety with different TTC threshold values.

Sites		TTC threshold values			
		2 s	3 s	4 s	5 s
Location 1	Median Lane	4.71%	6.87%	9.65%	13.04%
	Middle Lane	3.49%	5.19%	7.45%	10.31%
	Shoulder Lane	1.75%	2.98%	4.84%	7.47%
Location 2	Median Lane	5.77%	8.81%	12.81%	17.74%
	Middle Lane	4.47%	6.71%	9.64%	13.27%
Location 3	Shoulder Lane	2.62%	4.26%	6.55%	9.54%
	Median Lane	2.26%	3.32%	4.75%	6.60%

Location 4	Middle Lane	2.10%	3.60%	5.82%	8.89%
	Shoulder Lane	0.70%	1.56%	3.13%	5.73%
	Median Lane	3.33%	5.26%	7.91%	11.34%
Location 5	Middle Lane	3.41%	5.35%	8.00%	11.42%
	Shoulder Lane	2.21%	3.54%	5.42%	7.96%
	Median Lane	2.67%	3.72%	5.06%	6.74%
Location 5	Middle Lane	2.24%	3.46%	5.14%	7.37%
	Shoulder Lane	1.87%	3.13%	4.98%	7.52%

According to Tables 3, there are some important characteristics of expressway traffic safety can be found that:

- (1) With the increasing of TTC threshold values, the percentages of serious conflicts also increase. TTC threshold value has a great impact on assessment of safety and should be calibrated and validated by a large number of field data.
- (2) The percentage of serious traffic conflicts between different locations have a great difference. The percentages of serious traffic conflicts collected on the weaving segment are larger than these collected on mainlines. Furthermore, expressway traffic safety on the

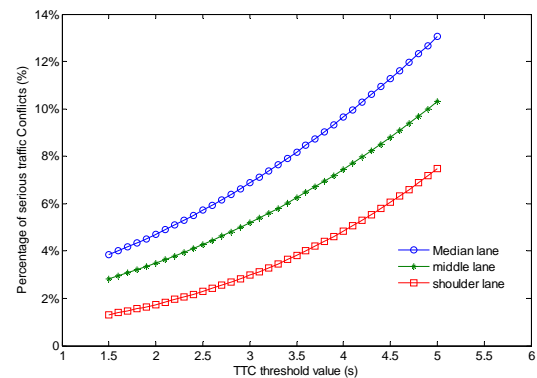


Fig. 4. TTC threshold values - percentages of serious traffic conflicts relationship.

downstream section of off-ramp is better than that on the upstream section of on-ramp. These results is mainly due to that traffic flow on the weaving segment is more unstable and a large number of lane changing behavior leading to less TTC value and more serious conflicts.

- (3) We further analyze the differences of expressway traffic safety between different lanes. Figure 4 shows the relationship between TTC threshold values and percentages of serious traffic conflicts with different lanes at location 1. From Figure 4, we can find that the percentage of serious traffic conflicts on median lane is larger than that on middle lane and shoulder lane. The



same results can also be found obviously at location 2-5. The main reason is that the on-ramp and off-ramp vehicles have greater effect vehicles on shoulder lane than these on median lane. Thus, vehicles on median lane have higher travel speeds and more serious traffic conflicts.

## 5. Conclusions

In this study, an assessment method of expressway traffic safety is proposed to estimate the vehicle crash frequency in different environments. TTC samples data was collected at 5 locations in Beijing expressway. Some basic descriptive statistics of 5 locations' TTC for different position, lane, and traffic flow level are shown and analyzed. We conclude that Gaussian mixture model distribution is the best-fitted distribution to TTC samples based on best-fit analysis and the K-S goodness of fit tests suggest the GMM performs very well. Accordingly, GMM is applied to establish the distribution of TTC samples under different traffic conditions. TTC samples are divided into three categories: dangerous situations, relative safe situations, and absolute safe situations, respectively. A novel concept of percentage of serious traffic conflicts is defined as the percentage of TTC samples below a predetermined threshold value in dangerous situation. Using this new safety indicator, assessment results of expressway traffic safety are presented. It is found that: (1) TTC threshold value has a great impact on assessment of safety. (2) expressway traffic safety on the weaving segment is lower than that on mainlines. (3) the percentage of serious traffic conflicts on median lane is larger than that on middle lane and shoulder lane. These conclusions are of great significance for analysis of expressway traffic safety and improvement of traffic safety.

## Acknowledgements

This study is supported by the National Natural Science Foundation of China (No. 70971053), the National High-Tech R&D Program (863) of China (No. 2011AA110304), and China Postdoctoral Science Foundation (No. 20100481419).

## References

1. Jeffery Archer., *Methods for the Assessment and Prediction of Traffic Safety at Urban Intersections and their Application in Micro-simulation Modeling*, (Doctoral Dissertation, Royal Institute of Technology, 2004, Stockholm, Sweden).
2. Miaou, S.P., and Lum, H., Modeling vehicle accidents and highway geometric design relationships, *Accident Analysis and Prevention*, **25**(6) (1993) 77–84.
3. Daniels, S., Brijs, T., Nuyts, E., Wets, G., Explaining variation in safety performance of roundabouts, *Accident Analysis and Prevention*, **42**(1) (2010) 393–402.
4. Guo, F., Wang, X., Abdel-Aty, M., Modeling signalized intersection safety with corridor spatial correlations, *Accident Analysis and Prevention*, **42**(1) (2010) 84–92.
5. Haque, M.M., Chin, H.C., Huang, H., Applying Bayesian hierarchical models to examine motorcycle crashes at signalized intersections, *Accident Analysis and Prevention*, **42**(1) (2010) 203–212.
6. Lord, D., Mannering, F., The statistical analysis of crash-frequency data: A review and assessment of methodological alternative, *Transportation research Part A*, **44**(5) (2010) 291–305.
7. Wang, W.H., Jiang, X.B., Xia, S.C., Cao, Q., Incident tree model and incident tree analysis method for quantified risk assessment: an in-depth accident study in traffic operation, *Safety Science*, **48**(10) (2010) 1248–1262.
8. Svensson, A., *A method for analyzing the traffic process in a safety perspective*, (Doctoral Dissertation. University of Lund, 1998, Lund, Sweden).
9. Michael, P.G., Leeming, F.C., Dwyer, W.O., Headway on urban streets: observational data and an intervention to decrease tailgating, *Transportation Research Part F*, **3**(2) (2000) 55–64.
10. Perkins, S.R., and Harris, J.I., Traffic Conflict Characteristics: Accident potential at intersections, *Highway Research Record*, **225** (1968) 45–143.
11. Hayward, J.C., *Near miss determination through use of a scale of danger* (traffic records 384), Highway Research Board. Washington, D.C., 1972.
12. Van Arem, B., and De Vos, A.P., *The Effect of a Special Lane for Intelligent Vehicles on Traffic Flows*, TNO-INRO Report 1997-02a, Delft, Netherlands, 1997.
13. Vogel, K., A comparison of headway and time to collision as safety indicators, *Accident Analysis and Prevention*, **35**(3) (2003) 427–433.
14. Minderhoud, M. M., and Bovy, P. H. L., Extended Time-to-Collision Measures for Road Traffic Safety Assessment, *Accident Analysis and Prevention*, **33**(1) (2001) 89–97.
15. FHWA, *Surrogate Safety Measures from Traffic Simulation Models*, Final Report, Publication No FHWA-RD-03-050, Federal Highway Administration, USA, 2003.



16. Vogel, K., *Modeling Driver Behavior – A Control Theory Based Approach*, (Doctoral Dissertation. Institute of Technology, University of Linköping, 2002, Sweden).
17. Farah, H., Bekhor, S., and Polus, A., Risk evaluation by modeling of passing behavior on two-lane rural highways, *Accident Analysis and Prevention*, **41**(4) (2009) 887–894.
18. Hydén, C., *The Development of a Method for Traffic Safety Evaluation: The Swedish Traffic Conflicts Technique*, (Doctoral Dissertation. University of Lund, 1987, Lund, Sweden).
19. Janssen, W.H., Nilsson, L., *An experimental evaluation of in-vehicle collision avoidance systems*, In: Proceedings of the Paper Presentation at the 24th ISATA International Symposium on Automotive Technology and Automation, 1991.
20. Van Winsum, W., Heino, A., Choice of time-headway in car-following and the role of time-to-collision information in braking, *Ergonomics* **39**(4) (1996) 579–592.
21. Hirst, S., Graham, R., *The format and presentation of collision warnings*, In: Noy, I.Y. (Ed.), *Ergonomics and Safety of Intelligent Driver Interfaces*, Lawrence Erlbaum, Mahwah, NJ., 1997.
22. Van Winsum, W., The human element in car following models, *Transportation Research Part F*, **2**(4) (1999) 207–211.
23. Jin, S., Wang, D.H., and Yang, X.R., Non-lane-based car following model using visual angle information, *Transportation Research Record: Journal of the Transportation Research Board*, 2011.
24. Jin, S., Huang, Z.Y., Tao, P.F., Wang, D.H., Car following theory of steady-state traffic flow using time-to-collision, *Journal of Zhejiang University-SCIENCE A (Applied Physics & Engineering)*, **12**(8) (2011) 645–654.
25. Bubb, H., Traffic Safety through Driver Assistance and Intelligence, *International Journal of Computational Intelligence Systems*, **4**(3) (2011) 287–296.
26. Maretzke, J., Jacob, U., *Distance warning and control as a means of increasing road safety and ease of operation*, In: Proceedings of the Paper Presentation at the FISITA'92: Safety, the Vehicle and the Road. XXIV FISITA Congress, London, 1992.
27. Hogema, J.H., Janssen, W.H., *Effects of intelligent cruise control on driving behavior*, TNO Human Factors, Soesterberg, The Netherlands, Report TM-1996-C-12, 1996.
28. Hsieh C.T., Lai E., Wang Y.C., Robust speaker identification system based on wavelet transform and Gaussian mixture model, *Journal of Information Science and Engineering*, **19**(2) (2003) 267–282.
29. Xiong, Z.Y., Zheng, T.F., Song, Z.J., Soong, F., Wu, W.H., A tree-based kernel selection approach to efficient Gaussian mixture model-universal background model based speaker identification, *Speech Communication*, **48**(10) (2006) 1273–1282.
30. Stauffer, C., Grimson, W.E.L., *Adaptive background mixture models for real-time tracking*, in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, **2**(1999) 246–252.
31. Zhou, X., and Wang, X., Optimisation of Gaussian mixture model for satellite image classification, *IEEE Proceedings-Vision Image and Signal Processing*, **153**(3) (2006) 349–356.
32. Kim, S.C., Kang, T.J., Texture classification and segmentation using wavelet packet frame and Gaussian mixture model, *Pattern Recognition*, **40**(4) (2007) 1207–1221.
33. Titterton, D., Smith, A., and Makov, U., *Statistical Analysis of Finite Mixture Distributions*, John Wiley & Sons, 1985.
34. McLachlan, G., *Mixture Models*. Marcel Dekker, New York, NY, 1988.
35. Dempster, A.P., Laird, N.M., Rubin, D.B., Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society*, **39**(1) (1977) 1–38.
36. Wu, C.F.J., On the convergence properties of the EM algorithm, *The Annals of Statistics*, **11**(1) (1983) 95–103.
37. Redner, R. and Walker, H., Mixture densities, maximum likelihood and the EM algorithm, *SIAM Review*, **26**(2) (1984) 195–237.
38. Jordan, M. and Jacobs, R., Hierarchical mixtures of experts and the EM algorithm, *Neural Computation*, **6**(2) (1994) 181–214.
39. Jordan, M. and Xu, L., Convergence results for the EM approach to mixtures of experts architectures, *Neural Networks*, **8**(9) (1996) 1409–1431.
40. Stephens, M.A., EDF Statistics for Goodness of Fit and Some Comparisons, *Journal of the American Statistical Association*, **69**(347) (1974) 730–737.