Survival Analysis of Urban Traffic Incident Duration: a Case Study at Shanghai Expressways

Yicheng Shi[1] Lun Zhang[2] Pei Liu[3]

[1] Key Laboratory of Road and Traffic Engineering of Ministry of Education, TONG JI University
Shanghai 201804, China
tongjisyc@163.com

[2] Key Laboratory of Road and Traffic Engineering of Ministry of Education, TONG JI University
Shanghai 201804, China
Lun_zhang@tongji.edu.cn

3Department of Transportation Technology and Management, Feng Chia University
Taichung 407, Taiwan, ROC
peiliufcu@gmail.com

Received 21 July 2014; Revised 19 September 2014; Accepted 19 December 2014

Abstract. Traffic incident duration is one of the most important parameters to describe traffic congestion intensity of expressways. Numerous measures have been developed to describe the characteristics of traffic incidents and the vast majority of these studies use data mining. Though they inform us the objective elements of the environment may influence traffic incident, the how question-via what way these physical features affect traffic incident-is largely unexplored. Based on the historical traffic data of viaduct expressways in Shanghai, this paper introduces survival analysis into the analysis of traffic incident mechanism, and a survival Analysis-Based Modeling of urban traffic incident duration is presented. This model first analyzes the time attributes of many traffic incident samples, and employs nonparametric regression based on Kaplan-Meyer model to estimate hazard-based traffic incident duration. Then, the key influence factors of traffic incident are divided into five types and the spatial-temporal distribution characteristics of traffic incident duration time are analyzed. Finally, COX regression is used to model the co-evolution between multidimensional influencing factors of traffic incident duration. The key characteristic parameters of expressway incident management in Shanghai are optimized to analyze the evolution mechanism of incident duration. The result shows that, for different type of influencing factors, the spatial-time distribution of traffic incident duration in Shanghai expressway exists significant difference, and factors like day & night, incident type, related vehicle number, related lane number, location, bottleneck and trailer will affect the incident duration significantly.

Key words: traffic incident, survival analysis, incident duration, COX regression, expressway

1 Introduction

Expressways are the crucial part of urban transportation system. In Shanghai, average daily traffic volume of urban expressway in 2009 is 28.8 million pcu kilometers, occupying 45% of the total volume [1]. As the urbanization and mechanization in China is getting faster, facilities and population are highly clustered in metropolises. The total possession of civil vehicles in Shanghai is more than 2 million and the total resident trips are up to 47 million. Presently, increasing traffic congestion has already become the major challenge of urban road traffic management that big cities all around China have to face [2]. Like the highway, urban expressway has no intersection and, generally, no signal control. However, expressways with no emergency strip have serried ramps, and always be in high-flow and high-density. All these features make non-recurrent traffic incidents happen randomly, and thus the traffic incident is hard to control and easy to cause large-scale traffic congestion. Statistics data in other countries show that the number of non-recurrent traffic congestion caused by traffic incidents is up to 60% of the total congestion [3]. It is necessary to use a statistical model to analyze the traffic incident duration data. If the formation mechanism and dissipation characteristics of incidents in urban expressways were learned, the spatial and temporal distribution of traffic incident can be analyzed, and then the crucial incident spots can be recognized, so as to provide decision supports to the intelligent management of urban transportation network. And this is the key to solve urban traffic congestion.

Presently, numerous measures have focus on describing serious accidents on highways or urban ground roads. There is, however, little work on traffic incident of urban expressways. For the studies on traffic incident duration models, Boyles *et al*. [4] propose a decision tree-based algorithm to predict traffic incident duration, and the experiment results show this model has good robustness. Zografos *et al.* [5] present an incident duration model based on hazard analysis. The studies show that the detection time, report time and response time of incident are consistent with Weibull model, while the clearing time is consistent with Log-logistic model. Garib *et al.* [6] develop a linear regression model to discuss traffic incident duration. And the results showed that incident features are represented by a series of binary and continuous variables. Kim *et al*. [7] employ fuzzy logic model to distribe incident duration , and refine the fuzzy sets to improve the prediction accuracy. All of these studies explain the process of traffic incidents by defined thresholds or pattern recognition. However, although they inform us the formation and propagation of traffic incidents in a large scale perspective, there are few quantitative analysis approaches to model the inner relationship between influencing factors of incident duration.

Survival analysis can reveal the characteristics of time variation in both qualitative and quantitative ways, and has a good ability to deal with censored data. This model can meet the inherent requirements of temporal and spatial distribution analysis of urban road traffic data [8]. It is widely used in medical area to research the efficiency of one therapy method, or the effect of one virus. Recent years, researchers have realized that the survival analysis can also be applied into the data mining of transportation, to describe the reasons of traffic phenomenon. Chung [9] has analyzed multiple factors of accident duration with censored data by using survival analysis model. Jovanis et al. [10, 11] develop a prediction model of traffic incident duration with survival analysis. Doohee et al. [12] apply hazard-based models to statistically evaluate the time the incidents takes to report, respond and clear in Washington. Zhou et al. [13] use survival analysis to analyze the spatial and temporal distribution of traffic congestion duration in Beijing. The vast majority of these studies have used semi-parametric model and accelerated failure time model to describe the relationship between multi-influencing factors of incident duration. However, these models haven’t separately modeled a single factor to the incident duration, and quantitative analysis of multi-factors of traffic incident as incident duration is not explored.

Because of its abnormal distribution and censored feature of traffic incident duration, the classical multivariate regression cannot accurately model these influencing factors. COX proportional hazard model, proposed by D.R.COX in 1972, can solve the above problems perfectly. This model is widely used in biomedical domain. New applications in the domain of transportation have been employed these years, such as pedestrian crossing waiting time [14], flight delay [15], pavement fatigue failure [16], accident duration in highway [12], pedestrian activity duration [17] and motorcycle age [18]. Note that Kang et al. [19] present a COX regression model to discuss the distribution regularities of traffic incident duration. However, they do not take the influencing factors of road environment under occurring incident into account. To find the formation mechanism and propagation characteristics of traffic congestion in urban expressways, single factor of incident duration should be analyzed both qualitatively and quantitatively, while multi-influencing relationship between these factors in complicated environment should also be modeled.

Along the line of previous studies, to address the critical issues above, this paper introduces survival analysis to model influencing factors of traffic incident duration. Two methods, Kaplan-Meyer and COX regression are applied to research these influencing factors. They offer an effective survival analysis-based process to quickly capture the space and time characteristics of traffic incident duration. The result will filter the influencing factors and leave us those critical ones. The remainder of this paper is organized as follows: Section II describes the traffic incident data in Shanghai expressways. Section III defines the survival analysis of traffic incident duration, and illustrates the data processing flow. Section IV develops the Kaplan-Meyer nonparametric model of incident duration to analyze the spatial and temporal distribution characteristics. Section V employs COX-based proportional hazard model to analyze the co-evolution between multidimensional influencing factors. Section IV concludes the work.

2 Data description

2.1 Traffic incident data source

The traffic incident data used in this paper is collected from the Road Traffic Real-time Information Monitoring System in Shanghai Traffic Information Center. As shown in Figure 1, the traffic incident data at 150 detected segments, covering the Inner Ring Viaduct (IRV) and most of the South-North Viaduct (SNV), are used to generate historical samples. The total length of the detected roads is 130km long, in which the IRV is 90km and the SNV is 40km. The IRV locates around the city center of Shanghai, while the SNV is the major road that runs through the municipality in the north-south direction. These two main expressways are both featured by many black spots of traffic congestion and traffic incidents.

Total 7354 samples of basic traffic incident data are collected from Nov. 2012 to Mar. 2013. With data pre-processing of rejecting the abnormal and duplicate data, there are 7203 valid examples. Each sample contains the information of incident duration, which reflects the details from different views. They are: incident time information, such as time, day, month, peak-time period; incident attribute information, such as incident type, involved vehicle number, incident zone, involved lane number, vehicle type; incident environment information, such as incident location, number of lanes, bottleneck states, weather condition, traffic state; incident rescue information, such as trailer state, ambulance state, fire engine state.

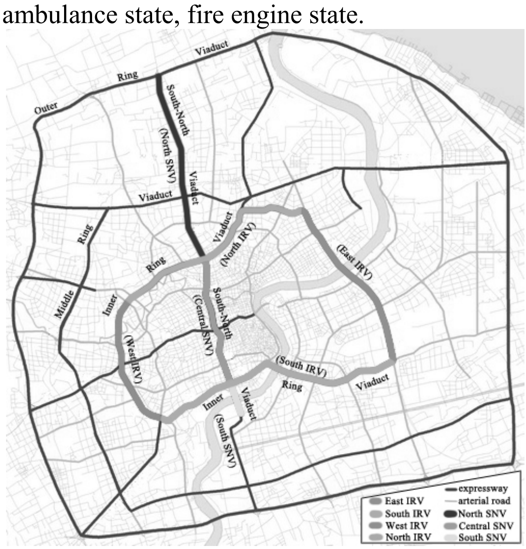


Fig. 1. Detected segments of Inner Ring and South-North Viaduct, Shanghai

2.2 Feature analysis of traffic incident data

Distribution feature of traffic incident data can be used to reveal the formation mechanism of traffic incidents and also offer basic-consideration for optimizing influencing factors of traffic incidents. To facilitate the model presentation, the notations used here-after and their statistics results from the above traffic incident samples are summarized in Table 1.

Table. 1. Summary of collected traffic incident data

Incident variables		value	Incident variables		value
Time variable	days	1, 2, 3, 4, 5, 6, 7	Rescue variable	trailer state	0=no, 1=yes
	peak-time period	0=off-peak hour, 1=peak hour (7am-10am, 4pm-7pm)		ambulance state	0=no, 1=yes
	day & night	0=day，1=night		fire engine state	0=no, 1=yes
Attribute variable	incident type	1=single-car breakdown, 2=single-car collision, 3=two-car collision, 4=multi-car collision, 5=goods drops, 6=others	Environment variable	incident location	1=East IRV, 2=South IRV, 3=West IRV, 4=North IRV, 5=North SNV, 6=Central SNV, 7=South SNV
	involved vehicle number	1, 2, 3+		number of lanes	2, 3, 4
	incident zone	1=middle lane, 2=side lane, 3=occupying lanes, 4=others		bottleneck state	0=non bottleneck, 1=bottleneck
	involved lane number	0, 1, 2, 3+		weather condition	1=no rain, 2=rain
	vehicle type	0=car, 1=freight		traffic state	0=smooth, 1=busy, 2=congested

3 Survival analysis based modeling of traffic incident duration

3.1 Definition

A general definition of survival analysis is that it is a statistical analysis approach to model the survival time of a specific event. According to the basic principles of survival analysis, the survival analysis of traffic incident duration is defined, particularly regarding following four elements.

(1) Survival time of traffic incident

Survival time is generally defined as the duration of a given event. And hence, the survival of traffic incident is the duration from the occurrence of a traffic incident to its end, when the road is cleaned up and traffic flow is recovered.

(2) Censored data of traffic incident duration

Compared with the complete data, censoring occurs when we have some information about individual survival time, but we don't know the survival time exactly. Censored data may occur for the following reason: 1) the object does not experience the event before the study ends; 2) the object is lost to follow-up during the study period; 3) the object withdraws from the study because of some other reasons.

(3) Survival function of traffic incident

The survival function, denoted by S(t), gives the probability that a traffic incident survives longer than some specified time t, that is, S(t) gives the probability that the random variable T exceeds the specified time t. The survival function is fundamental to a survival analysis, because obtaining survival probabilities for different values of t provides crucial summary information from survival data. The definition of S(t) is in equation (1).

$$S(t) = P(T > t) = \int_t^{\infty} f(x)dx \qquad \textbf{(1)}$$

In equation (1), the T is the duration of a traffic incident; f(t) is probability density function of traffic incident duration T; it shows that the S(t) is cumulative survival possibility. A sharp survival curve of S(t) means a low survival possibility, and vice versa.

(4) Hazard function

The hazard function, denoted by h(t), gives the instantaneous potential per unit time for the event to occur, given that the individual has survived up to time t. The equation of hazard function is shown in equation (2).

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t \leqslant T < t + \Delta t \mid T \geqslant t)}{\Delta t} = \frac{f(t)}{S(t)} = -\frac{d}{dt} S(t) \qquad \textbf{(2)}$$

In contrast to the survival function, which focuses on not failing, the hazard function focuses on failing, that is, on the event occurring. Thus, in some sense, the hazard function can be considered as giving the opposite side of the information given by the survival function. If the value of h(t) is bigger, the possibility that a traffic incident end in per unit time is higher.

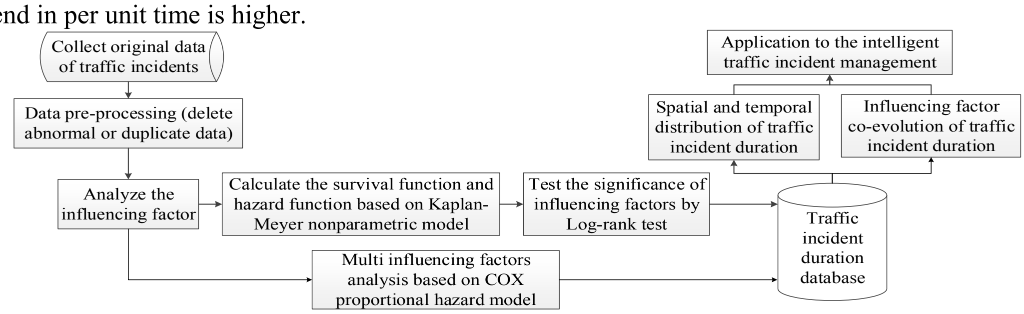


Fig. 2. Data processing flow of survival analysis based traffic incident management

3.2 Data processing flow

The data processing flow of survival analysis-based urban traffic incident management is shown in Figure 2, particularly regarding following steps:

(1) With the original data of traffic incidents, it just contains the real-time incident state at the incident spot. Thus, these data must be pre-processed to delete the abnormal or duplicate data from original database. And by associating those discrete data of each traffic incident at some time-step detection, evaluate the censoring status of traffic incident data and generate the valid sample of traffic incident duration with its influencing factors.

(2) Base on the sample data, analyze separately the multi-influencing factors of traffic incident duration, and employ the Kaplan-Meyer nonparametric model to calculate the survival function and hazard function of these factors respectively, while the significance of each influencing factors is tested through Log-rank test. Then the spatial and temporal distribution characteristics of traffic incident duration can be described.

(3) Use the COX proportional hazard model to analyze the co-evolution mechanism between multidimensional influencing factors of incident duration, and then the key characteristic parameters of expressway incident management can be optimized.

4 Kaplan-Meyer nonparametric-based model of traffic incident duration

4.1 Kaplan-Meyer based modeling

Kaplan-Meyer (K-M) model is a nonparametric model. If the K-M model is integrated in modeling traffic incident duration, there is no necessary assumption to the distribution of traffic incident duration, and the survival functions and hazard functions can be estimated directly. And hence, the spatial and temporal distribution characteristics of traffic incident can be observed. Denote T_i represents the traffic incident duration of the i^{th} sample, and the time-series sample set T satisfies the condition where $T_1<T_2<\ldots<T_n$, then the K-M based survival probability of traffic incident duration, denoted by $\hat{S}(t)$, is shown in equation (3).

$$\hat{S}(t)=\prod_{T_i^c\leq t}\frac{n-i}{n-i+1} \quad \textbf{(3)}$$

In equation (3), T_i^c represent the traffic incident duration of the i^{th} complete samples. Denote complete samples set as T^c, and $T_i^c \in T^c$. A complete sample must satisfy two following conditions: first, the T_i^c is a positive integer and less than t, that is $T_i^c \leq t,\ and\ T_i^c \in Z$; second, this sample is not a censored data. Thus, if all the observed samples are not censored, the $T^c = T$, otherwise discrete $T^c \subset T$.

4.2 Spatial and temporal distribution characteristics

(1) Overall distribution characteristics

Following the survival analysis based data process flow, SPSS PASW Statistics 18.0 is used to develop the survival analysis based model of traffic incident duration. It shows that in the IRV and SNV of Shanghai, the minimum incident duration is 1 minute, while the maximum is 127 minutes and the average is 9 minutes. The overall survival function (SF) and hazard function (HF) of the traffic incident duration in expressways is shown in Figure 3. It shows half of the traffic incident can be disposed in 6 minutes, while 96% of the traffic incident can be disposed in half an hour. In the studied five months, the number of those serious incidents, whose duration is more than 100 minutes, is no more than 13. The hazard value of those incidents cleared in an hour is less than 0.2, which indicates that the incidents occurred in expressways of Shanghai are mostly slight and can be disposed quickly. Following the findings above, they inform us that although most of Shanghai expressways are operating in fine conditions and most the incidents may be cleaned up in a short time, once an incident occurs in a certain situation, it may cause a serious traffic jam, even last for more than 2 hours.

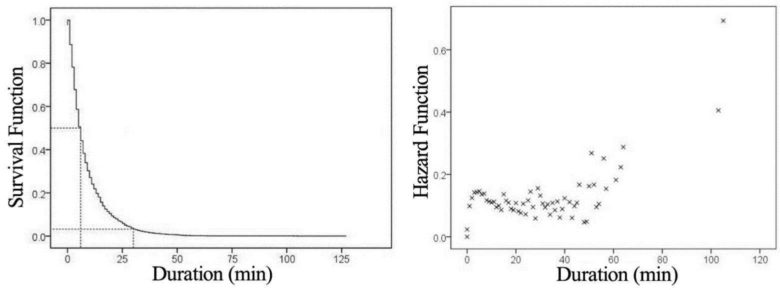


Fig. 3. Overall distribution of traffic incident duration in Shanghai expressways

(2) Temporal distribution characteristics

1) day

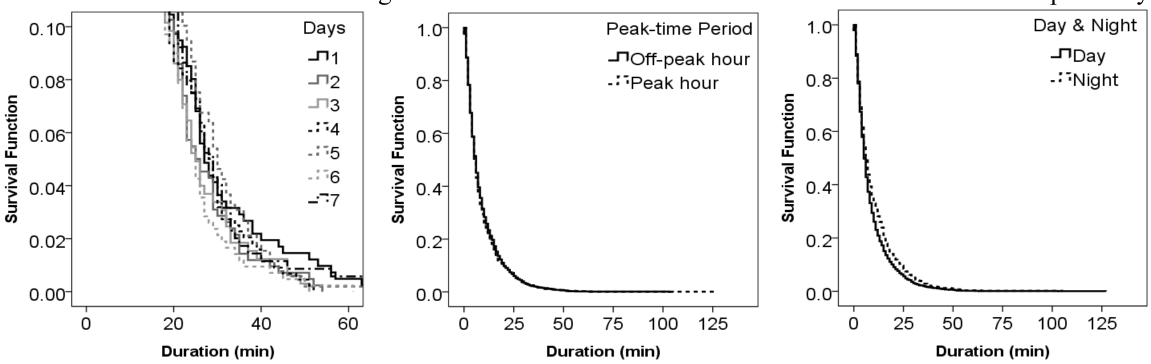
According to the different days of a week, the total sample set of traffic incident duration is divided into seven subsets from Monday to Sunday. The day-based survival function and cumulative hazard function is shown in Figure 4-a and Figure 4-b. It shows that there is little difference between the distributions of traffic incident duration in weekdays and weekends. Particularly, the incident duration in Saturday is shorter, while in Monday and Friday, the incident duration is relatively longer, and the probability that incident duration lasted more than half an hour is much higher. This may be because that both the traffic state in Monday and Friday is busy due to the increasing traffic volume for work business in the city and weekend holiday out of the city. And hence once traffic incident happened in Monday or Friday, it is prone to induce more serious traffic congestion in expressways.

2) peak-time period

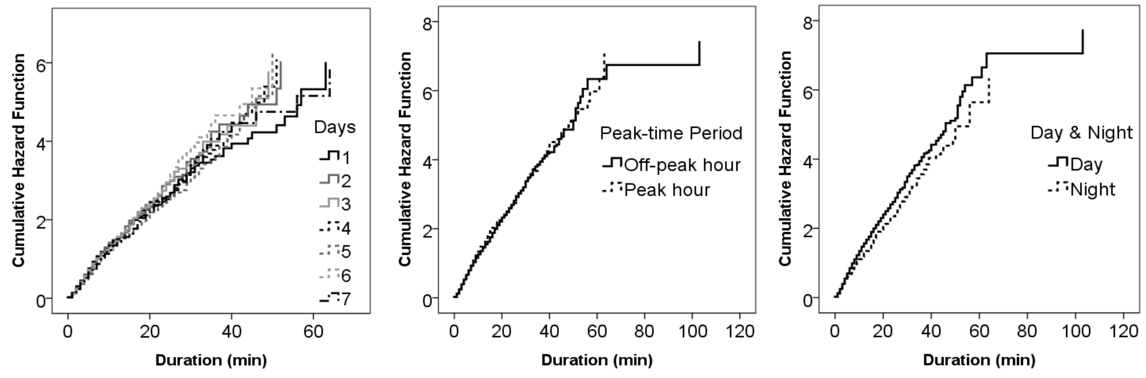
According to peak-time period of a day, the total set of traffic incident duration is first divided into two subsets by peak hour and off-peak hour. The peak-time period-based survival functions and cumulated hazard function (CHF) is shown in Figure 4-c and Figure 4-d. By using the Log-rank test to calculate the significance of difference between peak hour and off-peak hour, the result is 0.322, that is, the survival function of these two states is almost the equivalent. And hence it indicates that the peak-time period when a traffic incident occurs has little contribution to the duration of this traffic incident.

3) daytime and night

Like peak-time period-based analysis, according to the hours of the daytime and night, the total sample set of traffic incident duration is divided into two subsets by daytime (6am - 6pm) and night (6pm - 6am). The significance value of the difference between the daytime and night calculated by the Log-rank test is 0.001, that is, the shifting light change in the daytime and night have great significance to the traffic incident duration. And hence those traffic incidents occurred in the night are more serious and need more time to be cleared form expressways.



(a) day-based SF (c) peak hour-based SF (e) time period-based SF



(b) day-based CHF (d) peak hour-based CHF (f) time period- survial CHF

Fig. 4. Temporal distribution of traffic incident duration in Shanghai expressways

(3) Spatial distribution characteristics

1) incident location in the IRV

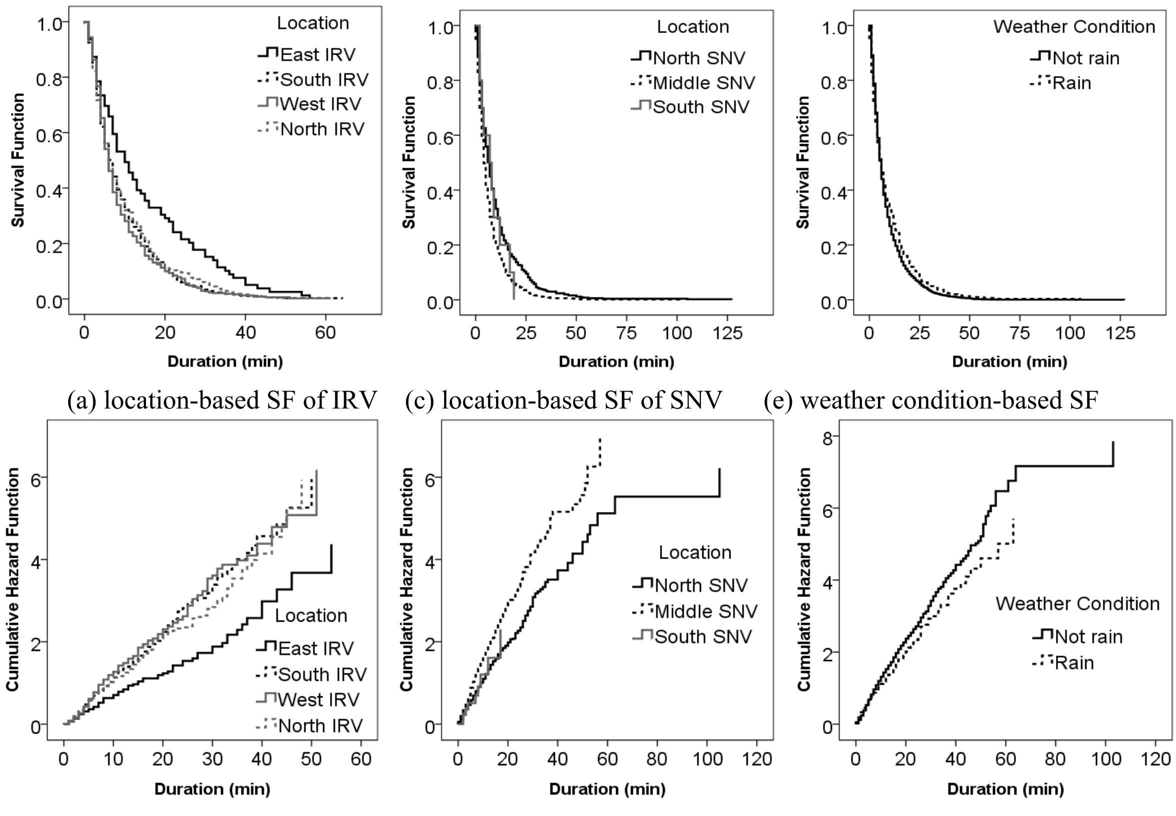
According to the different incident location in the IRV of Shanghai, the total incident duration samples are divided to four subsets: east IRV, south IRV, west IRV, and north IRV, as Figure 1 shows. The location-based survival function and cumulative hazard function of the IRV is shown in Figure 5-a and Figure 5-b. It shows that the survival curve of the east IRV is obviously above those of other segments, while the hazard curve of the east IRV is correspondingly below. These findings indicate that compared with other segments of the IRV, the incident duration in the east IRV is much longer, and the possibility that traffic incidents in the east IRV last longer time is higher. Mapping the east IRV into the Shanghai maps, it can be found that the east IRV is the major arteries of east Shanghai to connect the Yangpu District and Pudong New District, where there are large amounts of residential communities. And hence both the long and short distance trips between these two districts are quite frequent. This leads to the heavier traffic volume of this segment and increases the possibility of potential serious traffic incidents.

2) incident location in the SNV

Like the IRV, the incident duration samples from the SNV of Shanghai can be divided to three subsets: north SNV (north segment from the IRV to the Outer Ring Verdict), central SNV (inner segment between the IRV) and south SNV (south segment from the IRV to South Shanghai). The location-based survival function and cumulative hazard function of the SNV is shown in Figure 5-c and Figure 5-d. Due to few valid samples in the south SNV, this study did not take the south SNV into account in following analysis. The significance value of the difference between the north SNV and central SNV by Log-rank test is 3.45E-12, which is much smaller than 0.05. This still indicate that the incident duration distribution between the north SNV and central SNV is obviously difference. Compared with those of the central SNV, the probability that traffic incidents last longer time in the north SNV is higher. This is because that the number of lanes in the north SNV is less than those of the central SNV. In addition, the ramp deployment between the segments from the Middle Ring Verdict to the IRV is unreasonable because there are too many on ramps and off ramps in these segments.

3) weather condition

According to the different weather condition when traffic incident occurred in the expressways, the total incident duration samples are divided to two subsets: rain and no rain. The weather condition-based survival function and cumulative hazard function of incident duration samples is shown in Figure 5-e and Figure 5-f. It shows that the survival curve of rain is slightly above that of no rain. By Log-rank test, the significant value of the difference between these two weather condition states, is 0.09, which indicate that these two survival curves have no statistical difference, and the incident duration distribution the between rain and no rain is almost the same. In addition, it is worth noting that the incidents that lasted more than an hour almost occurred in rainy days. This is because that the visibility of drivers and the brake of vehicles are sharply deteriorated under the severe weather, which leads to slow running speed and frequent traffic congestion on expressways. Once a serious traffic incident occurs in this bad weather condition, the rescues may be delayed due to bad operating conditions of expressways, and hence the duration of a serious traffic incident in severe weather condition may be longer.



(a) location-based SF of IRV (c) location-based SF of SNV (e) weather condition-based SF

(b) location-based CHF of IRV (d) location-based CHF of SNV (f) weather condition-based CHF

Fig. 5. Spatial distribution of traffic incident duration in Shanghai expressways

5 COX proportional hazard-based model of traffic incident duration

5.1 The COX regression-based modeling

The COX regression is generally used for analyzing survival data and calculating the regression coefficient of multi-influencing factors, denoted by β_i. Different from the traditional regression analysis, COX regression doesn't use survival time t as the dependent variable of the regression equation, and each influencing factor is regarded as a covariate variable x. The quantitative contribution of each covariate variable to the survival time is described by the ratio of hazard function $h(t,x)$ to baseline hazard function $h_0(t)$, where the $h_0(t)$ is the inherent hazard function under the condition of no influencing factors. And hence the COX regression is also known as COX proportional hazard model. The COX regression has good suitability and could be used in the co-evolution analysis of multi-influencing factors of traffic incident duration, without any assumption of the distribution of survival time. The general COX regression model is in equation (4).

$$h(t,x) = h_0(t)\exp(\beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_i x_i) \tag{4}$$

With the logarithmic transformation in equation (4), the COX regression model is transformed into equation (5).

$$\ln\frac{h(t,x)}{h_0(t)} = \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_i x_i \tag{5}$$

Denote the relative risk by $RR=\frac{h(t,x)}{h_0(t)}$, then, the COX regression is the linear model of the logarithm of the *RR*. Under other covariate variables remaining constant, β_i describes that the logarithm change of the *RR* with the unit change of the i^{th} covariate variable. Based on the definition above, the COX regression has following properties:

1) If β_i>0, it means that the i^{th} variable is a risk factor and its hazard may be higher with the increasing time. And this indicates that the incident may be disposed quickly. 2) If β<0, it means this variable is a protective factor, and the traffic incident duration is longer, which indicates the incident cannot be disposed in time. 3) If β=0, it means this variable has nothing to do with the traffic incident duration.

5.2 Co-evaluation model of multi-influencing factor

The Log-rank test is used to evaluate the significance of those influencing factors in Table 1. By rejecting those factors that their significance satisfy the condition that P>0.05, the 10 influencing factors are remaining, each of which has significant effect on traffic incident time separately. These factors are *daytime & night*, *incident type*, *involved vehicle number*, *incident zone*, *involved lane number*, *incident location*, *number of lanes*, *bottleneck state*, *traffic state*, and *trailer state*. Three factors are not considered in this model for the unbalanced distribution of their parameters. They are *vehicle type*, *ambulance state* and *fire engine state*.

The COX regression above is employed to evaluate the co-evolution among these 10 influencing factors of traffic incident duration. This model uses the backward method to select significant covariate variables and estimates its regression coefficient filter the maximum-likelihood ratio. Though one influencing factor has great significance to the traffic incident duration separately, if taking other factors into account, it may have little effect under the combination of these multi-influencing factors. The COX regression based parameter estimation results is shown in Table 2. It shows that only 9 influencing factors are remaining in the COX regression model of traffic incident duration, and they are: daytime & night, incident type, involved vehicle number, involved lane number, incident location, bottleneck state and trailer state.

Table. 2. The estimated Parameters of COX Regression model

Influencing Factors	β	SE	Wald	P	exp(β)
x_1 daytime & night	-0.221	0.047	21.789	<0.001	0.802
x_2 incident type	0.278	0.065	18.348	<0.001	1.321
x_3 involved vehicle number	-0.527	0.121	19.037	<0.001	0.590
x_4 involved lane number	-0.251	0.077	10.743	0.001	0.778
x_5 incident location	0.088	0.012	50.980	<0.001	1.092
x_6 bottleneck state	0.432	0.104	17.291	<0.001	1.541
x_7 trailer state	-1.182	0.060	388.188	<0.001	0.307

In Table 2, SE denotes the standard error. Wald is used to describe the significant of the difference between β and 0. exp(β) represents the hazard ratio. Based on the findings above, the hazard function, that is, the COX regression based co-evolution model, of traffic incident duration in expressways is shown in equation (6) of is as follows:

$$\begin{aligned} h(t,x) = h_0(t)\exp(&-0.221x_1 + 0.278x_2 - 0.527x_3 - 0.251x_4 \\ &+0.088x_5 + 0.432x_6 - 1.182x_7) \end{aligned} \quad \textbf{(6)}$$

Particularly regarding the following findings:

1) For binary variables: daytime & night and trailer state, the regression coefficients of these variables are negative. This indicates that the hazard rate of the "death" of these traffic incidents may reduce with the increase of these covariate variables, that is, the traffic incident duration may be longer. And hence if a traffic incident occurs in night and the trailer is called out, this incident may last more time. Based on the exp(β) in Table 2, it shows that the probability of the "death" of a traffic incident at the night, with trailer is 0.802 and 0.307 times as smaller as that at the daytime and without trailer.

2) For the binary variable: bottleneck state, the regression coefficient of bottleneck is positive. This indicates that if an incident occurs at bottleneck segments, the hazard rate of the death of this incident may be higher, that is, traffic incidents at bottleneck segments may be disposed more quickly compared with those at non-bottleneck segments. In addition, it can be found that the probability of the death of a traffic incident at the bottleneck segment is 1.52 times as bigger as that at non bottleneck segments. One of the reasons may be that the traffic flow at the bottleneck segments is always more complex, and once an incident occurs at these segments, the car owners and emergency management department are requested to clear this incident as quickly as possible so as not to deteriorate the incident.

3) For continuous variables: involved vehicle number and involved lane number. The regression coefficients of these variables are negative. This indicates that with one unit increase of the involved vehicle number and lane number, the logarithm of the RR decreases by 0.615 and 0.215 respectively. The probability of the death of a traffic incident with one unit increase of the involved vehicle number and lane number is respectively 0.541 and 0.807 times as smaller as those with no increase, that is, the traffic incident duration under this condition will be prolonged. As a result, the effect of these traffic incidents will be more serious.

4) For the categorical variables: incident type and incident location. Different from the analysis before, these categorical variables can be modeled by the Categorical Variable Module in COX regression. Taking the single-car breakdown as the basic reference for incident type, as well as the East IRV for incident location, the COX regression based coefficient estimation of incident type and incident location is shown in Table 3.

Table. 3. The estimated parameters for the COX Regression of incident type and location

Incident type	β	SE	Wald	P	exp(β)	Location	β	SE	Wald	P	exp(β)
single-car breakdown			300.52	<0.001		East IRV			94.888	<0.001	
single-car collision	-1.197	0.130	84.865	<0.001	0.302	South IRV	0.533	0.336	2.517	0.113	1.704
two-car collision	-1.304	0.111	137.224	<0.001	0.272	West IRV	0.424	0.124	11.671	0.001	1.527
multi-car collision	-2.007	0.240	70.033	<0.001	0.134	North IRV	0.467	0.122	14.686	<0.001	1.595
goods drop	-0.683	0.107	41.049	<0.001	0.505	Central SNV	0.389	0.124	9.827	0.002	1.475
						North SNV	0.362	0.121	8.926	0.003	1.437
						South SNV	0.743	0.117	39.976	<0.001	2.101

It shows that the coefficient of each incident type is negative, and the incident duration indicated single-car breakdown < goods drop < single-car collision < two-car collision < multi-car collision. Like incident type, the coefficient of each incident location is positive, and the incident duration indicated south SNV < south IRV < north IRV < west IRV < central SNV < north SNV.

6 Conclusion and discussion

To evaluate the significant influencing factors of the duration of the traffic incidents in expressways, in this paper, the survival analysis is introduced to analyze the spatial-temporal distribution characteristics of traffic incident duration in expressways, and a co-evolution model of these influencing factors is established to discuss the key influence set to traffic incident duration.

The statistical results of Shanghai traffic incident data has shown that, the number of traffic incidents occurred in the IRV and SNV is about 48 per day. The probability of the occurrence of traffic incidents is higher in Monday, Tuesday and Friday because of the increasing traffic volumes in these days. Most of these incidents occurred at non-bottleneck segments, and the running speed of traffic is usually slow when an incident occurs.

The vast majority of traffic incidents are slight with two-car collision or one-car breakdown. Correspondingly, most of traffic incidents can be negotiated by drivers. For some serious incidents occurred at the bottleneck segments, trailers may be called occasionally, but ambulance and fire engine are rarely used to rescue.

Based on these findings above, the survival analysis based co-evolution model of influencing factors of traffic incident duration is further developed. The analysis results have demonstrated that there are obvious spatial and temporal characteristics of traffic incident duration in expressways. For the co-evolution of multidimensional influencing factors, the analysis results have shown that only 7 remaining factors have great significance to the duration of traffic incident in expressways. These co-evolution factors are daytime & night, incident type, involved vehicle number, involved lane number, incident location, bottleneck state and trailer state. In intelligent traffic incident management, these findings above have emphasized the need for consideration of the occurrence at night, two-car collision, occupied more lanes, involved more vehicles and trailer call.

The experiment results have demonstrated the rationality and feasibility of the developed model above. Compared with traditional methods, such as linear regression and logistic regression, survival analysis is able to deal with the data in both qualitative and quantitative ways. Survival analysis uses survival function and hazard function to display the temporal distribution of each influencing factors, which helps us to find the law of traffic incident duration. Still, there are some details to be further researched. Most of covariate variables in this study are mainly inherent properties and environment states when traffic incident occurred. The traffic incident duration is still dependent on the subjective influencing factors, such as the delay time before police call, the police call mode, etc. Such factors are closely related to the reaction of incident drivers when an incident occurs. Due to the limitation of traffic incident data, these factors, simplified or discarded, are keys to improve this co-evolution model. In future study, these factors are to be studied and relation of these factors to be elaborated by driver simulator.

References

[1] Office of the Fourth Comprehensive Traffic Survey in Shanghai, "Report on the fourth comprehensive traffic survey in Shanghai," *Shanghai: Shanghai Urban Construction and Communication Commission*, 2010.

[2] National Bureau of Statistics of China, "China Statistical Yearbook 2012," *Beijing: China Statistics Press*, 2012.

[3] U S Department of Transportation, "Traffic incident management handbook," *Washington D.C: Federal Administration*, 2010.

[4] S. Boyles, D. Fajardo, S. Waller, "A Naïve Bayesian classifier for incident duration prediction," *86th Annual Meeting of the Transportation Research Board*, 2007.

[5] K.G. Zografos, N.A. Konstantions, M.V. George, "A real-time decision support system for roadway incident response logistics," *Transportation Research. Part C*, Vol.10, No.1, pp.1, 2002.

[6] A.Garib, A.E. Radwan, H. Al-Deek, "Estimating magnitude and duration of incident delays," *Journal of Transportation Engineering*, Vol.123, No.6, pp.459, 1997.

[7] H.J. Kim, H. Choi, "A comparative analysis of incident service time on urban freeways," *Journal of International Association of Traffic and Safety Sciences*, Vol.25, No.1, pp.62-72, 2001.

[8] S.P. Washington, M.G. Karlaftis, F.L. Mannering, "Statistical and Econometric Methods for Transportation Data Analysis," *CRC Press*, 2009.

[9] Y. Chung, "Development of an accident duration prediction model on the Korean freeway systems," *Accident Analysis and Prevention*, Vol.42, pp.282, 2010.

[10] P. Jovanis, H. Chang, "Disaggregate model of highway accident occurrence using survival theory," *Accident Analysis & Prevention*, Vol.21, No.5, pp.445-458, 1989.

[11] A. Khattak, J. Schofer, M. Wang. "A simple time sequential procedure for predicting freeway incident duration," *Journal of Intelligent Transportation Systems*, Vol.2, No.2, pp.113-138, 1995.

[12] D. Nam, F. Mannering, “An exploratory hazard-based analysis of highway incident duration,” *Transportation Research Part A: Policy and Practice*, Vol.34, No.2, pp.85-102, 2000.

[13] Y.X. Zhou, X.B. Yang, M. Huan, B.F. Si, “Survival analysis approach for estimating urban traffic congestion duration,” *Applied Mathematics and Mechanics*. Vol.34, No.1, pp.98-106, 2013.

[14] G. Tiwari, S. Bangdiwala, A. Saraswat, S. Gaurav, “Survival analysis: Pedestrian risk exposure at signalized intersections,” *Transportation Research Part F: Traffic Psychology and Behavior*, Vol.10, No.2, pp. 77-89, 2007.

[15] J.T. Wong, S.C. Tsai, “A survival model for flight delay propagation,” *Journal of Air Transport Management*, Vol.23, pp.5-11, 2012.

[16] Y.H. Wang, K.C. Mahboub, D.E. Hancher, “Survival analysis of fatigue cracking for flexible pavements based on long-term pavement performance data,” *Journal of transportation engineering*, Vol.131, No.8, pp.608-616, 2005.

[17] J.P. Kharoufeh, K.G. Goulias, “Nonparametric identification of daily activity durations using kernel density estimators,” *Transportation Research Part B: Methodological*, Vol.36, No.1, pp.59-82, 2002.

[18] H.L. Chang, T.H. Yeh, “Regional motorcycle age and emissions inspection performance: A Cox regression analysis,” *Transportation Research Part D: Transport and Environment*, Vol.11, No.5, pp.324-332, 2006.

[19] G.X. Kang, S.E. Fang, “Application of Cox Regression Model in Traffic Incident Duration,” *Journal of Transportation and Safety*, Vol.29, No.2, pp.104-106, 2011.