

Research Article

Traffic Incident Clearance Time and Arrival Time Prediction Based on Hazard Models

Yang beibei Ji,¹ Rui Jiang,² Ming Qu,² and Edward Chung²

¹ School of Management, Shanghai University, Shangda Road 99, Shanghai, China

² Smart Transport Research Centre, Queensland University of Technology, Level 8, P Block, Brisbane, QLD 4001, Australia

Correspondence should be addressed to Yang beibei Ji; jybb@shu.edu.cn

Received 25 February 2014; Accepted 31 March 2014; Published 17 April 2014

Academic Editor: X. Zhang

Copyright © 2014 Yang beibei Ji et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Accurate prediction of incident duration is not only important information of Traffic Incident Management System, but also an effective input for travel time prediction. In this paper, the hazard based prediction models are developed for both incident clearance time and arrival time. The data are obtained from the Queensland Department of Transport and Main Roads' STREAMS Incident Management System (SIMS) for one year ending in November 2010. The best fitting distributions are drawn for both clearance and arrival time for 3 types of incident: crash, stationary vehicle, and hazard. The results show that Gamma, Log-logistic, and Weibull are the best fit for crash, stationary vehicle, and hazard incident, respectively. The obvious impact factors are given for crash clearance time and arrival time. The quantitative influences for crash and hazard incident are presented for both clearance and arrival. The model accuracy is analyzed at the end.

1. Introduction

Traffic incident is considered as one of the major factors which cause traffic congestion and delay. At the same time traffic incident has wide negative impact on both traffic system and related social activities. Many studies have been done to predict, estimate, or try to decrease traffic duration time just because the accurate prediction of incident time can (1) reduce incident duration time, (2) associate Traffic Incident Management (TIM) to quickly respond to incident to mitigate the impact of incidents (Chou [1]), and (3) improve travel time reliability by predicting travel time while occurrence of incident (Tsubota et al. [2]).

Traffic incident is nonrecurrent events which cause a capacity reduction or an abnormal increase in traffic demand, such as crash accident, stalled vehicles, debris, fire, construction, and sporting events. A general incident timeline as shown in Figure 1 reveals that incident duration can be divided into verification, response, clearance, and recovery period by recording timestamps at various stage of an incident.

However, most of the prediction models did not include all four parts or did not give the exact definition of incident

time. Each part of incident time has statistic distribution and has different influence factors. An example is given in Figure 2.

Figure 2 is the occupancy contour plot from 3:00 a.m. to 11:00 a.m., in which the y -axis denotes the distance from start point of the freeway and x -axis indicates the time. The green indicates low occupancy (free flow conditions), the yellow indicates increasing congestion, and the red represents much congestion on the links. The incident is marked as the red area where the incident occurred time (7:00 a.m.), cleared time (7:53 a.m.), and traffic recovered time (8:20 a.m.) are clearly marked. According to the incident data base (SIMS), a multiple vehicle crash incident on 22 July in 2011 was verified on 7:39 a.m. and was cleared on 7:53 a.m., and the tow assistance arrived on 7:47 a.m.. The total duration time was 80 minutes calculated based on real traffic flow data, while the duration time was only 14 minutes according to incident data base. The real duration time was almost 6 times of incident data base, resulting in errors of prediction model.

However, the accuracy of the incident prediction model cannot be improved, which is partly caused by the definition of the incident duration. It is difficult to get real incident duration time as shown in Figure 1, because the incident

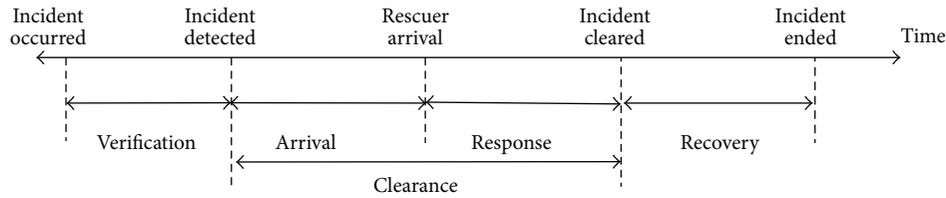


FIGURE 1: Traffic incidents timeline.

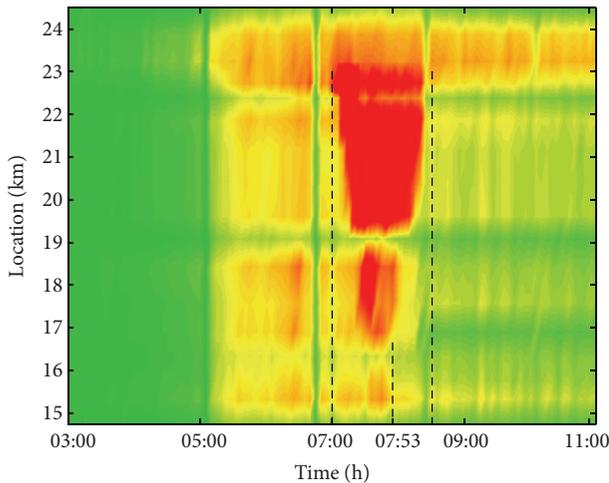


FIGURE 2: Contour plot on the incident scene.

occurred time and the incident ended time cannot be collected exactly. In this paper, the clearance time and arrival time are investigated individually. The major impact factors for the clearance time and arrival time are studied, and two prediction models are then developed. The comparison of different impact factors is put forward.

The rest of this paper is structured as follows. The introduction of Incident Management System (SIMS) in the Pacific Motorway, Brisbane, Australia, and the incident parameter properties are represented in Section 3. Section 4 describes model development: firstly the hazard based model is introduced; secondly, the best statistic fitting distribution for 3 types of incident is given for both clearance time and arrival time; thirdly, the obvious impact factors of the hazard based model are given separately after data filter; finally, the quantitative impact evaluation and precision are obtained. The conclusions are summarized in Section 5.

2. Literature Review

Over the past several decades, various models have been developed to predict traffic incidents. Most of these studies are based on statistics theory. These representative prediction models can be classified into the following categories: probabilistic distribution, linear regression analyses, time sequential methods, decision tree models, neural network

models, Bayesian classifier, hazard analysis model, and so forth.

Golob et al. [3] provided the first probabilistic distribution of incident duration. They found that each part of duration time was related to the previous duration time and demonstrated that the duration fits a lognormal distribution. Giuliano [4] extended Golob's work by applying a lognormal distribution among similar type of incidents and reported that whether involving truck was the key impact factors of incident duration. Garib et al. [5] and Sullivan [6] did similar research. Jones et al. [7] developed double logarithmic distribution of incident duration time. Nam and Mannering [8] found that the Weibull distribution can also be used to describe incident data.

Linear regression models have been widely used due to their simplicity and effectiveness. The relationship between impact factors and incident duration can be obviously reflected in regression models [5, 9]. Ozbay and Kachroo [10] developed incident clearance time based on regression method using 121 traffic incident data collected from Chicago, USA. Nine statistically significant variables were put forward, but the accuracy and effectiveness of the model were not given. Kau [11] estimated freeway incident duration using multiple linear regression method, with confidence interval of 95%. He defined the incident clearance time as the duration from the time that a police vehicle or freeway service patrol vehicle arrives at the scene until the vehicles are cleared from the scene.

The advantage of time sequential model is that it can do regression prediction using a few incident property variables at the early stage of the incident and can update the prediction result when more information is collected. Khattak et al. [12] found that incident type and severity were the most significant factors. He applied time sequential model based on small incident sample, but the practicability had not been demonstrated because of poor accuracy and little incident data.

The decision tree model is independent of distribution of dependent variables. Sethi et al. [13] indicated that the average duration time was 21 minutes based on 801 incident data, including data set of traffic interruption, disable vehicle, and severity. The results showed that the incident type was the most obvious factor on duration. Ji et al. [14] set up the decision tree model based on Bayesian using 1853 incident data in Utrecht, the Netherlands. The precision was improved to 73.39% compared to the decision model. Stephen et al. [15] developed an incident duration model based on a naïve Bayesian classifier. He emphasized that incident duration was

a highly variable quantity and although the model performed better than a linear regression, its classification was still correct only in half of the time.

Artificial neural networks (ANN) have been widely used for prediction and pattern classification problems. Lopes et al. [16] presented an adaptive model to forecast the clearance time of real time traffic incidents. The solutions included four models which were calibrated and tested by incident records from Portuguese highways. The performance showed that it was able to estimate 72% of incident with less than 10 minutes error and about 92% with less than 20 minutes error. Some other examples can be found in [17, 18].

The hazard analysis model has been used in traffic engineering, which is a common topic in many fields such as life sciences, biomedical, and reliability engineering. The model is more effective to analyze time-related problem, which is generally used to describe the analysis of data in the form of time from a well-defined time origin until the occurrence of some particular event of an end point [19]. Examples include the time between incident occurrence and its clearance [8, 20], the time between planning and execution of an activity [21], and the analysis of urban travel time [22].

Furthermore, hazard-based model has been used to model incident duration time. Chung [23] presented an accident duration model using 2-year-accident dataset from 2006 to 2007 in Korean freeway systems, and the Log-logistic distribution was selected for accelerated failure time metric model. Although the model had large prediction error, statistical test results indicated that this model was stable over time. Tavassoli et al. [24] developed parametric accelerated failure time survival models of incident duration. They found that the duration of each type of incident is uniquely different and responds to different factors.

One distinctive feature of hazard based model is that the model precision will be improved if the best fitting distribution of time variable is chosen. In this study, hazard based models, in particular the accelerated failure time (AFT) metric, are utilized to model both incident clearance time and arrival time.

3. Description of Incident Base

Incident data was collected by Queensland Department of Transport and Main Roads' STREAMS Incident Management System (SIMS) for South East Queensland urban networks from November 2009 to November 2010. SIMS is an incident management system which is used throughout Queensland, Australia, to capture traffic incidents which cause an impact on traffic flow on the road network. There are total 35103 incident data for one year, which can be classified into 9 types: alert, congestion, crash, fault, flood, hazard, planned incident, road works, and stationary vehicles. There are many detailed properties in SIMS incident data base, but not all of them are closely related to incident time prediction, such as location, SIMS ID, and status. Hence, the major properties of each incident data are shown in Table 1. However, not all these properties are recorded for each incident occurrence. For example, the parameters are only applicable for crash data

which are "number of vehicles involved," "number of people injured," and "number of fatalities".

Only 3 types of incidents: crash, stationary vehicle, and hazard are used to model development though 9 types of incident recorded in SIMS data base. Other incident type data are rare recorded. Consequently, the clearance time and arrival time prediction model are only developed for 3 types of incident.

4. Model Development

Hazard based time models were originally used for problems in biomedical, engineering, and social sciences, which are a class of statistical methods for studying the occurrence and timing of events. Recently, they were used to model time related issues in transportation. A review of the application of the hazard based duration models in transportation up to the early 1990s [25].

The incident time in hazard based model is a realization of a continuous random variable T , with a cumulative distribution function $F(t)$, which is called the failure function. A probability density function $f(t)$, survival function $S(t)$, and hazard function $h(t)$ are given as (2)–(4). The relationships between these four functions are formulated in (1)–(4), and $P(\cdot)$ means probability. The function of a random variable T is given by

$$F(t) = \int_0^t f(u) du = P(T < t), \quad (1)$$

$$f(t) = \frac{dF(t)}{dt} = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t)}{\Delta t}, \quad (2)$$

$$S(t) = P(T \geq t) = 1 - F(t), \quad (3)$$

$$h(t) = \frac{f(t)}{1 - F(t)} = \frac{f(t)}{S(t)} = \lim_{\Delta t \rightarrow 0} \frac{P(t + \Delta t \geq T \geq t)}{\Delta t}. \quad (4)$$

In (4), with fully parametric models, three distributional alternatives were considered, namely: Gamma, Log-logistic, and Weibull, for the hazard function and are tested to find the best fit to the incident clearance time and arrival time. The functional forms of the hazard function for each model can be derived by using each distribution model and general function.

4.1. Gamma Distribution. The Gamma distribution is briefly described as a two-parameter family of continuous probability distributions. The scale parameter is λ and the shape parameter is \bar{k} , where $\bar{k} > 0$ and $\lambda > 0$. The Gamma function is mathematically defined as [26]

$$\Gamma(\bar{k}) = \int_0^{\infty} t^{\bar{k}-1} e^{-t} dt. \quad (5)$$

After algebra transform, the p.d.f. (probability density function) of the Gamma distribution, generally written as $f[t; T \sim \Gamma(\lambda, \bar{k})]$, is given by

$$f[t; T \sim \Gamma(\lambda, \bar{k})] = \frac{\lambda(\lambda t)^{\bar{k}-1} e^{-\lambda t}}{\Gamma(\bar{k})}, \quad t > 0. \quad (6)$$

TABLE 1: Incident property information.

Parameter	Description information
Priority	
Classification	The classification to further define the type of incidents
Blockage type	The type of blockage caused by the traffic incident. Possible values are unknown, no blockage, partially blocked, blocked, both direction blocked
Number of lanes blocked	The number of lanes blocked by the traffic incident
Fire hazard	If hazard resulted in smoke/poor/visibility and/or fire
Intersection	Whether nearby the intersection
Major incident	Whether the incident is major
Lane blocked	How many lanes are blocked
Heavy traffic	Whether heavy traffic involved
Towing	Whether towing is needed
Number of vehicles involved	The number of vehicles involved in the crash. (only applicable for crash incident data)
Number of people injured	The number of people injured in the crash. (only applicable for crash incident data)
Number of fatalities	The number fatalities from the crash
Medical emergency	Yes/no
Weather	Possible values are cloudy, electrical storm, fine, fog, hail, heavy rain, showers, snow, unknown, wind. This column will always display "Unknown" for weather hazard incident.
Opp lanes blocked	The number of lanes blocked by the traffic incident, on the opposite link.
Diversion required	Indicates whether a diversion was required due to the traffic incident
Lateral position	Possible values are in bay, in lanes, in median, left shoulder, right shoulder, and unknown.
Chemical spill	This is only applicable to crash incidents and debris/obstruction/spill.

When $\tilde{k} = 1$, the Gamma density function reduces to the exponential density function, and the exponential distribution is also a special case of the Gamma distribution.

When $\lambda = 1$, (5) reduces to the one-parameter Gamma distribution, also referred to as the standard Gamma distribution of T , written as

$$f [t; T \sim \Gamma(\tilde{k})] = \frac{t^{\tilde{k}-1} e^{-t}}{\Gamma(\tilde{k})}, \quad t > 0 \quad (7)$$

with its c.d.f. (cumulative distribution function) defined as

$$F [t; T \sim \Gamma(\tilde{k})] = \int_0^t \frac{t^{\tilde{k}-1} e^{-t}}{\Gamma(\tilde{k})}, \quad t > 0. \quad (8)$$

Specification of the survival and hazard functions for the Gamma distribution are based on (8), which is called the incomplete Gamma function. The survival function $S(t)$ is given by the following equation:

$$S [t; T \sim \Gamma(\tilde{k})] = 1 - F [t; T \sim \Gamma(\tilde{k})] = 1 - \int_0^t \frac{t^{\tilde{k}-1} e^{-t}}{\Gamma(\tilde{k})}. \quad (9)$$

The Gamma distribution hazard function can be expressed as

$$h [t; T \sim \Gamma(\tilde{k})] = \frac{f [t; T \sim \Gamma(\tilde{k})]}{S [t; T \sim \Gamma(\tilde{k})]} \quad (10)$$

$$= \frac{\lambda(\lambda t)^{\tilde{k}-1} e^{-\lambda t}}{\Gamma(\tilde{k}) \{1 - F [t; T \sim \Gamma(\tilde{k})]\}}.$$

4.2. Weibull Distribution. The Weibull distribution model is almost the most widely applied parametric function in survival analysis because of its flexibility and simplicity among all the families of parametric time distributions [26].

The Weibull probability of event time T , a continuous function, is featured by the use of two parameters: a scale parameter λ and a shape parameter \tilde{p} . The survival function with the Weibull distribution is given by

$$S (t; \lambda, \tilde{p}) = \exp [-(\lambda t)^{\tilde{p}}], \quad t > 0. \quad (11)$$

Given the intimate associations among various lifetime measures, the hazard function in the Weibull distribution can be readily derived from the above equation.

Consider

$$h(t; \lambda, \tilde{p}) = -\frac{(d/dt) e^{-(\lambda t)^{\tilde{p}}}}{e^{-(\lambda t)^{\tilde{p}}}} = \frac{\lambda \tilde{p} t^{\tilde{p}-1} \exp [-(\lambda t)^{\tilde{p}}]}{\exp [-(\lambda t)^{\tilde{p}}]} \quad (12)$$

$$= \lambda \tilde{p} (\lambda t)^{\tilde{p}-1}.$$

The cumulative hazard function $H(t)$ can be expressed in terms of $S(t)$, given by

$$H(t) = -\log S(t) = -\log \left\{ \exp \left[-\int_0^t h(u) du \right] \right\}. \quad (13)$$

Therefore, the cumulative hazard function $H(t; \lambda, \tilde{p})$ can be written as

$$H(t; \lambda, \tilde{p}) = -\log S(t; \lambda, \tilde{p}) = -\log \left\{ \exp [-(\lambda t)^{\tilde{p}}] \right\} \quad (14)$$

$$= (\lambda t)^{\tilde{p}}.$$

Taking natural Log values on both sides of (14), (14) can be written as

$$\log [-\log S(t; \lambda, \tilde{p})] = \log \lambda + \tilde{p} \log t. \quad (15)$$

Specifications of $S(t)$ and $h(t)$ lead to the following equation for the Weibull p.d.f. function:

$$f(t; \lambda, \tilde{p}) = h(t) S(t) = \lambda \tilde{p} (\lambda t)^{\tilde{p}-1} \exp [-(\lambda t)^{\tilde{p}}]. \quad (16)$$

Likewise, the c.d.f. at time t is derived by

$$F(t; \lambda, \tilde{p}) = 1 - S(t; \lambda, \tilde{p}) = 1 - \exp [-(\lambda t)^{\tilde{p}}]. \quad (17)$$

Given $\lambda^* = 1/\lambda$, the Weibull hazard function can be reexpressed as

$$h(t; \lambda^*, \tilde{p}) = \frac{\tilde{p}}{\lambda^*} \left(\frac{t}{\lambda^*} \right)^{\tilde{p}-1}. \quad (18)$$

4.3. Log-Logistic Distribution. The lognormal distribution is widely used to describe events whose rate increases initially and decreases consistently afterwards. The Log-logistic distribution of T is the antilogarithm of the familiar logistic distribution. Let $Y = \log T$. The density function of Y is defined as the familiar logistic distribution [26]:

$$f(y) = \frac{\hat{b}^{-1} \exp [(y - \hat{\mu})/\hat{b}]}{\{1 + \exp [(y - \hat{\mu})/\hat{b}]\}^2}, \quad y \in (-\infty, \infty), \quad (19)$$

where $\hat{\mu}$ and \hat{b} are parameters for the logistic function of Y , described as $Y \sim \text{Logist}(\hat{\mu}, \hat{b})$. Let $\lambda = \exp(\hat{\mu})$ and $\hat{p} = \hat{b}^{-1}$. The antilogarithm of (19) is the density function of T :

$$f(t) = \frac{(\hat{p}/\lambda) (t/\lambda)^{\hat{p}-1}}{[1 + (t/\lambda)^{\hat{p}}]^2}, \quad t > 0, \quad (20)$$

where λ and \hat{p} are parameters of the Log-logistic distribution, written as $T \sim \text{LLogist}(\lambda, \hat{p})$. The c.d.f. of T is then given as

$$F[t; T \sim \text{LLogist}(\lambda, \hat{p})] = \frac{1}{1 + (t/\lambda)^{-\hat{p}}}. \quad (21)$$

Therefore, the survival and hazard rate functions of T can then be readily derived as follows:

$$S[t; T \sim \text{LLogist}(\lambda, \hat{p})] = 1 - F[t; T \sim \text{LLogist}(\lambda, \hat{p})]$$

$$= \frac{1}{1 + (t/\lambda)^{\hat{p}}}$$

$$h[t; T \sim \text{LLogist}(\lambda, \hat{p})] = \frac{f[t; T \sim \text{LLogist}(\lambda, \hat{p})]}{S[t; T \sim \text{LLogist}(\lambda, \hat{p})]}$$

$$= \frac{(\hat{p}/\lambda) (t/\lambda)^{\hat{p}-1}}{1 + (t/\lambda)^{\hat{p}}}. \quad (22)$$

5. Model Result

5.1. The Fitness of Distribution. Understanding of incident characteristics and patterns is essential to establish an appropriate prediction model; therefore, the statistical analysis is carried out firstly. There are 4966 crash records, 15791 stationary vehicle data, and 3847 hazard records for clearance time which are used to do distribution fitting analysis. Four probability density functions, which are Gamma, Log-logistic, Weibull, and lognormal, are fitted to the clearance time for crash, stationary vehicle, and hazard incidents, respectively, (see Figures 3(a), 3(b), and 3(c)). Thick full lines indicate the best fitness distribution. The figures indicate that each incident classification has its respective best fitness distribution function. Four parameters estimates of clearance time probability density distribution: Log likelihood, domain mean, and variance are listed in Table 2. Log likelihood and variance statistics indicate the goodness of fit distribution.

The less the variance is, the better the distribution fitting will be. For example the Gamma distribution variance for crash clearance time is 967.94, which is the least one comparing other distributions. It is clearly shown that Gamma distribution is best fit for crash clearance time, Log-logistic for stationary vehicle, and Weibull for the hazard.

There are totally 4569 crashes, 14665 stationary vehicle data, and 3382 hazard records for arrival time which are used in this distribution fitting. The number of arrival time record is less than the counterpart for clearance time, because there exists an abundant of invalid arrival time data records in SIMS. All the invalid and defective data are filtered. Figures 4(a), 4(b), and 4(c) represent the probability density distributions of arrival time for crash, stationary vehicle, and hazard separately. Table 3 lists the parameters estimates of arrival time probability density distribution for each incident type. Both the estimate parameters and the figure indicate that the Gamma distribution is best fit for crash arrival time, Log-logistic for stationary vehicle, and Weibull for the hazard.

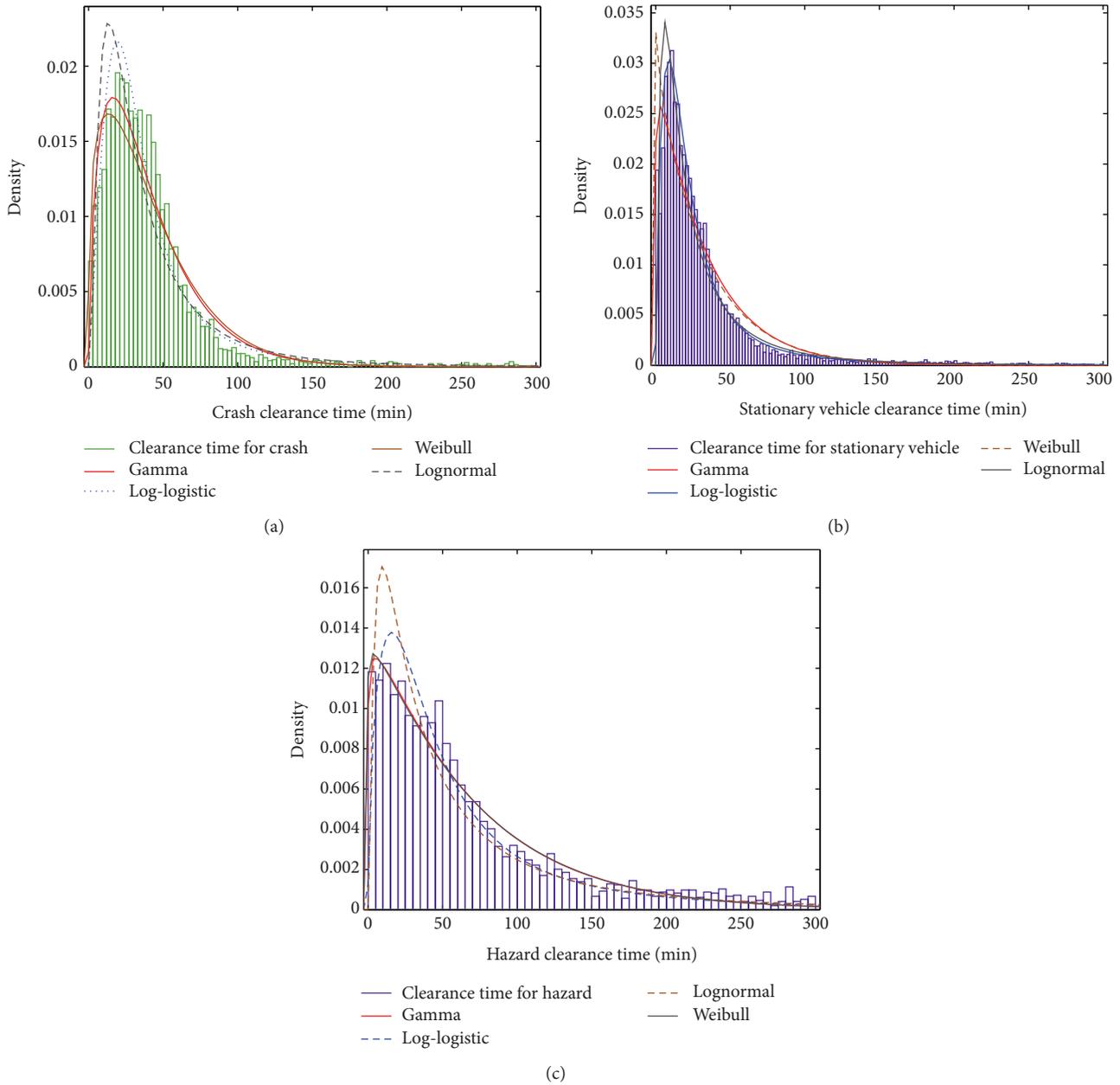


FIGURE 3: Probability density distribution of clearance time for crash (a), stationary vehicle (b), and hazard (c).

In summary, the clearance time and arrival time for the same incident classification follow the same probability distribution, but with different estimated parameters. Different incident classification has different probability distribution characteristics for both clearance time and arrival time. The best fitting model was selected for each incident type based on the results above to develop hazard based prediction model.

5.2. Hazard Based Model for Crash Clearance Time and Arrival Time. The crash clearance time and arrival time are developed based on the Gamma distribution survival model. Data filter is carried out before model development in order

to improve model precision. For example, the clearance times longer than 3 hours, accounting for less than 5% of the total crash dataset, are filtered as shown in the green bars in Figure 5 in order to improve the model accuracy.

Tables 4 and 5 list the parameter estimate results for model estimated for crash incident clearance time and arrival time. A positive sign of an estimate coefficient suggests an increase in the incident clearance time and a decrease in hazard function associated with an increase in that property variable.

All variables are statistically significant at a 95% confidence level. Therefore, all 14 significant property variables for crash clearance time are listed in Table 4. However, there are

TABLE 2: Parameters estimates of clearance time probability density distribution fitness for each incident type.

Incident type	Parameters	Gamma	Log-logistic	Weibull	Lognormal
Clearance time for crash	Log likelihood	-23032.7	-23007.3	-23128.9	-23202.9
	Domain	$0 < y < \text{Inf}$			
	Mean	40.3562	45.586	40.5906	42.8821
	Variance	967.94	11380.3	1020.94	2143.83
Clearance time for stationary vehicle	Log likelihood	-71707.57	-70991.8	-71724.7	-71306.9
	Domain	$0 < y < \text{Inf}$			
	Mean	33.2276	39.749	33.189	35.3951
	Variance	1032.58	975.23	1142.8	2897.44
Clearance time for hazard	Log likelihood	-20244.6	-20437.6	-20245.9	-20527.5
	Domain	$0 < y < \text{Inf}$			
	Mean	68.5289	103.329	68.5441	83.078
	Variance	4382.43	4478.4	4357.8	22634.2

TABLE 3: Parameters estimates of arrival time probability density distribution fitness for each incident type.

Incident type	Parameters	Gamma	Log-logistic	Weibull	Lognormal
Arrival time for crash	Log likelihood	-18182.7	-18211	-18215.4	-18207.7
	Domain	$0 < y < \text{Inf}$			
	Mean	19.9232	25.3389	19.979	21.4866
	Variance	323.564	Inf	343.475	885.809
Arrival time for stationary vehicle	Log likelihood	-58947.1	-58415.1	-58769.5	-58097.4
	Domain	$0 < y < \text{Inf}$	$0 < y < \text{Inf}$	$0 < y < \text{Inf}$	$1 < y < \text{Inf}$
	Mean	21.8141	32.3138	21.5822	23.2076
	Variance	559.408	Inf	628.02	1895.03
Arrival time for hazard	Log likelihood	-17132.8	-17269.8	-17127.8	-17250
	Domain	$0 < y < \text{Inf}$			
	Mean	58.4637	102.204	58.4066	70.5117
	Variance	3728.08	Inf	3914.51	20842.2

TABLE 4: Estimation results of survival AFT model for clearance time for crash.

Variable	Estimate coefficient	Wald Chi-square	Pr > ChiSq
Intercept	4.7907	8.74	0.0031
Priority	-0.1495	27.7239	<0.0001
Intersection	0.1737	5.3656	0.0205
Blockage type	0.1878	13.7469	0.0081
Weather	0.1103	15.1529	0.0341
Total opp lanes block	0.0753	11.5657	0.0412
Traffic disrupted	0.0628	4.6412	0.0312
Heavy traffic	0.1356	10.4797	0.0012
Major incident	0.2158	17.7047	<0.0001
Diversion required	0.2775	33.0936	<0.0001
Towing required	-0.1518	49.2159	<0.0001
Number of vehicles involved	0.2862	70.2871	<0.0001
Number of people injured	0.7411	26.1993	0.0101
Number of fatalities	1.0412	30.9592	<0.0001
Chemical spill	0.3045	17.4566	<0.0001
Log likelihood	-5509.6196		

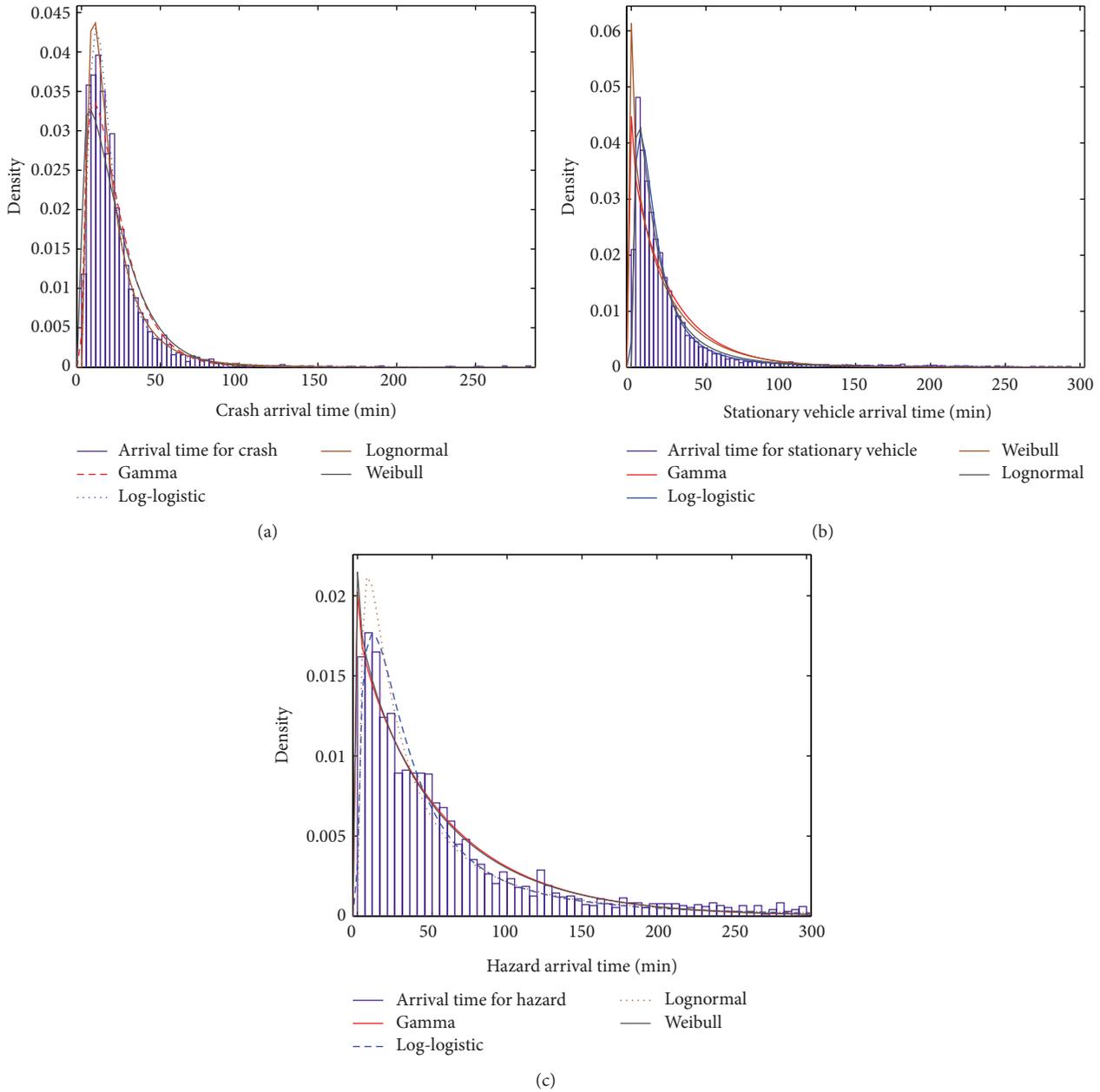


FIGURE 4: Probability density distribution of arrival time for crash (a), stationary vehicle (b), and hazard (c).

only 8 significant variables for arrival time, which is obviously less than that for clearance time. For example, priority, blockage type, weather, and so forth have a significant effect on the clearance time but not on the arrival time, because of different incident time characteristics. Another reason is that the information recorded in the SIMS for arrival time is less than clearance time.

5.3. Influence of Property Parameters on the Prediction Time. Table 6 represents the percentage change in clearance time and arrival time for crash and hazard incident. A negative percentage indicates a decrease in the clearance time with an increase in that property variable. Line “—” means that

the variable has no significant effect on the incident time. For instance, when priority is increased by one, the crash clearance time is 16.13% shorter, and the hazard clearance time is 7.03% shorter, but no significant influence on arrival time. As the number of injured people is increased by one, the crash clearance time is 109.83% longer, while crash arrival time is 57.02% longer, but no significant influence on hazard clearance time and arrival time.

The evaluation of the prediction accuracy for crash incident clearance time is given in Table 7 as an example. There are 4966 crash clearance time data which are used to set up hazard based model. 30% of them are used to evaluate the prediction accuracy. Results in Table 7 indicate the absolute

TABLE 5: Estimation results of survival AFT model for arrival time for crash.

Variable	Estimate coefficient	Wald Chi-square	Pr > ChiSq
Intercept	-0.5248	0.06	0.81
Intersection	0.246	3.6168	0.0472
Lateral position	0.2479	14.2799	0.0039
Total lanes blocked	0.3002	10.6649	0.0493
Traffic disrupted	0.0732	2.8418	0.0218
Towing required	0.2474	48.519	<0.0001
Number of vehicles involved	0.617	57.2253	<0.0001
Number of people injured	0.4512	18.3643	0.0351
Number of fatalities	0.5321	13.7241	0.0082
Log likelihood	-6416.712		

TABLE 6: Percentage change in clearance time and arrival time for crash and hazard.

Variable	Clearance time for crash	Clearance time for hazard	Arrival time for crash	Arrival time for hazard
Priority	-16.13%	-7.03%	—	—
Intersection	18.97%	73.13%	27.89%	71.77%
Lateral position	—	25.54%	28.13%	—
Blockage type	20.65%	47.38%	—	106.49%
Weather	11.66%	11.17%	—	6.56%
Total lanes blocked	—	—	35.01%	—
Total opp lanes blocked	7.82%	—	—	33.99%
Traffic disrupted	6.48%	—	7.59%	—
Heavy traffic	14.52%	—	—	—
Major incident	24.08%	81.56%	—	—
Diversion required	31.98%	—	—	130.23%
Towing required	-16.40%	—	28.07%	—
Number of vehicles involved	33.14%	—	85.34%	377.41%
Number of people injured	109.83%	—	57.02%	—
Number of fatalities	183.26%	—	70.25%	—
Chemical spill	35.59%	—	—	—

TABLE 7: Summary of evaluation of the prediction accuracy.

Performance measure	Value	Percentage
<10 min	543 incidents	39.7%
<15 min	768 incidents	56.1%
<20 min	963 incidents	70.4%
<30 min	1158 incidents	84.6%
<50 min	1280 incidents	93.5%

value of the difference between prediction clearance time and measured clearance time. For example there are 543 incidents whose difference between prediction and measured time is less than 10 minutes, which account for 39.68% of the total evaluation incident data. The accuracy of the model is similar with that of Chung [23].

6. Conclusions

Hazard based prediction model for both incident clearance time and arrival time are developed. Three types of incidents are investigated based on data collected from SIMS. Before

model development, the best fitting model was selected for each incident type based on the results of the likelihood ratio and variance. The results show the following.

- (1) Clearance time and arrival time follow the same distribution with the different estimated parameters for each incident type.
- (2) Gamma, Log-logistic, and Weibull distribution are best fit for crash, stationary vehicle, and hazard incident, respectively. After data filter, the hazard based prediction model is developed for crash incident as example.

There are 14 significant incident property variables for clearance time, while there are only 8 significant variables for arrival time. It clearly indicates that clearance time and arrival time have different impact factors.

The percentage changes in clearance time and arrival time for crash and hazard incident are given. The impact of each property variable on clearance time and arrival time is quantitatively provided. The model accuracy is given at the end of paper.

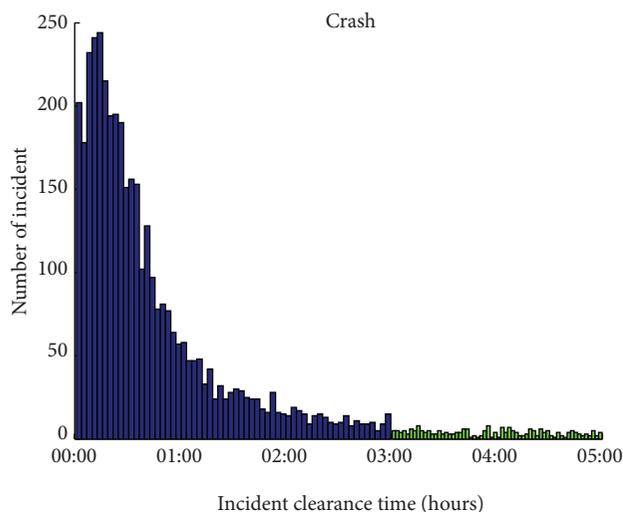


FIGURE 5: Distribution of crash clearance time.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This study has been substantially supported by the National Natural Science Foundation Council of China (no. 71201103). The first author finished the work during the postdoctoral research in Tongji University, and the support of a Project of Shanghai Shuguang Program (13SG23) is acknowledged. The first author thanks the support of the Smart Transport Research Center at the Queensland University of Technology by providing real traffic incident data.

References

- [1] C.-S. Chou, *Understanding the impact of incidents and incident management programs on freeway mobility and safety [Dissertation]*, University of Maryland, College Park, Md, USA, 2010.
- [2] T. Tsubota, H. Kikuchi, K. Uchiumi, H. Warita, and F. Kurauchi, "Benefit of accident reduction considering the improvement of travel time reliability," *International Journal of Intelligent Transportation Systems Research*, vol. 9, no. 2, pp. 64–70, 2011.
- [3] T. F. Golob, W. W. Recker, and J. D. Leonard, "An analysis of the severity and incident duration of truck-involved freeway accidents," *Accident Analysis & Prevention*, vol. 19, no. 5, pp. 375–395, 1987.
- [4] G. Giuliano, "Incident characteristics, frequency, and duration on a high volume urban freeway," *Transportation Research A: General*, vol. 23, no. 5, pp. 387–396, 1989.
- [5] A. Garib, A. E. Radwan, and H. A. Deek, "Estimating magnitude and duration of incident delays," *Journal of Transportation Engineering*, vol. 123, no. 6, pp. 459–466, 1997.
- [6] E. C. Sullivan, "New model for predicting freeway incidents and incident delays," *Journal of Transportation Engineering*, vol. 123, no. 4, pp. 267–275, 1997.
- [7] B. Jones, L. Janssen, and F. Mannering, "Analysis of the frequency and duration of freeway accidents in Seattle," *Accident Analysis & Prevention*, vol. 23, no. 4, pp. 239–255, 1991.
- [8] D. Nam and F. Mannering, "An exploratory hazard-based analysis of highway incident duration," *Transportation Research A: Policy and Practice*, vol. 34, no. 2, pp. 85–102, 2000.
- [9] S. Cohen and C. Nouveliere, "Modelling incident duration on an Urban expressway," in *Proceedings of the 8th International Federation of Automatic Control Symposium on Transportation Systems (IFAC '97)*, Chania, Greece, 1997.
- [10] K. Ozbay and P. Kachroo, *Incident Management in Intelligent Transportation Systems*, Artech House, Boston, Mass, USA, 1999.
- [11] V. H. L. Kau, *Estimating Freeway Incident Clearance Duration Using Multiple Linear Regression*, The University of Texas at Arlington, Arlington, Tex, USA, 2007.
- [12] A. J. Khattak, J. L. Schofer, and M. H. Wang, "A simple time sequential procedure for predicting freeway incident duration," *Journal of Intelligent Transportation Systems*, vol. 2, no. 2, pp. 113–138, 1995.
- [13] V. Sethi, F. S. Koppelman, C. P. Flannery, N. Bhandari, and J. L. Schofer, "Duration and travel time impacts of incidents-advance project," Tech. Rep. TRF-ID-202, Northwestern University, Evanston, Ill, USA, 1994.
- [14] Y. B. B. Ji, X. N. Zhang, and L. J. Sun, "Traffic incident duration prediction grounded on Bayesian decision method-based tree algorithm," *Journal of Tongji University: Natural Science*, vol. 36, no. 3, pp. 319–324, 2008.
- [15] B. Stephen, D. Fajardo, and S. T. Waller, "A naïve bayesian classifier for incident duration prediction," in *Proceedings of the 86th TRB Annual Meeting Compendium of Papers*, Washington, DC, USA, 2007.
- [16] J. Lopes, J. Bento, F. C. Pereira, and M. B. Akiva, "Dynamic forecast of incident clearance time using adaptive artificial neural network models," in *Proceedings of the 92nd TRB Annual Meeting Compendium of Papers*, Washington, DC, USA, 2013.
- [17] C.-H. Wei and Y. Lee, "Sequential forecast of incident duration using Artificial Neural Network models," *Accident Analysis & Prevention*, vol. 39, no. 5, pp. 944–954, 2007.
- [18] V. Gaetano, M. Lelli, and D. Cucina, "A comparative study of models for the incident duration prediction," *European Transport Research Review*, vol. 2, no. 2, pp. 103–111, 2010.
- [19] D. Collett, *Modelling Survival Data in Medical Research*, Chapman & Hall/CRC, Boca Raton, Fla, USA, 2nd edition, 2003.
- [20] A. Stathopoulos and M. G. Karlaftis, "Modeling duration of urban traffic congestion," *Journal of Transportation Engineering*, vol. 128, no. 6, pp. 587–590, 2002.
- [21] C. R. Bhat and A. R. Pinjar, "Duration modeling," in *Handbook of Transport Modeling*, D. A. Hensher and K. J. Button, Eds., Elsevier, Amsterdam, The Netherlands, 2nd edition, 2008.
- [22] P. C. Anastasopoulos, J. E. Haddock, M. G. Karlaftis, and F. L. Mannering, "An analysis of urban travel times: a random parameters hazard-based approach," in *Proceedings of the 91st TRB Annual Meeting Compendium of Papers*, Washington, DC, USA, 2012.
- [23] Y. Chung, "Development of an accident duration prediction model on the Korean Freeway Systems," *Accident Analysis & Prevention*, vol. 42, no. 1, pp. 282–289, 2010.
- [24] H. A. Tavassoli, L. Ferreira, S. Washington, and P. Charles, "Hazard based models for freeway traffic incident duration," *Accident Analysis & Prevention*, vol. 52, pp. 171–181, 2013.

- [25] D. A. Hensher and F. L. Mannering, "Hazard-based duration models and their application to transport analysis," *Transport Reviews*, vol. 14, no. 1, pp. 63–82, 1994.
- [26] L. Xian, *Survival Analysis: Models and Applications*, Higher Education Press, Beijing, China, 2012.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

