A Crash Risk Estimation Model for Urban Expressway Basic Segments Considering Geometry, Traffic Flow and Ambient Conditions

Yong WU [a], Hideki NAKAMURA [b], Miho ASANO [c]

[a,b,c] *Department of Civil Engineering, Nagoya University, Nagoya, 464-8603, Japan*
[a] *E-mail: yong@genv.nagoya-u.ac.jp*
[b] *E-mail: nakamura@genv.nagoya-u.ac.jp*
[c] *E-mail: asano@genv.nagoya-u.ac.jp*

Abstract: This study aims at developing a crash risk estimation model (CREM) considering the interaction of geometry, traffic flow and ambient conditions for basic segments of Nagoya Urban Expressway. A matched case-control study is firstly designed separately for three traffic conditions; low-density, high-density uncongested and congested regimes. Based on these case-control samples, conditional logistic regression is then applied and the CREM is finally developed. The results reveal that the model has statistical significance and acceptable goodness-of-fit, with 86.6%, 80.5% and 70.2% of predictive performance for each of the three traffic conditions, respectively. Regarding crash influencing factors, with the increase of traffic density, the significance of horizontal geometry affecting crash becomes lower. In contrast, the contribution of vertical geometry to crash risk is on the rise. Meanwhile, the effect of average speed on crashes gets more significant. Besides, nighttime or holiday can increase the relative risk compared to daytime or weekday, respectively.

Keywords: Crash Risk Estimation Model (CREM), Matched Case-control Study, Urban Expressway, Odds of Crash Occurrence, Traffic Conditions

1. INTRODUCTION

Growing concern over traffic safety has led to research efforts directed towards predicting crash occurrence in advance as dynamic traffic management (DTM) appears. Several studies have suggested that crashes are associated with the interaction of geometry, traffic flow and ambient conditions (Rengarasu *et al.*, 2009; Bajwa *et al.*, 2010). However, few studies have incorporated these factors in a single model to investigate their combined effects on crashes. Furthermore, as traffic conditions change, traffic characteristics are varied and the influence of traffic flow on crashes may also differ, while existing studies paid little attention to this regard. In addition, although crash characteristics are dependent on facility types (Wu *et al.*, 2012), most previous studies focused on the whole routes of intercity expressway.

Comparing to intercity expressway, urban expressway is often forced to have tight geometric design due to urban limited space, such as higher access density and smaller curve radius. Besides, traffic characteristics, *e.g.* the vehicle composition and the driver population, are different between two types of expressways. Necessarily, crash characteristics and their related influencing factors may be different as well.

This paper aims at developing a crash risk estimation model considering the interaction of geometry, traffic flow and ambient conditions on urban expressway. Meanwhile, only basic segments are focused on. Crash rate characteristics and their related influencing factors, as a proactive analysis, have been introduced in our another paper (Wu *et al.*, 2013). Based on this analysis, a matched case-control study is adopted in this study, for comparing the influencing factors between crashes and the corresponding non-crash samples. Then, conditional logistic regression is applied for quantifying the effects of influencing factors on crashes.

The rest of this paper is organized as follows. The following section briefly summarizes the major issues shared by previous studies on crash analysis. Section 3 describes the features of the study sites in detail. Then the methods for data processing, the theories of matched case-control study and conditional logistic regression are explained in Section 4. Section 5 demonstrates the crash risk estimation model (abbreviated to CREM) by traffic conditions. Finally, Section 6 offers conclusions and suggestions for future research.

2. LITERATURE REVIEW

Numerous studies have established statistical links between crash rate/frequency and various influencing factors (Golob *et al.*, 2004; Lord *et al.*, 2005). In those studies, traffic conditions are generally represented by low-resolution traffic data, such as hourly or daily flows. Besides, geometric features are primarily reflected in light of the hierarchy of radius or slope (Shivery *et al.*, 2011). Even so, most existing models were separately developed based on a single factor, which may undermine the validity of models (Abdel-Aty and Pande, 2007).

In view of the insufficiency of aggregated statistics in reflecting the nature of individual crashes, some studies have tried to develop crash models at individual crash level, in an effort to reliably predict crash risk on a real-time basis. Based on an urban freeway in Toronto, Lee *et al.* (2002) confirmed that crash occurrence is significantly affected by short-term turbulence of traffic flow. Several studies followed and extensively developed the method of crash prediction analysis for intercity expressway, while different factors were involved in their models. Oh *et al.* (2005) suggested that the standard deviation of speed is the best indicator of a “disruptive” traffic flow leading to crashes. Comparatively, Abdel-Aty and Pemmanaboina (2006) found that the 5-min average occupancy, standard deviation of volume and the coefficient of variation in speed can affect crash occurrence most significantly.

However, the combined effects of geometry, traffic flow and ambient conditions on crashes have not been well investigated in the above studies. Besides, most existing models are not facility type-specific. Regarding the methods of previous studies, statistical methods, such as logistic regression (Bajwa *et al.*, 2010) and probit model (Christoforou *et al.*, 2011), were generally utilized. Artificial intelligence, *e.g.* neural networks (Oh *et al.*, 2005) and classification trees (Pande and Abdel-Aty, 2006), was also applied for predicting crash risk.

Such traffic variables as flow rate q and speed v are highly correlated with each other. Neural network-based methods can accommodate these variables, while they expect sufficient prior knowledge regarding the problem exhibited through the interrelationship of the predictors (Hossain and Muromachi, 2012). As for statistical methods, the significance and independence of explanatory variables should be identified in advance for the reliability of statistics.

Given the problems of existing studies, the objective of this paper is to develop a CREM for urban expressway basic segments, considering the combined effects of geometry, traffic flow and ambient conditions on crashes. These explanatory factors are first identified and the CREM is finally developed by focusing on traffic conditions.

3. STUDY SITES AND DATABASES

3.1 Study Sites

As shown in Figure 1, Nagoya Urban Expressway network (NEX) is involved in this study. Up to December 31, 2009, this network was about 69.2km in total length with over 250 ultrasonic detectors installed in approximately 500m intervals on mainline. Most routes are 4-lane roadways (2-lane/dir) except for Inner ring (route No.R) which is one-way roadway and where the number of lane differs (2~5) with the change of ramp-junctions.

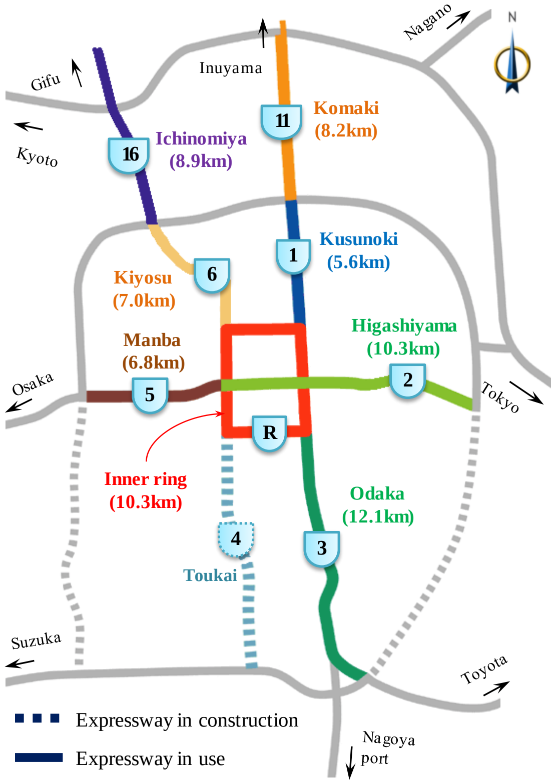

Figure 1. Schematic map of Nagoya Urban Expressway network (2009)
(Source: Nagoya Expressway Public Corporation, modified by authors)

Basic segment is selected for the following analysis. It is defined as the segments that are outside the influence areas of merging, diverging and weaving maneuvers. In this study, basic segment is extracted outside the 500m up- and downstream of ramp-junctions. Besides, a special geometric design, tight curve with a radius smaller than 100m, is excluded in advance due to its high crash rates than ordinary segments (Wu *et al.*, 2013). In view of the limitation of segment samples, basic segments on Inner ring are not considered. As a result, a total length of 56.63km 4-lane basic segments is used for the following analysis.

3.2 Databases

Five databases are prepared in this study; 1) crash records with the occurrence time in minute and the location in 0.1km, 2) detector data including traffic volume q, average speed v and occupancy *occ* per 5 minutes on cross section basis, 3) geometric design and the locations of detectors in 0.01km, 4) the locations and periods of temporal lane/cross-section closures, and 5) daily sunrise/sunset time records in Nagoya. The period of the data above is for three years (2007-2009) except for those on Kiyosu (route No.6) that opened from December 1, 2007.

4. METHODOLOGY

4.1 Data Processing and Matching

4.1.1 Detector data

In principle, detectors can count the number of vehicles at their locations only. Hence, the "coverage area" of detector should be defined, for estimating traffic conditions at crash locations by detector data. The boundary of consecutive coverage areas is defined at the midpoint between two neighboring detectors. Note that, the time of crash in the study dataset is not the exact occurrence time, since it was recorded by road administrators after the crash occurrence. For this reason, detector data within small time before crash should be rejected to avoid mixing up crash-influencing and crash-influenced data. Thus, the latest data at least 5 minutes before the recorded time are accepted in this study. The invalid data and the data within lane/cross-section closure intervals are excluded from the study dataset.

Traffic density has been proposed as the service measure of traffic flow for basic segments in some literatures (HCM, 2010). In this regard, average density k_e estimated by 5-min traffic volume q and average speed v is adopted to be the measure of effectiveness for classifying traffic conditions in this study.

$$k_{ei} = \frac{12 \times q_i}{v_i} \tag{1}$$

Where, q_i, v_i and k_{ei} denote traffic volume, average speed and the estimated traffic density in 5 minutes at detector ID i, respectively.

4.1.2 Geometric features

Design consistency is the conformance of geometry of a highway with driver expectancy, and its importance and significant contribution to safety is justified by understanding the driver-vehicle-roadway interaction (Ng and Sayed, 2004). Considering the interaction that is truly characterized in location-specific, geometric variation in the upstream of crash location is proposed to reflect the effect of geometry on crashes. In view of the length of coverage area, the following variables in 500m distance are used (Hikosaka and Nakamura, 2001).

1) Variation in road elevation h between the crash location and its 500m upstream, and the maximal elevation H during the 500m upstream distance (Figure 2).

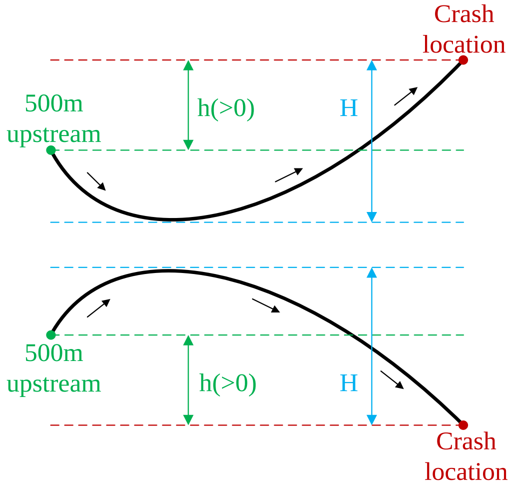

Figure 2. Variation in road elevation

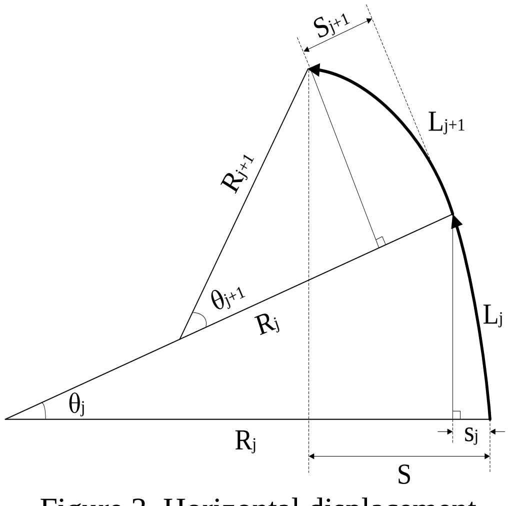

Figure 3. Horizontal displacement

2) Horizontal displacement S. Radius is impossible to describe a section composed of various curves. Besides, centrifugal force is also essentially associated with the horizontal displacement S in the direction of tangent to the curve (Figure 3). Instead, S in the 500m distance is adopted and calculated by the following equations.

$$\theta_j = \frac{L_j}{R_j} \qquad (0 \le \theta_j \le \frac{\pi}{2}) \tag{2}$$

$$s_j = R_j(1 - \cos\theta_j) \tag{3}$$

$$S = \sum s_j \tag{4}$$

Where, j is the ID of curve. R_j, θ_j, L_j and s_j corresponds to the radius, central angle, arc length and horizontal displacement of curve j, respectively.

3) Index of centrifugal force I_{CF}. Speed v always has a square relation with centrifugal force. This study designs I_{CF} to reflect the combined effect of speed v along with horizontal displacement S, while it is not the correct value of centrifugal force.

$$I_{CF} = S \times v^2 \tag{5}$$

4) Index of space displacement I_{SD}. In the following, I_{SD} is utilized to reveal the comprehensive geometric features induced by horizontal and vertical variation.

$$I_{SD} = S \times H \tag{6}$$

The geometric data above are collected every 0.1km as crash is recorded in a unit of 0.1km. Besides, those data are also extracted at the location of detector that is the common link between crash and detector data. Table 1 exemplifies the process of such data collection.

Table 1. Example of geometric variations collection

Route #	Direction	Kilo-post (km)	h (m)	H (m)	S (m)	I_{SD} (m^2)	Note
1	Southbound	0.0	-4.63	5.49	0.78	4.30	
1	Southbound	0.1	-7.90	8.49	3.91	33.2	
1	Southbound	0.2	-10.6	11.5	6.08	69.9	
1	Southbound	0.21	-11.5	11.8	8.88	104.7	Detector #0101
1	Southbound	0.3	-15.3	15.3	9.60	146.9	
1	Southbound	…	…	…	…	…	
1	Southbound	6.4	10.2	10.9	5.15	56.1	
1	Northbound	0.0	0.53	0.98	12.6	12.3	
…	…	…	…	…	…	…	
16	Northbound	8.1	-1.31	1.40	17.0	23.8	

4.1.3 Ambient conditions

Commonly prevailing and uncontrolled environment and weather conditions are defined as ambient conditions in this study. They are 1) ambient light classified into daytime/nighttime that is the time period from sunrise to sunset and from sunset to sunrise, respectively, 2) sunny/cloudy/rainy weather conditions at the time of crash, 3) dry/wet pavement conditions at the location of crash, and 4) day type on crash days including holiday/weekday. Here, holiday includes all weekends, all national and traditional holidays like the Golden Week in May and the Obon Week in August in Japan.

4.1.4 Data matching

The related detector data, geometric variations and ambient conditions for individual crashes are matched as demonstrated in Table 2. The crashes matched with invalid detector data and within traffic regulation period such as lane and cross-section closure intervals are also excluded in advance. Consequently, a total of 457 crashes remain for the following analysis.

Table 2. Example of data matching at individual crash level

Crash ID	Detector data			Geometric features				Ambient conditions			
	q (veh/5min)	v (km/h)	k_e (veh/km)	h (m)	H (m)	I_{CF} (km^3/h^2)	I_{SD} (m^2)	Ambient light	Weather condition	Pavement condition	Day type
1	58	86.4	8	4.50	4.85	57.4	37	Night	Sunny	Dry	Holiday
2	267	38.6	83	-1.55	3.69	4.93	12	Day	Rainy	Dry	Weekday
3	2	50.4	1	1.64	2.14	87.3	74	Night	Cloudy	Dry	Weekday
4	60	77.0	9	4.44	4.44	61.9	46	Night	Sunny	Wet	Holiday

4.2 Matched Case-Control Design

The case-control design is an efficient method to study rare event that is particularly prevalent in epidemiology. In theory, it is an observational-retrospective study: it identifies the cases (a group with outcome) and the controls (a group without outcome), and then traces back to investigate the exposures which are related to outcomes (Lewallen and Courtright, 1998).

For crash analysis, the case is a crash event which may be associated with various exposure factors. The matched controls correspond to the crash scenes or similar conditions but not involved in a crash. If the factors interested for analysis are defined, they would be incorporated in analysis and other conditions should be controlled.

Even controls should like the cases in many ways, it is possible to over-match, where the factors interested for analysis are controlled. Over-matching can result in underestimation on influences. Another important technique for adding power to this method is to enroll more than one control for each case (Lewallen and Courtright, 1998).

As a rule of thumb, a case-to-control ratio around 1:4 is recommended as the statistical power does not increase significantly under a 1:4 ratio (Zheng *et al.*, 2010). Number of factors analyzed in these studies is smaller than 4 while more than 4 variables may be involved in this study. Hence, another way to decide the reliable case-to-control ratio is proposed in this study through examining the *Odds ratios* of individual factors explained in 4.3 whether they significantly change or not as control samples increase.

4.3 Conditional Logistic Regression

In essence, crash is a binary outcome event (crash *vs.* non-crash). If the outcome is binary, the prevalent method to measure the effects of several independent variables on it is logistic regression (Hosmer and Lemeshow, 2004). Meanwhile, conditional logistic regression is a popular method to analyze the relationship between an outcome and a set of explanatory factors in matched case-control studies. Thus, this study adopts conditional logistic regression to predict the probability of crash occurrence. Generally, the probability of crash occurrence (Y=1) considering various impact factors on crash (X=x_1, x_2, ..., x_n) can be expressed as:

$$P = P(Y = 1 \mid x_1, x_2, ..., x_n) \qquad (0 < P < 1) \tag{7}$$

Crash has two distinct outcomes, thus the probability value estimated by Equation (7) is actually a pseudo-value in practice. However, in theory, this value may be used to reveal the relative risk of crash occurrence for a given condition compared to the conditions involved crashes. For this purpose, the *Odds* of crash occurrence are applied. As P is defined as the probability of crash occurrence, 1-P is regarded as the probability of crash not occurring. Then, the *Odds* of a crash in a given condition can be defined as P/(1-P). As described before, several factors should be involved to represent this exposure. The joint effect of all the factors on the *Odds* of crash occurrence put together can be expressed mathematically as:

$$Odds = \frac{P}{1-P} = \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_n x_n) \tag{8}$$

Here, β_0 is the intercept value and β_1, β_2,…, β_n correspond to the coefficients estimated for individual variables.

Towards revealing the effects of individual factors on the *Odds* of crash occurrence, *Odds ratio* (*OR*) is employed. *OR* is defined as the ratio of the *Odds* in favor of a crash from one factor (x_k) to the *Odds* in favor of a crash from another factor (x_l).

$$Odds \;\; ratio = \frac{Odds \; (x_k)}{Odds \; (x_l)} \tag{9}$$

To reduce the number of comparison, *Odds* contributed by x_l is taken as 1.0 in this study. Note that, x_l is actually an assumed factor that is designed to serve for simplifying comparison. In this way, *OR* of x_k is equal to its corresponding exp (β_k). Hence, *OR* can be regarded as the variation of the *Odds* of crash occurrence induced by the increase in a unit of x_k. In substance, it implies the relative contribution of x_k to crash risk in a given condition.

5. CRASH RISK ESTIMATION MODEL

In this section, the required independent variables by traffic conditions are analyzed according to the results of Principal Component Analysis (PCA) in Wu *et al.* (2013). Followed these findings, the matched non-crash samples are designed for individual crashes. Then, based on these case-control samples, a crash risk estimation model (CREM) is finally developed.

5.1 Selection of Independent Variables

As concluded in Wu *et al.* (2013), traffic conditions could be classified into three regimes in terms of the monotonicity of crash rate tendency (Figure 4); 1) low-density uncongested regime where speed is higher than the critical speed v_c of 60km/h (the boundary between un- and congested flows) and traffic density is lower than 25veh/km, 2) high-density uncongested regime in which speed is still higher than 60km/h while traffic density is larger than 25veh/km, and 3) congested regime where speed is lower than 60km/h.

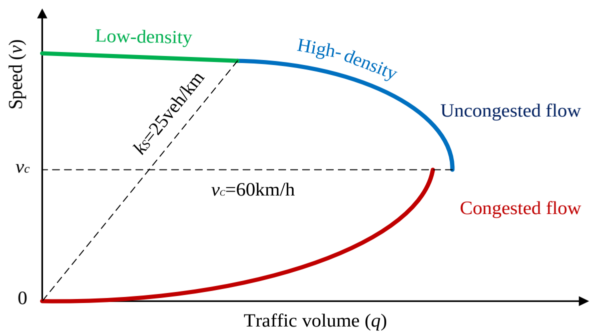

Figure 4. Classification of traffic conditions

The following variables were identified through using PCA in Wu *et al.*, (2013). Both speed *v* and traffic density k_e were selected to represent traffic conditions. Index of centrifugal force I_{CF} and variation in elevation *h* were used to describe horizontal and vertical alignment variation, respectively. Besides, index of space displacement I_{SD} was regarded as an index of comprehensive geometric feature. Dummy variables were referred in order to incorporate ambient conditions into PCA: *Light*=1 if nighttime, 0 otherwise and *Day*=1 if holiday, 0 otherwise. Meanwhile, weather conditions were replaced by pavement conditions (*Pave*= 1 if wet pavement, 0 otherwise). The analysis results are briefly summarized in Table 3.

Table 3. Summary of the significance of individual factors*

Variable	Uncongested regime		Congested regime	Definition
	Low-density	High-density		
ke	+	+	+	Traffic density
v	+	+	+	Average speed
I_{CF}	+	+	+	Index of centrifugal force: $I_{CF}=S\times v^2$
h	+	+	+	Variation in road elevation
I_{SD}	+	+	+	Index of space displacement: $I_{SD}=S\times H$
Light	+	+	-	Ambient light conditions
Pave	+	-	+	Pavement conditions
Day	-	+	+	Day type

* +/-: denotes significant/not significant

For the reliability of statistics, these explanatory factors should be selected in terms of significance and independence. For traffic variables, speed *v* and traffic density k_e are used together in low-density uncongested regime since both factors belong to different components (Wu *et al.*, 2013). In other regimes, a single variable is applied since it can generally reflect average traffic conditions. Considering the application in dynamic traffic management, speed *v* is selected. In view of the independence of various geometric factors, the comprehensive index I_{SD} is rejected. Horizontal displacement *S* is utilized to replace index of centrifugal force I_{CF}, while variation in road elevation *h* is still kept. Ambient conditions are chosen in terms of their significance. However, pavement conditions are not adopted in the following analysis, since these data are not virtually available for non-crash days.

5.2 Control (Non-Crash) Samples Design

The variables above should be matched for crash records and their related non-crash samples. Traffic variables are extracted on the day of crash and from all corresponding non-crash days. The correspondence here means that non-crash days around the day of crash on the same day-of-week, to control the monthly/weekly variation in traffic characteristics. Meanwhile, regarding the daily variation in traffic conditions, these data (non-crash samples) are collected around the time of crash, and half hour prior to the time of crash is accepted in this study.

Towards reflecting the variation in geometric features, non-crash samples are also designed at other locations of basic segments. In the purpose to reveal the effects of alteration in ambient conditions on crash risk, some non-crash samples should be designed in the other ambient conditions relative to the conditions of a crash.

As illustrated in Figure 5, a crash occurred on Wednesday, May 6, 2009, 15:12 at 8.6km on Northbound of Odaka line. Referring to the methods explained in Section 4.1, data from detector #0335 (nearest to 8.6km) at 15:05 (at least 5 minutes before 15:12) can be used to represent traffic condition prior to this crash. Actually, it belongs to low-density uncongested regime (see Table 4). The geometric variation at crash location is extracted from geometric database. The ambient conditions at 15:12 can be referred to crash records. As for its control samples, the corresponding days are regarded as other Wednesdays (and non-crash days) before/after the day of crash in one month. Then, traffic data are randomly collected from the detectors on Odaka line. Meanwhile, some data are extracted in nighttime. Of course, these control samples are extracted at basic segment only and the data not belonging to low-density uncongested regime are excluded. The design process above is summarized in Table 4.

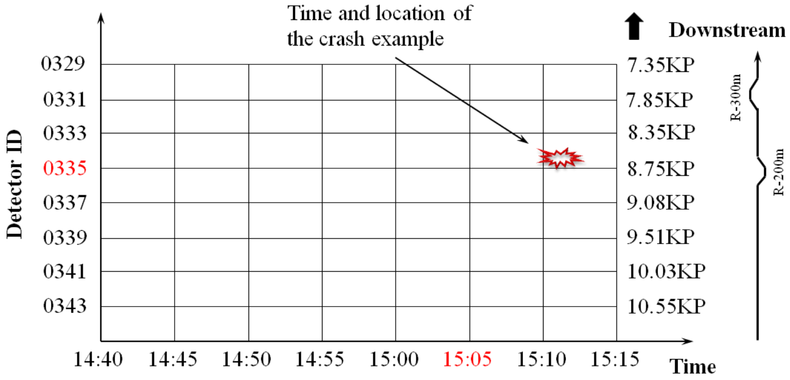

Figure 5. Time-space allocation of the crash example

Table 4. Example of case-control design (low-density uncongested regime)

Y*	Day	Time	Detector ID	v (km/h)	k_e (veh/km)	S (m)	h (m)	*Light*
1	05/06/2009	15:05	0335	72.6	20	26.99	0.08	0
0	04/22/2009	3:00	0303	79.8	1	6.00	0.65	1
0	04/22/2009	22:25	0315	81.5	12	0.00	4.23	1
…	…	…	…	…	…	…	…	…
0	04/29/2009	11:00	0339	85.5	16	6.28	0.99	0
0	04/29/2009	23:00	0341	89.1	5	13.35	-8.16	1
…	…	…	…	…	…	…	…	…
0	05/20/2009	7:00	0313	88.9	18	4.94	-4.61	0

* Y=1: crash event; Y=0: the matched non-crash samples.

5.3 Model Development

5.3.1 Minimal required control sample size

To simplify the process of case-control samples, this study designs a method to examine the minimal required control samples. If *OR* of individual variables does not significantly change after inputting more than m control samples, m is regarded as the minimal required control sample size. Correspondingly, the maximal required case-to-control ratio can be taken as 1: m.

Table 5. *OR* of individual variables based on varied case-to-control ratios

Case-to-control ratio (1:m)	v	k_e	S	h	*Light*
1:1	1.033	0.999	1.026	1.065	1.473
1:3	1.025	0.998	1.020	1.055	1.375
1:5	1.019	0.997	1.020	1.035	1.327
1:7	1.014	0.995	1.018	1.031	1.309
1:9	1.010	0.991	1.019	1.029	1.287
1:10	1.009	0.990	1.018	1.028	1.285
1:15	1.009	0.990	1.018	1.027	1.283
1:20	1.008	0.989	1.018	1.028	1.282
…	…	…	…	…	…
1:50	1.008	0.989	1.018	1.028	1.282

Table 5 gives the *OR* of individual variables with the increase in control samples in low-density uncongested regime. After inputting more than 10 control samples, even if these values vary among different factors, they do not significantly change for the same variable. For the sake of security, the case-to-control ratio is finally defined as 1:20. Based on this ratio, a total of 193, 87 and 79 crashes combining with their corresponding controls are designed successfully in low-density, high-density uncongested and congested regimes, respectively.

5.3.2 Crash risk estimation model (CREM)

Table 6 summaries the results of CREM by traffic condition and they are significant at 95% level. Meanwhile, all of the independent variables are of statistical significance (*Sig.*<0.05), while only the estimated coefficients *Coef.* and their significances are illustrated in this table.

Table 6. Summary of crash risk estimation models (CREM)

Exposures	Variables	Uncongested regime				Congested regime	
		Low-density		High-density			
		Coef. (β)*	*Sig.*	*Coef.* (β)	*Sig.*	*Coef.* (β)	*Sig.*
Traffic flow	v	1.83×10^{-2}	0.020	-2.82×10^{-2}	0.000	5.39×10^{-2}	0.001
	k_e	-1.64×10^{-2}	0.026	-	-	-	-
Geometric features	S	1.09×10^{-2}	0.000	2.24×10^{-3}	0.036	1.25×10^{-3}	0.045
	h^{**}	$\pm7.48\times10^{-3}$	0.025	$\pm1.21\times10^{-2}$	0.024	$\pm2.18\times10^{-2}$	0.001
Ambient conditions	*Light*	2.08×10^{-1}	0.035	1.55×10^{-1}	0.000	-	-
	Day	-	-	1.63×10^{-1}	0.007	1.85×10^{-1}	0.014
Intercept value		-1.5	0.000	1.65	0.000	-2.45	0.000
Model test	-2Log Likelihood	1828.2		515.9		407.9	
	Chi-square	51.13		83.59		11.05	
	Significance	0.000		0.000		0.036	

* *Coef.*: estimated coefficient; **+/- correspond to upgrade and downgrade, respectively.

In low-density uncongested regime, CREM involves five factors; speed v, traffic density k_e, horizontal displacement S, variation in road elevation h and ambient *Light*. Amongst them, variables except of k_e have positive contributions on the *Odds* of crash. By contrast, crash risk would be increasing with the decrease in k_e. Such findings indicate that high speed coupled with nighttime and frequent variation in geometric features would increase crash risk. In this sense, the ways to remind driver attention (*e.g.*, driver warning system) and to control speed (*e.g.*, variable speed limit) can be regarded as two effective countermeasures at low flow rate, especially at the segments with poor design consistency and in nighttime conditions.

When traffic flow increases, the inter-vehicle interaction becomes more intensive. In high-density uncongested regime, speed v has a negative contribution to crash risk. Combined with the findings in low-density uncongested regime, the results can support the convex downward crash rate tendency following traffic density in uncongested regime (Lord *et al.*, 2005; Wu *et al.*, 2013). Meanwhile, based on the absolute value of coefficient, the contribution of speed v on crashes gets more significant. On the contrary, the effect of horizontal geometry on crashes is on decrease while the change of vertical geometry becomes more sensitive to crash risk, with the variation of traffic conditions.

With the further increase of traffic density, congested flow appears. In this condition, speed v is more significantly related to crash risk compared to uncongested flow. In the meantime, its coefficient gets positive again, which may be induced by the fact that a major crash events are virtually observed around the onset of breakdown. In contrast to speed v, the effect of horizontal geometry on crashes is further decreasing as opposed to a significant rise in contribution of vertical geometry on crash risk. In addition, the effect of day type on crashes shows more important in comparison with high-density uncongested regime. On holidays, there are more drivers unfamiliar with driving conditions. Therefore, one possible cause is the inappropriate reaction of drivers to the unexpected variation in traffic conditions.

The statistic significance of CREM in congested flow seems poor in comparison to uncongested flow, which is likely induced by the limitation of variables. Generally, a typical traffic characteristic in congested flow is traffic oscillation characterized by recurring decelerations followed by accelerations. In this regard, average traffic variables may not be the optimal index to reveal the natural traffic characteristics of congested flow.

5.4 Model Validation

As explained by Hosmer and Lemeshow (2004), R^2 of logistic regression is a pseudo value and the goodness-of-fit of models should be assessed by comparing the observed value with the predict value. Hence, the CREM is validated through calculating the relative risk of observing a crash versus not. To minimize the misclassification rate, the *Odds* value (for Y=1) of 1.0 is taken as a threshold *Odds* for hazardous condition (hazard if *Odds*≥1.0 and vice versa).

Based on the coefficients in Table 6 and Equation (8), the *Odds* of crash before the time of crash sample (in Section 5.2) are estimated as shown in Figure 6. In this figure, geometric design of this section is roughly demonstrated in the left and these *Odds* values are classified into several levels represented by different colors as shown in right. Meanwhile, the temporal-variations of traffic variables (*i.e.*, v and k_e) at detector #0335 which is closest to the location of crash are exhibited in the above of this figure.

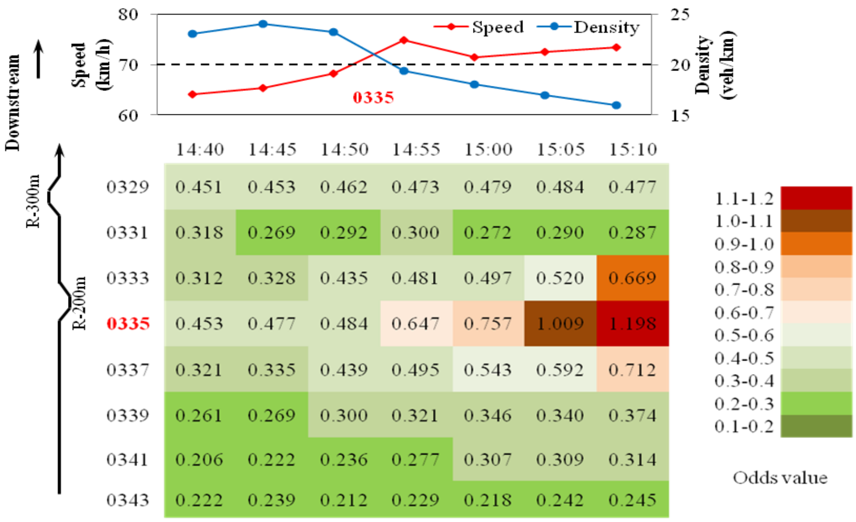

Figure 6. Time-space variation of the *Odds* of crash preceding crash

It is obvious that the *Odds* value at detector #0335 gets greater than 1.0 from 15:05. Compared to the recorded time (15:12), nearly 7 minutes before crash occurrence can be found to be hazardous conditions. For traffic management, this finding is significant since it may provide leverage in terms of time to be able to predict and avoid an impending crash.

The *Odds* values in the same traffic condition prior to other crashes are estimated as shown in the example of Table 7. Based on the estimated crash risk, it is observed that 86.6% of total crash events can be correctly predicted in hazardous conditions. In the same way, the predictive performances of CREM in high-density uncongested and congested regimes are finally discovered to be 80.5% and 70.2%, respectively.

Table 7. Estimation on the *Odds* of crash in low-density uncongested regime

ID of crash	Observed values					Estimated *Odds*	Hazardous condition (*Odds*>1.0)
	v(km/h)	k_e(veh/km)	S(m)	h(m)	*Light*		
1	86.4	8	7.69	-4.50	1	1.23	Yes
2	80.1	20	5.97	-0.40	0	0.741	No
3	92.0	13	11.1	0.28	0	1.10	Yes
4	62.7	20	0.00	-15.0	0	0.620	No
5	87.3	13	0.00	1.37	0	1.06	Yes
6	88.7	19	0.00	1.50	0	1.05	Yes
7	103.4	7	30.2	1.14	1	2.28	Yes
8	97.5	10	12.1	2.50	1	1.61	Yes
…	…	…	…	…	…	…	…
Number of crash estimated in hazard				Total number of crash			Percentage
209				181			86.6%

Figure 7 illustrates the time-space variation of the *Odds* values on May 15, 2009 (non-crash day) at the same section during 8:00 to 8:30. Likewise, geometric design of this section and the temporal variation in speed at detector #0335, where some high *Odds* values (in high-density regime) are observed, are also shown. Most conditions are found to be not hazardous (*Odds*<1.0) except some conditions at detector #0335 that locates in a small curve. Furthermore, the *Odds* are discovered to be relatively lower at sections far from small curves. In this way, the potential hazardous locations can be identified, and they may be flagged with warnings by variable message signs (VMS) in order to remind drivers to pay attention.

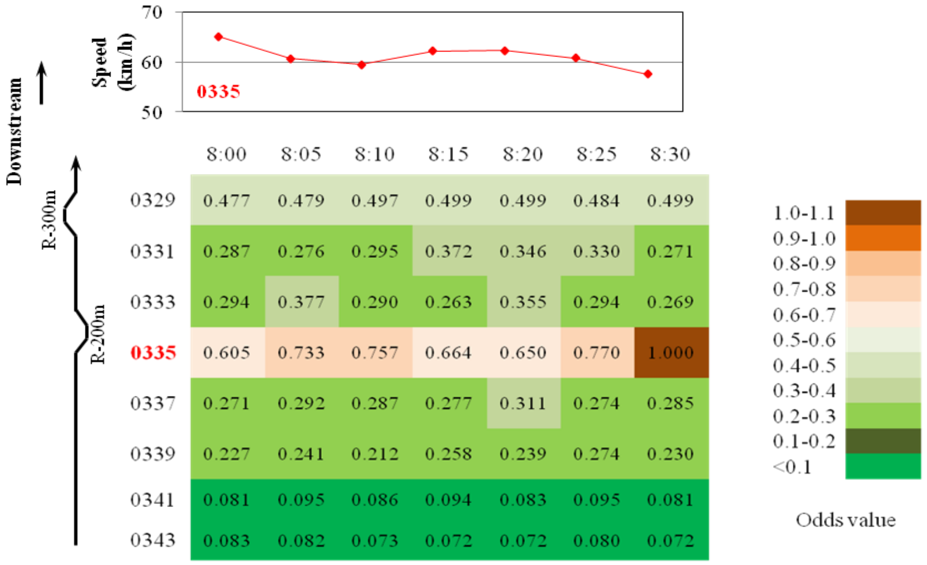

Figure 7. Time-space variation of the *Odds* of crash on a non-crash day

Predictive performance of the above CREM may seem not so perfect but it is worth mentioning that this model does not consider any variables related to driver factors and errors. In this study, it is difficult to obtain these variables based on the original dataset which are mostly collected at aggregated level. For the same reason, the short-term turbulence of traffic flow cannot be reflected appropriately by average traffic variables in 5-min. In this sense, the predictive performance of this model seems to be acceptable. Furthermore, it is evident that the modeling strategy by traffic conditions is more reliable compared to the previous studies in terms of predictive performance (*e.g.*, the model of 59% predictive power in Abdel-Aty and

Pemmanaboina, 2006), not to mention more influencing factors are involved in this study. Therefore, it is reasonable to believe that the analysis in this study including concept and technique is promising considering the application in traffic management, even if a substantial effort is further required to adapt such analysis into practice.

6. CONCLUSIONS AND FUTURE WORK

This paper presented a crash risk estimation model (CREM) for basic segment of Nagoya Urban Expressway, considering the interaction of geometry, traffic flow and ambient conditions. For identifying the effects of various factors on crashes, a matched case-control study was adopted. Based on these case-control samples, the CREM was finally developed by applying conditional logistic regression. The model was further found to be significant in statistics and of acceptable goodness-of-fit, with 86.6%, 80.5% and 70.2% of predictive performance in low-density, high-density uncongested and congested regimes, respectively.

With the increase in traffic density, the significance of horizontal geometry affecting crash is on the decrease. In contrast, the effect of vertical geometry on crash risk becomes more significant. Due to the more powerful inter-vehicle interaction, operating speed gets more sensitive to crash risk when traffic density increases. Ambient conditions are another non-negligible exposure. Generally, nighttime and holiday may increase crash risk relative to daytime and weekday, respectively.

The potential benefits of integrating the model in dynamic traffic controls for safety are numerous. Based on the model by traffic conditions, crash risk can be estimated on a real-time basis. Once a condition is identified as hazardous, it may be flagged with warnings by variable message signs (VMS). Furthermore, the concept of variable speed limit could be used in order to countervail crash risk. Meanwhile, the findings of quantitative effects of influencing factors on crashes may help prioritize countermeasures and recommend some specific methods for smoothing hazardous conditions. In addition, the safety performance of an adopted countermeasure may be estimated in advance by referring to this model.

For more accurate identification of the causal relationship between crash risk and various factors, further studies are required by using high-sample-size data (*e.g.,* crash events and geometric features). Besides, as explained before, the variables to reflect the short-term turbulence of traffic flow are highly recommended for improving the current CREM. Furthermore, this study did not distinguish between individual lanes, while the variations in speed and flow separately across lanes are significantly related to crashes (Golob and Recker, 2004). Finally, as suggested by Christoforou *et al.* (2011), the conditions preceding crash events are different by type of crash, future analysis in this regard is necessary.

ACKNOWLEDGMENTS

Authors are very grateful to Nagoya Expressway Public Corporation for their data provision and other supports for this analysis.

REFERENCES

Abdel-Aty, M., Pemmanaboina, R. (2006) Calibrating a real-time traffic crash risk estimation model using archived weather and ITS traffic data. *IEEE Transactions on Intelligent Transportation System*, 7(2), 167-174.

Abdel-Aty, M., Pande, A. (2007) Crash data analysis: Collective vs. individual crash level approach. *Journal of Safety Research*, 38(5), 581-587.

Bajwa, S., Warita, M., Kuwahara, M (2010) Effects of road geometry, weather and traffic flow on road safety. *Proceedings of the 15th International Conference of Hong Kong Society for Transportation Study*, 773-780.

Christoforou, Z., Cohen, S., Karlaftis, M. G. (2011) Identifying crash type propensity using real-time traffic data on freeways. *Journal of Safety Research*, 42, 43-50.

Golob, T. F., Recker, W. W. (2004) A method for relating type of crash to traffic flow characteristics on urban freeways. *Transportation Research Part A*, 38(1), 53-80.

Hikosaka, T., Nakamura, H. (2001) Statistical analysis on the relationship between traffic accident rates and traffic flow conditions in basic expressway sections. *Proceedings of the 21st Japan Society of Traffic Engineering Meeting*, 173-176.

Hossain, M., Muromachi, Y. (2012) A bayesian network based on framework for real-time crash prediction on the basic freeway segments of urban expressways. *Accident Analysis & Prevention*, 45, 373-381.

Lee, C., Saccomanno, F., Hellinga, B. (2002) Analysis of crash precursors on instrumented freeways. *Transportation Research Record*, 1784, 1-8.

Lewallen, S., Courtright, P. (1998) Epidemiology in practice: case-control studies. *Community Eye Health*, 11(28), 57-58.

Lord, D., Manar, A., Vizioli, A. (2005) Modeling crash-flow-density and crash-flow-V/C ratio relationships for rural and urban freeway segments. *Accident Analysis & Prevention*, 37, 185-199.

Ng, J. C. W., Sayed, T. (2004) Effect of geometric design consistency on road safety. *Canadian Journal of Civil Engineering*, 31, 218-227.

Oh, C., Oh, J., Ritchie, S. (2005) Real-time hazardous traffic condition warning system. *IEEE Transactions on Intelligent Transportation System*, 6(3), 8.

Hosmer, D. W., Lemeshow, S. (2004) Applied Logistic Regression. *Wiley-Interscience*.

Panda, A., Abdel-Aty, M. (2006) Comprehensive analysis of the relationship between real-time traffic surveillance data and rear-end crashes on freeways. *Transportation Research Record*, 1953, 31-40.

Rengarasu, T. M., Hagiwara, T., Hirasawa, M. (2009) Effects of road geometry and cross section variables on traffic accidents using homogeneous road segments. *Transportation Research Record*, 2102, 34-42.

Shively, T. S., Kockelman, K., Damien, K. (2010) A Bayesian semi-parametric model to estimate the relationships between crash counts and roadway characteristics. *Transportation Research Part B*, 44(5), 699-715.

Transportation Research Board. (2010) Highway Capacity Manual. Washington D.C.

Wu, Y., Nakamura, H., Asano, M. (2012) A comparative study on crash rate characteristics at different intercity expressway facility types. *Proceedings of the 46th IP meeting of Japan Society of Civil Engineering*, CD-ROM.

Wu, Y., Nakamura, H., Asano, M. (2013) A comparative study on crash characteristics between urban and intercity expressway basic segments. *Proceedings of the 13th Word Conference on Transport Research*, CD-ROM.

Zheng, Z. D., Ahn, S., Monsere, C. M. (2010) Impact of traffic oscillation on freeway crash occurrences. *Accident Analysis & Prevention*, 42, 626-636.