

A comparative study on crash-influencing factors by facility types on urban expressway

Yong Wu · Hideki Nakamura · Miho Asano

Received: 1 July 2013 / Revised: 15 October 2013 / Accepted: 18 October 2013 / Published online: 30 November 2013
© The Author(s) 2013. This article is published with open access at Springerlink.com

Abstract This study aims at identifying crash-influencing factors by facility type of Nagoya Urban Expressway, considering the interaction of geometry, traffic flow, and ambient conditions. Crash rate (CR) model is firstly developed separately at four facility types: basic, merge, and diverge segments and sharp curve. Traffic flows are thereby categorized, and based on the traffic categories, the significances of factors affecting crashes are analyzed by principal component analysis. The results reveal that, the CR at merge segment is significantly higher than those at basic and diverge segments in uncongested flow, while the value is not significantly different at the three facility types in congested flow. In both un- and congested flows, sharp curve has the worst safety performance in view of its highest CR. Regarding influencing factors, geometric design and traffic flow are most significant in un- and congested flows, respectively. As mainline flow increases, the effect of merging ratio affecting crash is on the rise at basic and merge segments as opposed to the decreasing significance of diverging ratio at diverge segment. Meanwhile, longer acceleration and deceleration lanes are adverse to safety in uncongested flow, while shorter acceleration and deceleration lanes are adverse in congested flow. Due to its special geometric design, crashes at sharp curve are highly associated with the large centrifugal force and heavy restricted visibility.

Keywords Crash-influencing factors · Crash rates · Principal component analysis · Facility types · Urban expressway

1 Introduction

Improving traffic safety is a worldwide issue to be relieved urgently. Crash characteristics and their influencing factors, as the theoretical basis for safety improvement, may provide direction for policies and countermeasures aimed at smoothing hazardous conditions. For a better understanding of crash-influencing factors, researchers have continually sought ways through an extensive array of approaches, and the most prominent one is crash data analysis [1]. The conventional approaches have established statistical links between crash rate (CR) and its explanatory factors [2, 3]. In the analyses, traffic flows are generally represented by low-resolution data that is collected at a highly aggregated level, e.g., hourly or daily flows. Geometric features are primarily considered the hierarchy of radius or slope [4, 5]. Meanwhile, several studies have suggested that crashes are associated with the interaction of geometry, traffic flow, and ambient conditions [6]. However, most existing studies investigated the factors individually and the related CR models were developed based on single factor only. As a result, it is inadequate to identify the nature of individuals through aggregated analysis only, since the conditions preceding individual crashes are virtually different from each other [1].

Considering the insufficiency of CR analysis above, some studies have tried to identify crash characteristics at individual level, in an effort to predict crash risk on a real time basis [7–9]. Through these studies, the effect of traffic flow on crash risk has been well analyzed. In theory, the concept of real-time crash prediction exhibits huge promise for the application of proactive traffic management strategies for safety.

However, the combined effects of geometry, traffic flow, and ambient conditions on crashes still have not been well

Y. Wu (✉) · H. Nakamura · M. Asano
Department of Civil Engineering, Nagoya University,
C1-2(651) Furo-cho, Chikusa-ku, Nagoya 464-8603, Japan
e-mail: gerry_woo@126.com

analyzed through the above studies. Furthermore, these papers primarily developed crash model for the whole traffic conditions, which may conflict with the fact that the influence of traffic flow on crashes may vary when traffic conditions change. In addition, even if crash characteristics are found out to be dependent on facility type that is composed of uniform segment individually, e.g., basic, merge, and diverge segments [2], the existing studies are focused on the entire route of intercity expressway without segmentation.

Another cause for the limited predictive performance of existing models is the inadequacy of analytic process [10]. As for statistical methods, the significance and independence of explanatory variables should be identified in advance for the reliability of statistics. Whereas, many previous studies paid little attention to this point and incorporated the potential influencing factors into crash modeling directly.

Urban expressway is one common type of separated highway with full control of access in large cities in Japan. Generally, it is composed of various facility types where geometric features and traffic characteristics are often different from each other. Correspondingly, crash characteristics and their influencing factors may also be different by facility type. In the meantime, compared to intercity expressway, crash characteristics and their related influencing factors of urban expressway are different [11]. Necessarily, urban expressway deserves to be analyzed independently and its crash characteristics should be identified based on specific facility type.

Given the problems of existing studies, the objective of this paper is to investigate crash characteristics based on CR models and their influencing factors by facility type on urban expressway. Meanwhile, the causes are identified by considering the interaction of geometry, traffic flow, and ambient conditions. Besides, geometric features are identified considering the driver-vehicle-roadway interaction. The significances of these factors affecting crashes are compared at different facility types using principal component analysis (PCA). Their influencing mechanisms are further discussed. In essence, this study can be regarded as a proactive analysis for crash risk prediction model in the future.

2 Study sites and datasets

2.1 Study sites

The test bed of this study is Nagoya Urban Expressway network (NEX) as shown in Fig. 1. Up to December 31, 2009, this network was about 69.2 km × 2 (two directions) in total length with over 250 ultrasonic detectors

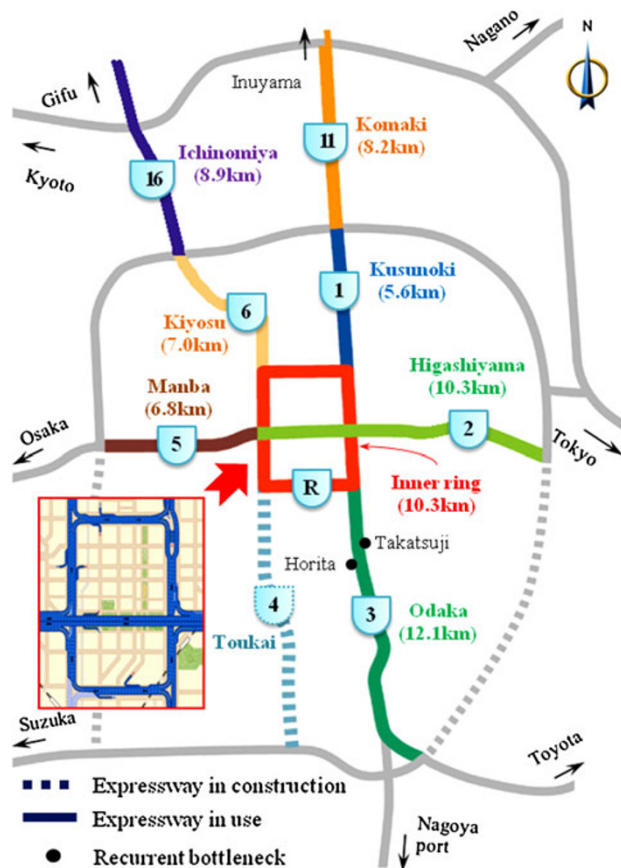


Fig. 1 Schematic map of NEX network (2009) (Source Nagoya Expressway Public Corp., modified by authors)

installed with an average spacing of 500 m (varied in 250–750 m) on mainline. Most routes are 4-lane roadways (2-lane/dir), except the inner ring (Route no. R) that is one-way roadway and where the number of lanes differs (2–5) with the change of ramp junctions. In the limited areas, such as the links of other routes to the inner ring, small curves are designed. In this network, two recurrent bottlenecks are located along Odaka line (Route no. 3).

Five databases are used in this study; (1) crash records with the occurrence time in minutes, the location in 0.1 km and the weather and pavement conditions; (2) detector data including traffic volume q , average speed v , and occupancy occ per 5 min; (3) geometric design and the location of detector in 0.01 km; (4) traffic regulation records for incidents (e.g., crash, working, and inclement weather) including the locations and periods of temporal lane and cross-section closures; and, (5) daily sunrise and sunset time records in Nagoya. Here, it is worth noting that detector data are processed for the whole cross section of each direction. The period of the data above is for 3 years (2007–2009) except for those on Kiyosu line (Route no. 6) that was opened from December 1, 2007.

2.2 Segmentation of facility types

Basic segment is extracted outside the 500 m up- and down-stream of ramp junctions considering the experience in Japan [12]. Correspondingly, merge or diverge segments are regarded as the sections inside the 500 m up- and down-stream of on- and off-ramps, respectively. The segmentation methods are shown in Fig. 2. Other than these segments, there is a special geometric design in NEX, curve with small radius. Figure 3 explains CR statistics dependent on radius. Obviously, compared to other segments, much higher CR exists in the curves with radius smaller than 100 m. Thus, these curves are defined as sharp curves and regarded as another distinct facility type of NEX. Given the limitation of segment samples available, basic, merge, and diverge segments, and sharp curve will be analyzed in this study.

The cross sections of inner ring are diverse and the length of individual layouts, i.e., 2-, 3-, 4-, or 5-lane, is not

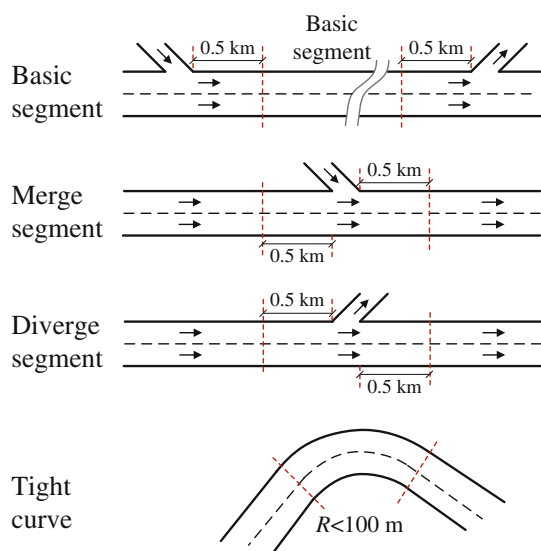


Fig. 2 Segmentation of facility types

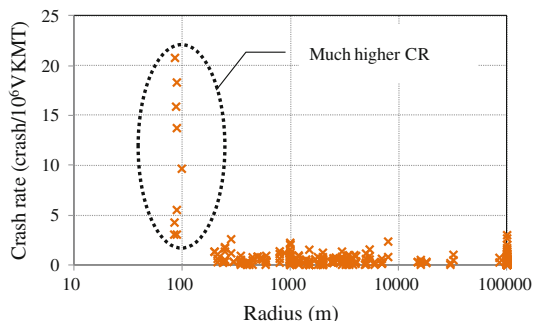


Fig. 3 Distribution of crash rate to radius

Table 1 Geometric statistics of facility types

| Facility type | No. of segments | Total length (km) |
|-----------------|-----------------|-------------------|
| Basic segment | 38 | 56.6 |
| Merge segment | 28 | 20.9 |
| Diverge segment | 35 | 25.2 |
| Sharp curve | 9 | 2.5 |

enough to be separately analyzed. Meanwhile, all of the sharp curves along Inner ring are 2-lane roadway. In this regard, only 2-lane segments are analyzed in this study and the geometric statistics by facility types are summarized in Table 1.

3 Methodology

3.1 Data extraction

3.1.1 Detector data

In principle, detectors can count the number of vehicles at their locations only. In such case, the “coverage area” of detector is defined for estimating traffic conditions at crash locations through detector data. At basic segment, the boundary of two consecutive coverage areas is defined at the midpoint between two neighboring detectors. At merge and diverge segments, it is bounded at the ramp-junction point, and one segment can be divided into up- and down-stream areas. Each sharp curve can be matched with a single detector. Note that the time of crash is recorded by road administrators after the crash occurrence. In reality, it does not correspond to occurrence time exactly. For this reason, data within small time before crashes should be rejected to avoid mixing up crash-influencing and crash-influenced data. Therefore, the latest data at least 5 min before the recorded time are accepted after the exclusion of invalid data and the data within lane and section-closure intervals in advance.

3.1.2 Geometric features

Design consistency is the conformance of geometry of a highway with driver expectancy, and its importance and significant contribution to road safety is justified by understanding the driver–vehicle–roadway interaction [13] that may vary at individual locations in nature. In this regard, geometric variation in the upstream of crash location is proposed to reflect the effect of geometry on crashes. Considering the length of detector coverage area, the following variables in 500 m distance are extracted [12].

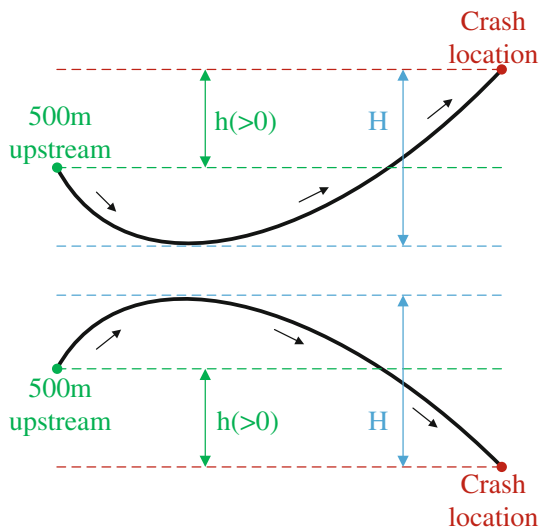


Fig. 4 Variation in road elevation

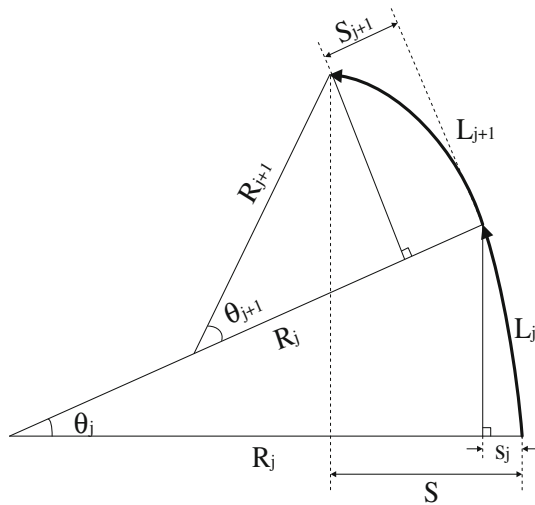


Fig. 5 Horizontal displacement

(1) Variation in road elevation h between the crash location and its 500 m upstream, and the maximum elevation difference H in this 500 m distance (Fig. 4).

(2) Horizontal displacement S . Radius is impossible to describe a section composed of various curves. On the other hand, centrifugal force is also associated with the horizontal displacement s in the direction of tangent to the curve j (Fig. 5). In such case, S in the 500 m distance ($\sum s_j$) is adopted and calculated by the following equations.

$$\theta_j = \frac{L_j}{R_j} \quad (0 > ; \theta_j \leq \pi/2), \tag{1}$$

$$s_j = R_j(1 - \cos \theta_j), \tag{2}$$

where j is the ID of curve. R_j , θ_j , L_j , and s_j correspond 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} ($I_{CF} = Sv^2$) to reflect the combined effect of speed v and horizontal displacement S , while it is not centrifugal force.

(4) Index of space displacement I_{SD} . I_{SD} ($I_{SD} = SH$) is used to reveal the comprehensive geometric features induced by horizontal and vertical variation in this study.

The geometric data above are collected every 0.1 km as crash is recorded in a unit of 0.1 km. Besides, these data are also extracted at the location of detector that is the common link between crash and detector data. Table 2 summarizes the process of data collection.

3.1.3 Ambient conditions

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

3.1.4 Data matching

The related detector data, geometric features, and ambient conditions for individual crashes are matched as exemplified in Table 3. The crashes matching with invalid detector data and within lane and cross-section closure intervals are excluded in advance. As a result, a total of 1,591 crashes remain for the following analysis.

Table 2 Example of geometric variations collection

| Route no. | Direction* | Kilopost** | h (m) | H (m) | S (m) | I_{SD} (m ²) |
|-----------|------------|------------|---------|---------|---------|----------------------------|
| 1 | SB | 0.0 | -4.63 | 5.49 | 0.78 | 4.30 |
| 1 | SB | 0.1 | -7.90 | 8.49 | 3.91 | 33.2 |
| 1 | SB | 0.2 | -10.6 | 11.5 | 6.08 | 69.9 |
| 1 | SB | 0.21 | -11.5 | 11.8 | 8.88 | 104.7 |
| 1 | SB | 0.3 | -15.3 | 15.3 | 9.60 | 146.9 |
| 1 | SB | - | - | - | - | - |
| 1 | SB | 6.4 | 10.2 | 10.9 | 5.15 | 56.1 |

*SB South-bound

** 0.21: the Kilopost of detector #0101

Table 3 Examples of data matching for individual crashes

| Crash ID | Traffic characteristics | | | | Geometric features | | | | | | | Ambient conditions | | | |
|----------|-------------------------|------------|------|------|----------------------------|---------|---------|---|----------------------------|-------------|-------------|--------------------|---------|----------|----------|
| | q (veh/5 min) | v (km/h) | MR | DR | Facility type ^a | h (m) | H (m) | I_{CF} (km ³ /h ²) | I_{SD} (m ²) | L_A^b (m) | L_D^b (m) | Light | Weather | Pavement | Day type |
| 1 | 139 | 88.7 | – | – | B | 1.5 | 1.5 | 0 | 0 | – | – | Day | Sunny | Dry | Holiday |
| 2 | 96 | 95.0 | – | 0.02 | D | 0.5 | 1.3 | 302 | 45 | – | 220 | Day | Sunny | Dry | Weekday |
| 3 | 29 | 80.3 | – | – | S | 10.3 | 12.7 | 126 | 249 | – | – | Night | Cloudy | Wet | Holiday |
| 4 | 154 | 82.9 | 0.07 | – | M | 1.5 | 1.5 | 0 | 0 | 200 | – | Day | Cloudy | Dry | Weekday |

B, M, D, and S basic, merge, and diverge segments, and sharp curve, respectively, L_A length of acceleration lane at merge segment, L_D length of deceleration lane at diverge segment

3.2 Classification of traffic conditions

Congested flow, characterized by traffic oscillation, has different features from uncongested flow. It is necessary to make a distinction between two traffic regimes. Figure 6 shows the traffic volume–speed diagram at Horita on-ramp junction, one typical bottleneck in NEX. The speed of 60 km/h, corresponding to maximal flow is defined as the critical speed v_c that is used for classifying un- and congested flows [2, 14]. Besides, the corresponding value at another bottleneck (Takatsuji on-ramp junction) is also found out around 60 km/h.

The value of 60 km/h would be regarded as the related index at basic and diverge segments, since no bottleneck can be virtually found at both segments in NEX. At sharp curve, a threshold speed of 45 km/h is selected in general for classifying two traffic regimes based on traffic flow-speed diagram at Tsurumai curve (Fig. 7). The value is further checked at other sharp curves, and it is found out to be reliable for classifying un- and congested flows basically.

To reflect the variation in traffic characteristics, each traffic regime is further sub-classified. It is evident that speed has a high variance at low flow rates (see Figs. 6 and 7). Besides, occupancy is not a commonly used index. Thus, traffic density k calculated by Eq. (3) is proposed to be the measure of effectiveness to further classify the traffic conditions. In view of the number of crash samples available, the aggregation intervals of k are set as 10 and 30 veh/km for un- and congested flows, respectively.

$$k_{ei} = \frac{12 \times q_i}{v_i}, \tag{3}$$

where q_i and v_i denote traffic flow and average speed in 5 min # i , respectively. k_{ei} corresponds to the calculated traffic density in this 5 min.

3.3 Calculation of crash rate (CR)

CR for traffic condition n can be calculated by the following equation:

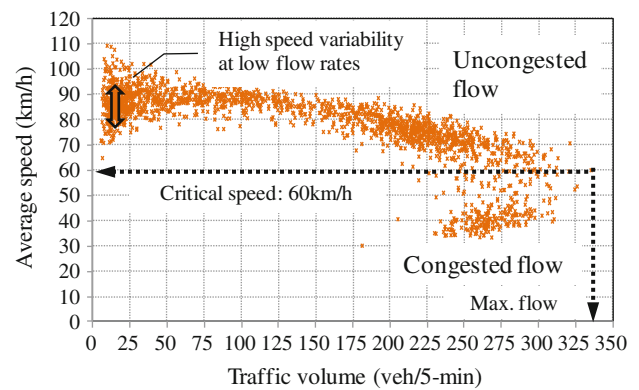


Fig. 6 q - v diagram at Horita on-ramp junction

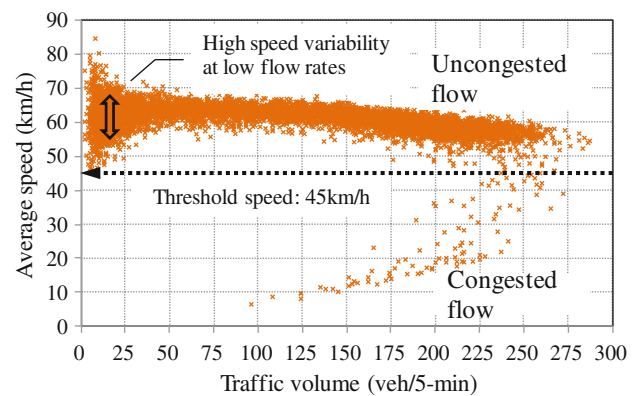


Fig. 7 q - v diagram in Tsurumai curve

$$CR_n = \frac{NOC_n \times 10^6}{\sum Q_{nl}L_l}, \tag{4}$$

where n and l are the ID of traffic condition and coverage area, respectively; NOC_n is the number of crashes for traffic condition n . $Q_{nl}L_l$ is the value of vehicle kilometers traveled (VKMT) in detector coverage area l for traffic condition n .

3.4 Principal component analysis (PCA)

PCA is a powerful tool for reducing a large number of observed variables into a small number of artificial variables that account for most of the variance in the dataset [15]. In general, through orthogonal transformation, a set of observations of possibly correlated variables can be converted into a set of values of linearly uncorrelated variables. Those converted values are defined as principal components. Technically, a principal component can be regarded to be a linear combination of optimally weighted observed variables [15]. As a result, the components are ranked in the order of accounting amount of total variance in the observed variables. Then, two criteria are generally available to select the number of component extracted: (1) 80 % rule, the extracted components should be capable to explain at least 80 % of the variance in the original dataset. (2) Eigen value rule, only components whose eigen values are over 1.0 can be retained.

4 Crash rate estimation models

In the following, the differences of crash characteristics by facility type are investigated by comparing CR models based on traffic conditions.

4.1 Uncongested flow

Figure 8 gives the CR tendency following traffic density k by facility type in uncongested flow. It is evident that sharp curve has a special characteristic compared to other segments. Its CR is the highest among four facility types at low-density stages. Then, the value follows a decreasing tendency to k . In contrast, the CR at other segments increases as k increases. Such phenomenon may be related to the design of small radius for sharp curve. Such geometric design can result in high centrifugal force that can act on the vehicle and tries to push it to the outside of the curve. Furthermore, higher speed may result in higher centrifugal force.

Regarding the differences at other segments, CR at merge segment increases rapidly at high-density stages and gets much higher compared to basic and diverge segments. The results of paired t -test at the three facility types in Table 4 also reveal that CR at basic/diverge segments is significantly lower than that at merge segment, while they are not significantly different from each other between basic and diverge segments. At merge segment, merging maneuvers can result in slow-down and lane-changing behaviors for mainline traffic. These interruptions may increase the possibility of vehicle conflicts. Such possibility can further increase with an increase in k .

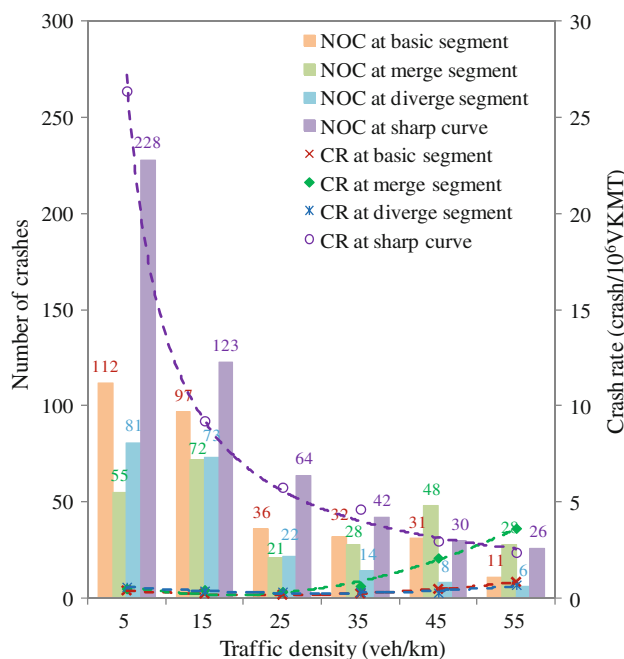


Fig. 8 CR comparison in uncongested flow

Table 4 T -test of CR in uncongested flow

| Paired | t -Value | df | Sig. |
|------------------------------------|------------|----|-------|
| Pair 1: basic and merge segments | -2.781 | 5 | 0.019 |
| Pair 2: basic and diverge segments | -1.070 | 5 | 0.310 |
| Pair 3: merge and diverge segments | 2.320 | 5 | 0.043 |

Table 5 summarizes the CR regression models as function of k as well as the goodness-of-fit of models at four facility types. At sharp curve, the model is power function while they are quadratic functions at other facility types. All of the models and variables are significant at 95 % confidence level (not shown in Table 5). Regarding quadratic models, CR at merge segment is most sensitive to the increase in k , more than three times of CR increases as that at basic and diverge segments by the increase in one unit of k .

4.2 Congested flow

Figure 9 describes the differences of CR distribution to k by facility type in congested flow. It appears that CR follows increasing tendencies to k at four facility types. In contrast to other segments, sharp curve still has the highest CR in congested flow while no statistical regression model is developed at this facility type due to the limited crash samples. Since the differences of CR tendency at other segments are not clear in Fig. 9, a paired t -test is conducted as shown in Table 6. The results indicate that there is no

Table 5 CR regression models in uncongested flow

| Facility type | Sample size | Model ^a |
|---------------|-------------|--|
| B | 319 crashes | $CR = 6.81 \times 10^{-4}k^2 - 3.23 \times 10^{-2}k + 0.541, R^2 = 0.998, k(CR_{\min}) = 24$ |
| M | 251 crashes | $CR = 2.55 \times 10^{-3}k^2 - 9.29 \times 10^{-2}k + 1.01, R^2 = 0.983, k(CR_{\min}) = 20$ |
| D | 204 crashes | $CR = 5.68 \times 10^{-4}k^2 - 3.34 \times 10^{-2}k + 0.747, R^2 = 0.849, k(CR_{\min}) = 28$ |
| S | 513 crashes | $CR = 132.2k^{-0.983}, R^2 = 0.992$ |

^a $k(CR_{\min})$: traffic density corresponding to the minimal CR

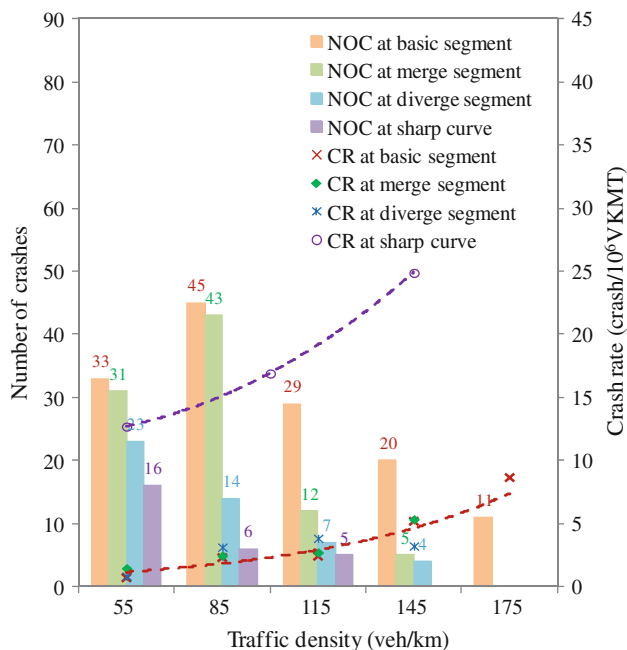


Fig. 9 CR comparison in congested flow

Table 6 *T*-test of CR in congested flow

| Paired | <i>t</i> -Value | df | Sig. |
|------------------------------------|-----------------|----|-------|
| Pair 1: basic and merge segments | -2.448 | 4 | 0.092 |
| Pair 2: basic and diverge segments | -0.153 | 4 | 0.888 |
| Pair 3: merge and diverge segments | 0.325 | 4 | 0.767 |

significant difference of CR at basic, merge, and diverge segments. Such finding may imply that the effect of facility type on crashes is reduced in congested flow. For this reason, CR model is developed by combining the three facility types in order to increase the number of crash samples for reliability. As demonstrated in Table 7, an exponential function is adopted and it fits well to the combined CR tendency. The model and its variables are also significant at 95 % confidence level, while the results are not shown in Table 7.

Table 7 CR regression model in congested flow

| Facility type | Sample size | Model |
|---------------|-------------|--|
| B + M + D | 513 crashes | $CR = 4.87 \times 10^{-1}e^{0.0155k}, R^2 = 0.924$ |

5 Effects of influencing factors

The analyses above reveal that CR characteristics are different by facility type, which may be related to the different geometric designs and traffic characteristics. However, CR analysis is insufficient to examine a variety of factors by a single model. Instead, PCA is applied and the affecting mechanisms of individual factors are further investigated.

5.1 Introduction of variables

Table 8 explains individual variables combining with its type and some summary statistics. In nature, traffic flow diagram is two-dimensional, and k and v are used together to describe traffic conditions. As for geometric features, h , I_{CF} , and I_{SD} are picked out to reflect the vertical, horizontal, and comprehensive geometric variations, respectively. Dummy variables are referred to incorporate ambient conditions into PCA. A dummy variable usually takes 0 and 1. In this case, weather conditions (over 2 categories) are replaced by pavement conditions (only 2 categories), since two conditions are usually highly related to each other.

At merge and diverge segments, ramp traffic is a significant influencing factor on crashes [16]. This study employs ramp flow ratio to illustrate the interaction between ramp and mainline traffic. Merging ratio (MR) or diverging ratio (DR) is defined as the proportion of on- or off-ramp traffic out of the sum of ramp and mainline traffic, respectively. Meanwhile, the length of acceleration lane L_A or the length of deceleration lane L_D is adopted to reveal the space available provided for merging or diverging maneuvers, respectively.

Table 8 Introduction of individual variables

| Variables | Statistics ^a | | Description |
|-----------|-------------------------|------|---|
| | Max. | Min. | |
| k | 238 | 0 | Traffic density (veh/km) |
| v | 141.0 | 4.7 | Average speed (km/h) |
| MR | 0.78 | 0.00 | Merging ratio |
| DR | 0.60 | 0.00 | Diverging ratio |
| I_{CF} | 1997 | 0.1 | Index of centrifugal force (km ³ /h ²) |
| h | 14.8 | 0.1 | Vertical variation (m) |
| I_{SD} | 1112.7 | 0.0 | Index of space displacement (m ²) |
| L_A | 250 | 100 | Length of acceleration lane (m) |
| L_D | 432 | 100 | Length of deceleration lane (m) |
| Pave | $f(1) = 27.7\%$ | | Equal to 1 if wet pavement, 0 otherwise |
| Light | $f(1) = 27.1\%$ | | Equal to 1 if nighttime, 0 otherwise |
| Day | $f(1) = 29.4\%$ | | Equal to 1 if holiday, 0 otherwise |

f frequency

^a Max./Min.: maximal and minimal values, respectively

5.2 PCA among various facility types

In essence, PCA rotates data by using a linear transformation. Consequently, only the monotonic loadings of factors can be reflected by this approach. For this reason, uncongested flow is further classified into low-and high-density conditions at approximately 25 veh/km in view of the value of k (CR_{min}) as shown in Fig. 10, since there are different monotonicities of CR model in two conditions. As a result, three traffic conditions are analyzed, i.e., low- and high-density uncongested flow as well as congested flow.

5.2.1 Low-density uncongested flow

Table 9 demonstrates PCA results at basic segment in low-density uncongested flow. In terms of the rules introduced in Sect. 3.4, four components are selected and all of the factors can explain at least 80 % of variance in the original dataset in terms of the value of cumulative percent.

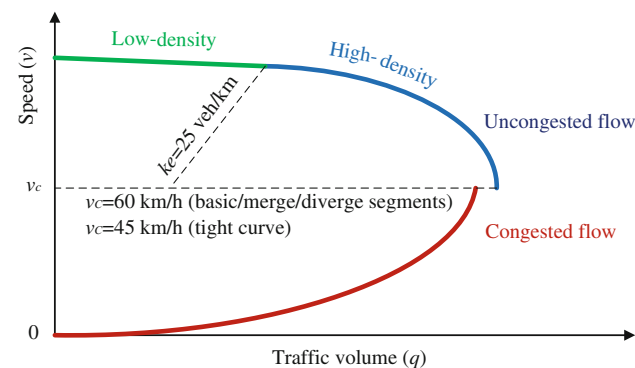


Fig. 10 Classification of traffic conditions

Table 9 PCA results at basic segments

| Variables | Component | | | |
|---------------------|--------------|---------------|--------------|--------------|
| | 1st | 2nd | 3rd | 4th |
| k_e | -0.194 | -0.852 | -0.119 | 0.103 |
| v | 0.285 | 0.182 | 0.798 | -0.086 |
| I_{CF} | 0.953 | 0.005 | 0.053 | -0.122 |
| I_{SD} | 0.959 | 0.011 | -0.044 | 0.090 |
| h | 0.119 | 0.149 | 0.228 | 0.973 |
| Pave | 0.294 | 0.214 | 0.783 | -0.095 |
| Day | 0.139 | 0.093 | 0.190 | -0.468 |
| Light | -0.164 | 0.838 | -0.130 | 0.143 |
| Initial Eigenvalue | 2.12 | 1.54 | 1.37 | 1.12 |
| Percent of variance | 30.3 | 20.2 | 18.2 | 15.1 |
| Cumulative Percent | 30.3 | 50.5 | 68.7 | 83.8 |

The variables highly related to each component are in bold

In low-density uncongested flow, crashes at basic segment are found to be significantly associated with geometric variation (I_{CF} and I_{SD}), traffic density along with ambient light, speed coupled with pavement, and vertical variation h . Geometric variation is the 1st component, as great variation may result in frequent speed reduction. Accordingly, the difficulty for drivers to control vehicle behaviors increases. At low traffic density k , driver’s attention is not high, and some discretionary behaviors may be operated. Such condition combining with the poor ambient light is possible to increase crash risk. Meanwhile, due to the reduced value of tire-pavement friction, high speed v combining with wet pavement can reduce the roadability. In such cases, k and v are two separate components, which can further demonstrate that both variables are not highly interrelated at low flow rate. In addition, vertical variation h has a positive loading because of the increased visibility restriction and the difficulty in maintaining vehicle behaviors for drivers.

Principal components at other segments are analyzed as shown in Table 10. The variables that are significantly related to each component are selected based on their loadings. For judging the relative significance of the same component by facility type, the percent of variance explained by each component is provided as well.

One difference at merge segment from basic segment is that MR combining with the length of acceleration lane L_A becomes a principal component. Meanwhile, day type is found to be significant. In terms of the percent of variance accounted by components, the significance of geometric variation gets lower in contrast to basic segment. Merging traffic is an important influencing factor, since it can induce interruption to mainline traffic. Such interruption may get stronger as MR increases. Besides, higher L_A can provide more space for ramp and mainline traffic to adjust for

Table 10 PCA results in low-density uncongested flow

| Facility type (number of crash) | Item | Principal component | | | | | | | | | |
|---------------------------------|---------------------|---------------------|-------|----------|--------|-------|--------|------|-------|-------|-------|
| | | 1st | | 2nd | | 3rd | | 4th | | 5th | |
| | | F | L | F | L | F | L | F | L | F | L |
| Basic segment (225) | Component | I_{CF} | 0.953 | k | -0.852 | v | 0.798 | h | 0.973 | | |
| | | I_{SD} | 0.959 | Light | 0.838 | Pave | 0.783 | | | | |
| | Percent of variance | 30.3 | | 20.2 | | 18.2 | | 15.1 | | | |
| Merge segment (140) | Component | I_{CF} | 0.867 | k | -0.841 | MR | 0.808 | v | 0.721 | Day | 0.874 |
| | | I_{SD} | 0.901 | Light | 0.869 | L_A | 0.747 | Pave | 0.742 | | |
| | Percent of variance | 20.1 | | 18.8 | | 15.4 | | 13.9 | | 11.9 | |
| Diverge segment (167) | Component | DR | 0.807 | I_{CF} | 0.948 | k | -0.807 | v | 0.892 | L_D | 0.797 |
| | | h | 0.845 | I_{SD} | 0.746 | Light | 0.860 | Pave | 0.848 | | |
| | Percent of variance | 18.9 | | 17.6 | | 16.2 | | 15.3 | | 13.3 | |
| Sharp curve (319) | Component | I_{SD} | 0.854 | I_{CF} | 0.948 | k | -0.820 | Pave | 0.929 | Day | 0.981 |
| | | h | 0.978 | v | 0.753 | Light | 0.794 | | | | |
| | Percent of variance | 23.7 | | 17.5 | | 16.4 | | 13.6 | | 12.7 | |

F factor, L loading

merging behaviors. Regarding the influence of day type on crashes, it may be related to the different vehicle compositions and driver populations between holiday and weekday, while such influence needs a further study to investigate vehicle behaviors at merge segment. As for geometric variation, on ramps in NEX are virtually allocated far from poor alignment like small curve. Thus, it is considered reliable that the significance of geometric variation affecting crashes is lower at merger segment compared to basic segment.

At diverge segment, the most significant difference from basic and merge segments is that the DR and the vertical variation h are related to the 1st component. Higher DR can significantly interrupt mainline traffic since it is necessary to pass through several lanes to move onto the deceleration lane for driving vehicles. Furthermore, higher h can make lane-changing maneuvers more difficult.

Generally, sharp curve has much worse design consistency compared to other segments. Crashes at sharp curve are found to be associated with poor vertical consistency (I_{SD} and h), high horizontal variation I_{CF} along with speed v , low traffic density k in nighttime, wet pavement, and holiday. In NEX, sharp curve is often designed to connect routes with different elevations. Thus, the vertical consistency is fairly poor. Smaller radius along with high v may cause notable centrifugal force. The affecting mechanisms of other component are similar to these at basic, merge, and diverge segments.

5.2.2 High-density uncongested flow

As traffic density increases, the inter-vehicle interaction gets more intensive. The corresponding results of PCA in

high-density uncongested flow are summarized in Table 11. All of the components are of statistical significance.

In the case of high-density uncongested flow, it is distinct that traffic-related variables including k and v become an independent component, as a reflection of the increased interaction of vehicles. Furthermore, in terms of the value of loading, high density not low density is adverse to safety. The finding can further support the results of CR models: CR is decreasing to k in low-density uncongested flow, while it is increasing in high-density uncongested flow.

With respect to the differences by facility type, at merge segment, MR gets to be a factor related to the 1st component due to the increased interruption of ramp traffic with the increase of traffic density. During the variation in traffic conditions, the significance of DR becomes lower than geometric variation at diverge segment. However, in high-density uncongested flow, L_D is more important in contrast to low-density uncongested flow. Once a driver feels the difficulty for lane-changing maneuvers in diverging area, they may move onto the nearest lane to off-ramp in advance in the upstream of diverging area. As a result, the impact of lane-changing maneuvers on mainline traffic gets relatively low. In a sharp curve, crashes are still found to be probable with a decrease in k , which is similar to the tendency of CR model.

5.2.3 Congested flow

With the further increase of traffic density, congested flow appears. In the same way, Table 12 summarizes the results of PCA by facility type in congested flow.

Table 11 PCA results in high-density uncongested flow

| Facility type (number of crash) | Item | Principal component | | | | | | | | | |
|---------------------------------|---------------------|---------------------|-------|-------|--------|-------|--------|-------|-------|------|-------|
| | | 1st | | 2nd | | 3rd | | 4th | | 5th | |
| | | F | L | F | L | F | L | F | L | F | L |
| Basic segment (94) | Component | I_{CF} | 0.983 | k | 0.858 | Day | 0.776 | h | 0.925 | | |
| | | I_{SD} | 0.974 | v | -0.885 | Light | 0.781 | | | | |
| | Percent of variance | 28.9 | | 21.8 | | 17.1 | | 14.4 | | | |
| Merge segment (112) | Component | MR | 0.818 | k | 0.936 | L_A | 0.842 | Day | 0.810 | Pave | 0.963 |
| | | I_{CF} | 0.923 | v | -0.899 | | | Light | 0.713 | | |
| | Percent of variance | 28.3 | | 18.0 | | 13.5 | | 12.3 | | 10.4 | |
| Diverge segment (37) | Component | I_{CF} | 0.970 | DR | 0.904 | k | 0.793 | h | 0.733 | Pave | 0.913 |
| | | I_{SD} | 0.977 | L_D | 0.772 | v | -0.704 | | | | |
| | Percent of variance | 21.8 | | 19.8 | | 16.8 | | 13.7 | | 13.5 | |
| Sharp curve (122) | Component | I_{CF} | 0.723 | k | -0.901 | Pave | 0.901 | Day | 0.869 | | |
| | | I_{SD} | 0.962 | v | 0.872 | | | | | | |
| | Percent of variance | 33.2 | | 21.9 | | 15.3 | | 13.9 | | | |

Table 12 demonstrates that the effect of traffic flow on crashes get more important in congested flow, compared to that in uncongested flow. Except merge segment, the significance of traffic flow affecting crashes is the highest. Based on the percent of variance, the influence of geometric design is further decreasing.

Regarding the differences by facility type, crashes at merge segment are found to be positively associated with smaller L_A , not higher L_A . For congested flow, smaller L_A may increase the difficulty of adequate speed adjustment for merging and lane-changing maneuvers. Besides, based on the loading of day type, weekday not holiday is a significant factor. It is likely related to higher percentage of heavy vehicles on weekday that may induce more frequent shockwave in congested flow. At diverge segment, as similar to merge segment, weekday is also a significant factor. Meanwhile, smaller L_D not higher L_D is adverse to safety. At sharp curve, poor ambient light can significantly restrict visibility, while visibility is critical for driving in small inter-vehicle spacing. Thus, ambient light becomes another important factor in congested flow compared to high-density uncongested flow.

From the analyses above, geometric features are found out to be the most significant influencing factor in uncongested flow. In this sense, the different CR characteristics by facility type in uncongested flow may be significantly associated with the variation in geometry. Poor design consistency induced by small radius is the potential cause for the highest CR in sharp curve. Ramp traffic can interrupt mainline traffic, and longer acceleration lane may provide longer interruption area. Both features can increase

crash risk at merge segment. A lot of diverging traffic may move onto the lane nearest to deceleration lane in advance in the upstream of diverging area, since urban expressway carries a lot of commuters and many drivers are familiar with road structure. Hence, even if DR and L_D are found out as significant influencing factors, CR at diverge segment is not significantly higher than that at basic segment.

As traffic density increases, the effects of traffic-related variables increase and get more significant than geometry in congested flow. In this condition, once a breakdown initiates at bottlenecks, it can propagate to upstream section that may consists of several facility types, where traffic conditions are not significantly different. As a result, the difference of CR characteristics at basic, merge, and diverge segments is not significant. However, due to the heavily restricted visibility induced by the special geometric design, sharp curve still has higher CR than other facility types.

6 Conclusions and future work

This paper identified the different CR characteristics by facility type of Nagoya Urban Expressway. In uncongested flow, CR at basic, merge, and diverge segments appears convex downward to traffic density. In contrast, the value at sharp curve follows a decreasing tendency. In congested flow, CR at four facility types increases as traffic density increases. In both un- and congested flows, sharp curve has the worst safety performance in view of its highest CR among the four facility types. As for other segments, merge

Table 12 PCA results in congested flow

| Facility type (number of crash) | Item | Principal component | | | | | | | | | |
|---------------------------------|---------------------|---------------------|--------|----------|--------|----------|--------|-------|--------|-------|--------|
| | | 1st | | 2nd | | 3rd | | 4th | | 5th | |
| | | F | L | F | L | F | L | F | L | F | L |
| Basic segment (138) | Component | k | 0.950 | I_{CF} | 0.842 | h | 0.699 | Day | 0.942 | | |
| | | v | -0.947 | I_{SD} | 0.743 | Pave | 0.814 | | | | |
| | Percent of variance | 26.0 | | 20.3 | | 18.9 | | 15.4 | | | |
| Merge segment (91) | Component | MR | 0.871 | k | 0.883 | I_{CF} | 0.672 | L_A | -0.871 | Day | -0.933 |
| | | h | 0.790 | v | -0.884 | Light | 0.745 | Pave | 0.716 | | |
| | Percent of variance | 22.1 | | 18.3 | | 14.6 | | 13.9 | | 10.1 | |
| Diverge segment (48) | Component | k | 0.868 | I_{CF} | 0.849 | DR | 0.854 | Pave | 0.721 | Light | 0.951 |
| | | v | -0.801 | I_{SD} | 0.879 | L_D | -0.783 | Day | -0.716 | | |
| | Percent of variance | 21.9 | | 19.6 | | 15.4 | | 13.8 | | 11.6 | |
| Sharp curve (27) | Component | k | -0.772 | I_{CF} | 0.909 | Pave | 0.799 | Day | 0.912 | | |
| | | v | 0.950 | I_{SD} | 0.875 | 0.949 | | | | | |
| | Percent of variance | 30.5 | | 26.6 | | 22.5 | | 11.1 | | | |

segment has higher CR compared to the basic and diverge segments in uncongested flow. Comparatively, CR at three facility types is not significantly different in congested flow.

The causes of the differences were further investigated by focusing on traffic conditions and considering the interaction of geometry, traffic flow, and ambient conditions. Generally, geometric features are the most significant factors in uncongested flow. With the increase of traffic density, the effects of traffic-related variables increase and become most significant in congested flow. For ramp traffic, the significance of MR affecting crashes is on the rise as mainline flow increases. In contrast, the significance of DR gets decreasing. In addition, higher L_A and L_D are adverse to safety for uncongested flow, while smaller L_A and L_D are adverse for congested flow. Crashes at sharp curve are highly associated with the after effects of its special geometric design, such as large centrifugal force and heavy restricted visibility.

The potential benefits of integrating these findings in safer geometric design and traffic control are numerous. The analysis can provide a basis for geometric audit for safety regarding design consistency. Meanwhile, based on the estimated CR models, road administrators can easily image the safety performance with the variation of traffic conditions at a given facility type. Furthermore, PCA results may help prioritize countermeasures and further estimate the safety performance of an adopted countermeasure.

For more accurate analysis of crash characteristics, data in smaller time window, e.g., 1 min even 30 s, are highly recommended to improve the reliability of statistics, since crash occurrence is significantly associated with the short-term turbulence of traffic flow [1]. Furthermore, it is better to examine the effect of inter-lane interaction on crashes if the lane-based data is available. In this study, ramp traffic is found out to play a significant role for safety at merge and diverge segments. Thus, a microscopic analysis on driver behavior is needed. In essence, PCA is a qualitative analysis and the results are insufficient for applying specific countermeasures for a given case. Future studies are expected to acquire the quantitative effects of various influencing factors on crashes.

Acknowledgments The authors gratefully acknowledge the support of Nagoya Expressway Public Corporation for the data provision and other helps for this study.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References

1. Abdel-Aty M, Pande A (2007) Crash data analysis: collective versus individual crash level approach. *J Saf Res* 38(5):581–587
2. Wu Y, Nakamura H, Asano M (2012) A comparative study on crash rate characteristics at different intercity expressway facility

- types. In: Proceedings of the 46th IP meeting of Japan Society of Civil Engineering (JSCE), Saitama University, 2–4 Nov 2012
3. Golob TF, Recker WW (2004) A method for relating type of crash to traffic flow characteristics on urban freeways. *Transp Res Part A* 38(1):53–80
 4. Rengarasu TM, Hagiwara T, Hirasawa M (2009) Effects of road geometry and cross section variables on traffic accidents using homogeneous road segments. *Transp Res Rec* 2102:34–42
 5. Shively TS, Kockelman K, Daminen K (2010) A bayesian semi-parametric model to estimate relationships between crash counts and roadway characteristics. *Transp Res Part B* 44(5):699–715
 6. Bajwa S, Warita H, Kuwahara M (2010) Effects of road geometry, weather and traffic flow on safety. In: Proceedings of the 15th international conference of Hongkong Society for transportation study, The Hong Kong University of Science and Technology, Hong Kong, 11–14 Dec 2010
 7. Abdel-Aty M, Pemmanaboina R (2006) Calibrating a real time traffic crash prediction model using archived weather and ITS traffic data. *IEEE Trans Intell Transp Syst* 7(2):167–174
 8. Caliendo C, Guida M, Pasis A (2007) A crash-prediction model for multilane road. *Accid Anal Prev* 39(4):657–670
 9. Christoforou Z, Cohen S, Karlaftis MG (2011) Identifying crash types propensity using real-time traffic data on freeways. *J Saf Res* 42:43–50
 10. Hossain M, Muromachi YA (2012) A bayesian network based framework for real-time crash prediction on basic freeway segments of urban expressways. *Accid Anal Prev* 45:373–381
 11. Wu Y, Nakamura H, Asano M (2013) A comparative study on crash characteristics between urban and intercity expressway basic segments. In: Selected proceedings of the 13th world conference on transport research, 15 pp
 12. Hikosaka T, Nakamura H (2001) Statistical analysis on relationship between crash rate and traffic flow condition in basic expressway sections. In: Proceedings of the 21st Japan Society of Traffic Engineering Meeting, Tokyo, 29–30 Oc 2001. (in Japanese)
 13. Ng JC, Sayed T (2004) Effects of geometric design consistency on road safety. *Can J Civ Eng* 31:218–227
 14. Shawky M, Nakamura H (2007) Characteristics of breakdown phenomena in Japan urban expressway merging sections. In: Proceedings of TRB 86th annual meeting, Washington DC, 21–25 Jan 2007
 15. Orme JG, Combs-Orme T (2009) Multiple regression with discrete dependent variables. Oxford University Press, New York
 16. Zhou HG, Chen HY, Zhao JG et al (2010) Operational and safety performance of left-sides off-ramps at freeway diverge areas. In: Proceedings of TRB 89th annual meeting, Washington DC, 10–14 Jan 2010