Accident Analysis and Prevention 133 (2019) 105290

Contents lists available at ScienceDirect

Accident Analysis and Prevention

journal homepage: www.elsevier.com/locate/aap

Risk analysis and management for highway operations safety using a covariate-balanced determinant detector

Chengqian Li, Qi Fang, Lieyun Ding*, Botao Zhong

Department of Construction Management, School of Civil Engineering & Mechanics, Huazhong University of Science & Technology, Wuhan, Hubei, China

ARTICLE INFO

Keywords:
Highway safety
Risk management
Covariate-balanced determinant detector
Interaction effect
Mitigation strategies

ABSTRACT

Highway operations are marred with inherent risks of injury or death, making risk management critical for ensuring the adequate safety of the people involved. This paper investigates the interaction between various highway safety risk factors and effective risk mitigation strategies related to such interaction. The Covariate-Balanced Determinant Detector (CBDD) technique is used to estimate the quantity of both individual and combined risks, and their effect on highway operations safety. Through this technique, the most dangerous risk combinations have been identified and corresponding risk mitigation scenarios have been developed. The results illustrate that the most dangerous scenarios probably result from the interactive effect of risk factors rather than individual factors, and the effect of mitigation strategies should be evaluated in response to a risk scenario before it is implemented.

1. Introduction

Highway traffic accidents have become a major concern for the highway construction and operation administration. Traffic accidents are consistently one of the top three causes of death for people aged 5–44 years old (Qu et al., 2014), imposing a huge economic loss to countries (Zhang et al., 2016).

There is still much room for safety improvement on existing highways, even though design specifications stipulate the desirable range of road geometry parameters to ensure operation safety (Li et al., 2019). However, meeting the requirements of the codes does not mean that the risk of traffic accidents has been removed entirely, especially in precipitous terrain areas. Severe accidents frequently occur on highway segments with constrained road geometry, adverse weather conditions, and poor visual distance (Ahmed et al., 2011), which implies an urgent need for improvements in highway operations.

To improve highway operations safety, risk analysis and mitigation evaluation plays a very important role in providing theoretical support. This paper aims to follow the risk analysis framework to first identify the leading risk scenarios and then investigate the effect of mitigation strategies related to such risk scenarios. The application value of this paper is twofold: firstly, to help engineers designing a similar highway project to avoid significant risk scenarios in the design and to adopt proper mitigation measures to improve the operational safety of the project; secondly, to assist highway operation managers in identifying accident hot spots and formulating decisions on effective mitigation strategies.

To solve the above problems, this paper proposes a new Covariate-Balanced Determinant Detector (CBDD) technique to identify the leading risk scenarios of the highway design as well as the effective risk mitigation strategies for risk mitigation. The CBDD technique is a causal-effect analysis method used to explore the association between different variables, and one of its greatest strengths is that it has little model assumption and does not need to specify an explicit formula between the dependent and independent variables. It can bring great convenience to the data analysts if the function relations between the dependent and independent variables is very unsure or hard to determine.

This work advances research on highway safety analysis in two ways: first, a CBDD technique, which improves Geographical Detector by balancing the covariates vector to eliminate the bias (detailed in Section 3.1), is proposed in this paper as a new causal-effect analysis tool in the toolbox for analysts; second, the proposed method is applied to a high-risk highway section of China in precipitous terrain and adverse weather environment to help engineers identify the most risky scenario and the corresponding effective engineering countermeasures.

2. Risk analysis and techniques

Risk is generally defined as the probability of an accident adversely

* Corresponding author at: Huazhong University of Science & Technology, Wuhan, Hubei, China.
E-mail address: dly@hust.edu.cn (L. Ding).

https://doi.org/10.1016/j.aap.2019.105290
Received 23 April 2019; Received in revised form 24 July 2019; Accepted 2 September 2019
Available online 01 October 2019
0001-4575/

or unfavorably affecting project objectives due to uncertainty (Al-Bahar and Crandall, 1990). The effective systematic framework to manage risk usually follows three processes (Al-Bahar and Crandall, 1990; Degn Eskesen et al., 2004): risk identification, risk assessment and evaluation, and risk mitigation and control. Therefore, exploring the causal relationship between the accidents and the risk factors, and estimating the effect of risk mitigation strategies on road safety, are the core of risk management in highway safety.

There are many methods available to conduct risk analysis but the most commonly used methods can be summarized into the following four techniques: Fault Tree Analysis (FTA) (Chi et al., 2014), Bayesian Networks (BNs) (Wu et al., 2015), the Regression Model (Chen et al., 2017), and the Geographical Detector Model (Wang et al., 2010).

Both FTA and BN are the ‘decomposition tools’ for accidents, since they usually construct a tree-like network (Friedman et al., 1997) for an accident event. In these techniques, the root node (or top event in FTA) is the accident itself; leaf nodes (or basic events in FTA) describe the root causes for the accident and the intermediate nodes are used to describe how the root causes lead to an accident. FTA and BN are widely used for risk analysis of engineering accidents, especially in accident chain analysis. However, the chain reaction of basic events (i.e., the network structure) needs to be specified manually, which greatly depends on the accuracy of expert judgment. The causal relationship deduction may be misleading due to model misspecification.

The Regression Model (Shankar et al., 1995) primarily consists of three main steps: the process starts with including all potential explanatory variables into the regression model, then examining the variables’ impact on the crash rate with a marginal effects analysis or statistical significance test (Chen et al., 2019) and, finally, estimating the average treatment effect using a before-after analysis (Buddhavarapu et al., 2015). However, the regression method sometimes cost the analysts a lot of energy and time in determining the proper formula relations between the explanatory and response variables. It may be a heavy task to find a correct model to satisfy the model assumption for the researchers (Wang et al., 2016a). On the other hand, when applying the regression model, all possible risk factors are assumed to be independent, which may not exist in some empirical cases. This hypothesis may lead to a selection of only a few independent variables according to the expert knowledge or a supplementary method to reduce the number of variables, such as Principal Component Analysis (Luo, 2017).

The Geographical Detector Model (GDM) is a powerful method proposed by Wang et al. (2010) and it has been widely used to identify the key drivers of various kinds of geographical phenomenon in the fields of energy, sociology, medical science and ecology. The main advantages of GDM lie in the fewer model assumptions it makes than when compared with other methods and its capacity to reveal the possible interactive relationship between different potential factors (Wang and Xu, 2017).

However, there are still limitations to the original GDM method. The most important criticism of GDM is that it neglects the covariate information when calculating the contributing impacts of driving factors. The method does not consider the value of covariates between different units, which can cause bias when estimating the variables’ causal impacts. Covariate balance (Brooks and Ohsfeldt, 2013; Hansen and Bowers, 2008) is an important concern when estimating the treatment effect of a policy or risk mitigation strategies; otherwise, the computing results may be biased and unreliable if there are unmeasured confounders (Imai and Ratkovic, 2014) (The covariate balance problem will be detailed in Section 3.1).

Based on this, we proposed a new Covariate-Balanced Determinant Detector (CBDD) technique to identify the leading risk factors and effective risk mitigation strategies for highway safety based on the accident data. The CBDD has improved GDM by constructing a weight for the Sum of Squares (SS), such that the observed covariates are balanced between different SS’. Moreover, a risk analysis framework is also presented based on CBDD, which follows the risk management process: risk identification, evaluation and mitigation.

3. Methodology

This paper investigates the most dangerous risk scenarios for highway safety improvement by quantifying the individual and interactive effects of risk factors, and then identifies the effective strategies in response to the above risk scenarios.

Given an input data set for risk analysis with N records $\{\boldsymbol{X_n^r}, \boldsymbol{X_n^m}, Y_n\}_{n=1}^N$, the CBDD technique is proposed to detect the significant dominants among p_r risk factors $\boldsymbol{X^r} = (\boldsymbol{x_1^r}, \boldsymbol{x_2^r}, ..., \boldsymbol{x_{p_r}^r})$ and effective strategies among p_m mitigation measures $\boldsymbol{X^m} = (\boldsymbol{x_1^m}, \boldsymbol{x_2^m}, ..., \boldsymbol{x_{p_r}^m})$ that may influence the accident rate y. The CBDD technique is mainly composed of three components: an individual effect detector, an interactive effect detector, and a mitigation effect estimator.

In the individual effect detector, the impacts of all the potential factors on the dependent variable are computed respectively, and the leading factors are identified according to the relative magnitude of their impacts. The factors detected with a high impact value are recognized as leading factors, while those with a low impact value are not.

In the interactive effect detector, the interaction effects of different combinations of risk factors are quantified. Some factors may show strongly enhanced impacts through interaction with each other, which was usually neglected in the previous research.

In the mitigation effect estimator, various kinds of risk scenarios are recognized and the effects of the mitigation strategies are measured, which will help support decisions made by the operations managers.

The overall framework of the risk analysis based on the CBDD technique, is shown in Fig. 1. First, the potential risk factors are identified according to expert knowledge. Second, the individual effect detector and interactive effect detector are employed to estimate the risk factor and risk scenario effects. The risk factors (or risk combinations) detected with the highest impact on the accident occurrence are identified as leading risk scenarios. Third, the effect of the risk mitigation strategies are estimated. The risk mitigation strategies that significantly reduce the accident rate in a risk scenario are regarded as the effective mitigation strategies.

3.1. Individual effect detector

Here, we suppose all risk factors $\boldsymbol{X^r} = (x_1^r, x_2^r, ..., x_{p_r}^r)$ are categorical variables instead of continuous variables (some discretization methods (Cao et al., 2013) could be applied to transfer the explanatory factors into categorical variables if they are continuous variables by nature), and the dependent variable y is a continuous variable. Based on this assumption, we can further calculate three important indicators for a factor: x_u^r, i.e., $SSW(x_u^r)$, $SSB(x_u^r)$ and $SST(x_u^r)$. The individual effect of x_u^r is computed according to these three indicators.

x_u^r is discretized into $h = 1,2, ..., \mathrm{L}$ levels, and h_{th} level is composed of N_h records. We divide the dependent variables (accident rate) y into L zones according to discrete values of x_u^r, $\boldsymbol{y} = \{\boldsymbol{y_{h_1}}, \boldsymbol{y_{h_2}}, ..., \boldsymbol{y_{h_L}}\}$. We define the sum of squares within the same level as $SSW(x_u^r)$,

$$SSW(x_u^r) = \sum_{h=1}^{L} \sum_{i}^{N_h} (y_{hi} - \bar{y}_h)^2 \tag{1}$$

the sum of squares between different levels as $SSB(x_u^r)$,

$$SSB(x_u^r) = \sum_{h=1}^{L} \sum_{i}^{N_h} (\bar{y}_h - \bar{y})^2 \tag{2}$$

and the total sum of squares $SST(x_u^r)$,

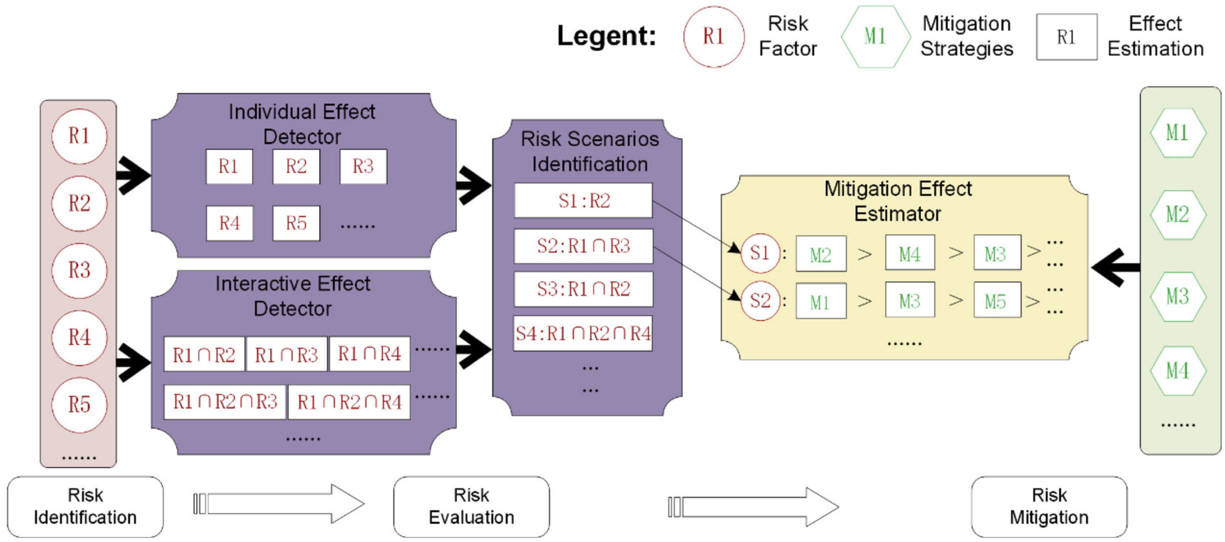


Fig. 1. Framework of the risk analysis based on the CBDD technique.

$$SST(x_u^r) = \sum_i^N (y_i - \bar{y})^2 \tag{3}$$

where $y_{h,i}$ and y_i denote the value of record i in the level h and in the population, respectively; $\bar{y}_h = (1/N_h)\sum_1^{N_h} y_{h,i}$ and $\bar{y} = (1/N)\sum_1^N y_i$ are the level mean and population mean, respectively.

In previous research (Wang et al., 2016b), it has been shown that there is a relation between these indicators, $SST(x_u^r) = SSW(x_u^r) + SSB(x_u^r)$. Furthermore, the individual effect of x_u^r can be computed by the Geographical Detector method (Hu et al., 2011; Huang et al., 2014; Wang et al., 2010) according to the following equation:

$$q(x_u^r) = 1 - SSW(x_u^r)/SST(x_u^r) \tag{4}$$

where $q(x_u^r) \in [0,1]$ is the estimated impact of x_u^r, which is an effective indicator showing the explanatory power of the influencing factor. If the value of $q(x_u^r)$ equals 1, it means that x_u^r totally explains the variation of the accident rate. If the value equals 0, it means that there is no relationship between x_u^r and the accident rate.

To better understand the q -statistics, we introduce the definition of variogram(Barnes, 1991) into Eqs. (1) and (3):

$$SSW(x_u^r) = \sum_{h=1}^{L} \sum_{i=1}^{N_h} (y_{hi} - \bar{y}_h)^2 = \sum_{h=1}^{L} \left[\frac{\sum_{i=1}^{N_h} \sum_{j=1}^{N_h} (y_{hi} - y_{hj})^2}{2N_h} \right] \tag{5}$$

$$SST(x_u^r) = \sum_{i=1}^{N} (y_i - \bar{y})^2 = \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} (y_i - y_j)^2}{2N} \tag{6}$$

Therefore, $SSW(x_u^r)$ refers to the variation of accident rates within the same level of x_u^r, and $SST(x_u^r)$ refers to the variation of the accident rate in a population.

Inspired by Cang's idea (Cang and Luo, 2018), the weighted means of Eqs. (5) and (6) can be calculated in Eqs. (7) and (8)

$$WSSW(x_u^r) = \sum_{h=1}^{L} \left[\frac{1}{2N_h} \frac{\sum_{i=1}^{N_h} \sum_{j=1}^{N_h} w_{hi,hj} (y_{hi} - y_{hj})^2}{\left(\sum_{i=1}^{N_h} \sum_{j=1}^{N_h} w_{hi,hj}\right) / \left(\sum_{i=1}^{N_h} \sum_{j=1}^{N_h} 1\right)} \right]$$

$$= \sum_{h=1}^{L} \left[\frac{N_h}{2} \frac{\sum_{i=1}^{N_h} \sum_{j=1}^{N_h} w_{hi,hj} (y_{hi} - y_{hj})^2}{\sum_{i=1}^{N_h} \sum_{j=1}^{N_h} w_{hi,hj}} \right] \tag{7}$$

$$WSST(x_u^r) = \frac{N}{2} \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} w_{i,j} (y_i - y_j)^2}{\sum_{i=1}^{N} \sum_{j=1}^{N} w_{i,j}} \tag{8}$$

where $w_{i,j}$ is the weight of the squared difference between y_i and y_j, and $w_{hi,hj}$ refers to the special $w_{i,j}$ if records i and j belong to the same level h. It's noted that the weighted means in Eqs. (7) and (8) are equal to the variance in Eqs. (5) and (6) respectively, if $w_{i,j} = 1, \forall$ i, j$\in N$.

One of the major limitations of the Geographical Detector method is that it neglects the covariate balance problem when estimating the individual effect of an influencing factor.

For example, when estimating the impact of x_1^r in Table 1, it is difficult to determine whether the difference in $(y_1 - y_2)^2$ is caused by x_1^r, since x_1^r and x_2^r vary at the same time between records 1 and 2. In contrast, the difference in $(y_1 - y_3)^2$ is more probably caused by x_1^r, since the covariate, x_2^r, is the same between records 1 and 3. There should be a higher weighting for $(y_1 - y_3)^2$ than $(y_1 - y_2)^2$, since the former better explains that the change of accident rate is caused by target factors instead of the covariates.

Assuming that the covariate vectors for $x_{u,i}^r$ are represented as $\ddot{\boldsymbol{x}}_{\boldsymbol{u},i} = \{x_{v,i}^r\}, \forall\ v \neq u$. We define $w_{i,j}$ as a similarity measure for the covariates vectors between items i and j, i.e., $\ddot{\boldsymbol{x}}_{\boldsymbol{u},\boldsymbol{i}}$ and $\ddot{\boldsymbol{x}}_{\boldsymbol{u},\boldsymbol{j}}$. $w_{i,j}$ can be computed by the Pareto distributions (Newman, 2005) in Eq. (9), which means that the explanatory power is concentrated on a small number of record comparisons in which the covariates are well balanced;

$$w_{i,j} = \frac{k}{(d_{i,j} + 1)^{k+1}} \tag{9}$$

where k is a pre-defined parameter showing the model sensitivity to covariate balancing, and $d_{i,j}$ is a Mahalanobis distance between the covariates vectors of item i and j.

$$d_{ij} = \sqrt{(\ddot{\boldsymbol{x}}_{\boldsymbol{u},\boldsymbol{i}} - \ddot{\boldsymbol{x}}_{\boldsymbol{u},\boldsymbol{j}})^T W (\ddot{\boldsymbol{x}}_{\boldsymbol{u},\boldsymbol{i}} - \ddot{\boldsymbol{x}}_{\boldsymbol{u},\boldsymbol{j}})}$$

Table 1
An illustrative example of the covariate balance problem.

Record ID i	Dependent variable y_i	Risk factors	
		Target factor $x_{1,i}^r$	Covariate $x_{2,i}^r$
1	0	0	0
2	1	1	1
3	1	1	0

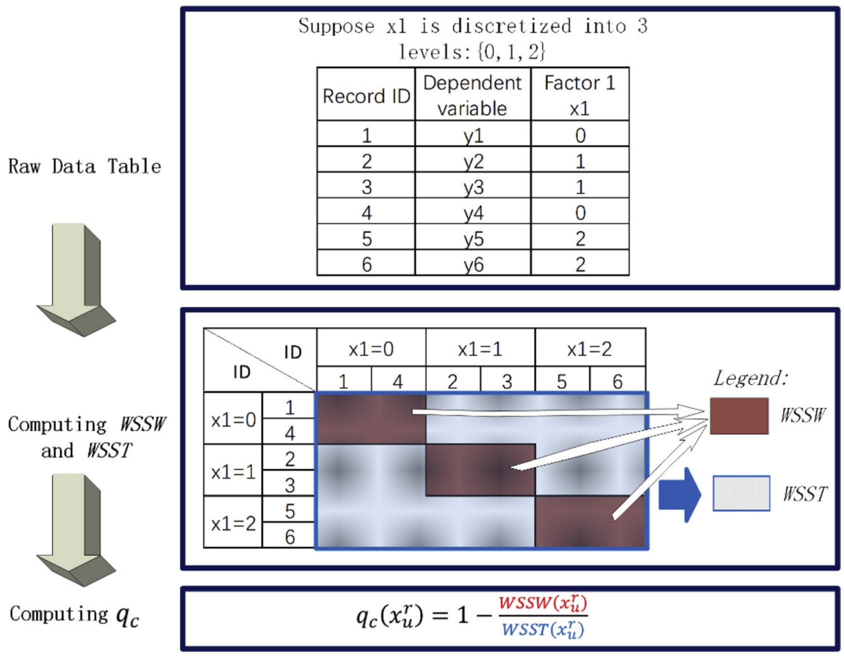


Fig. 2. The calculating process of the individual effect detector.

$$= \sqrt{\sum_{l=1,v\neq u}^{L-1} \frac{(x_{v,i}^r - x_{v,j}^r)^2}{s_v{}^2}} \tag{10}$$

W is the covariance matrix, s_v is the standard deviation of $\boldsymbol{x}_v^r$, $\boldsymbol{x}_v^r = \{x_{v,1}^r, x_{v,2}^r, ..., x_{v,N}^r\}$.

$w_{i,j}$ is a strictly monotonically decreasing positive function of d_{ij}. In extreme cases, $w_{i,j} = k$ if $d_{ij}= 0$, and $w_{i,j} = 0$ if $d_{ij} \to +\infty$.

Given the weighted means of SSW and SST, a new covariate balancing estimator q_c, to predict the effect of an influencing risk factor x_u^r, can be calculated as follows:

$$q_c(x_u^r) = 1 - WSSW(x_u^r)/WSST(x_u^r) \tag{11}$$

where $q_c(x_u^r)$ identifies the degree of how much a potential risk factor x_u^r accounts for the changes in a highway accident rate, the value range of q_c is [0,1], and its value indicates that x_u^r explains the $(100 \times q_c)\%$ change of accident rate. In extreme cases, $q_c(x_u^r) = 0$ means that there is no relationship between x_u^r and the accident rate, and $q_c(x_u^r) = 1$ means that x_u^r completely explains the change in accident rate. The complete calculation process of the individual effect detector is presented in Fig. 2.

3.2. Interactive effect detector

An interactive effect detector identifies the interaction of different explanatory factors on the road accident rate by comparing the sum contribution of two individual factors with the contribution of these two factors when taken together. Different types of interactions are shown in Table 2 according to the comparison results. The symbol '∩' denotes the interaction between these two factors. The calculation of the interactive effect follows the same procedure of individual effect, the only difference is that $(x_u^r \cap x_v^r)$ is discretized into more levels. Suppose there are L_u and L_v levels for x_u^r and x_v^r respectively, the interaction of $(x_u^r \cap x_v^r)$ can be discretized into $L_u \times L_v$ levels. A measure of the interactive effect for x_u^r and x_v^r is defined by:

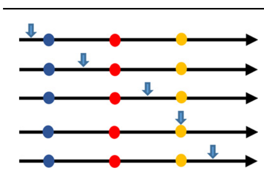
$$\mathrm{I}_q(x_u^r \cap x_v^r) = q_c(x_u^r \cap x_v^r) - [q_c(x_u^r) + q_c(x_v^r)] \tag{12}$$

if $\mathrm{I}_q = 0$, x_u^r and x_v^r are independent from each other. Otherwise, it indicates an interactive effect on the factors combination. The greater the deviation of $\mathrm{I}_q(x_u^r \cap x_v^r)$ from 0, the more significant the interactive effect.

3.3. Mitigation effect estimator

The mitigation effect of risk mitigation strategies on a road accident can be estimated by the following framework using the CBDD technique, as shown in Fig. 3. First, all the explanatory factors are divided into two categories, i.e., Risk Factors and Risk Mitigation Strategies. The former refers to the operational environment that is unavoidable or difficult to change, e.g., weather or tunnel environment. The latter refers to the risk mitigation strategies that are easily employed in practice, such as the improvement of pavement materials and road barriers. Then, steps proceed to identify Leading Risk Factors (LRF), Leading Risk Combinations (LRC), Risky but Overlooked Combinations (ROC) and Effective Mitigation Strategies (EMS), respectively. The main objective of these steps is to identify which level of the explanatory variables has the highest (or lowest) mean accident rate compared to others and if it is statistically significant.

Table 2
Interaction types of two risk factors (Jiang et al., 2018; Zhao et al., 2017).

Graphical representation	State	Interaction type
	$q_c(x_u^r \cap x_v^r) < \min(q_c(x_u^r), q_c(x_v^r))$	Weaken, nonlinear
	$\min(q_c(x_u^r), q_c(x_v^r)) < q_c(x_u^r \cap x_v^r) < \max(q_c(x_u^r), q_c(x_v^r))$	Weaken, univariate, nonlinear
	$q_c(x_u^r \cap x_v^r) > \max(q_c(x_u^r), q_c(x_v^r))$	Weaken, linear, bivariate
	$q_c(x_u^r \cap x_v^r) = q_c(x_u^r) + q_c(x_v^r)$	Independent
	$q_c(x_u^r \cap x_v^r) > q_c(x_u^r) + q_c(x_v^r)$	Enhance

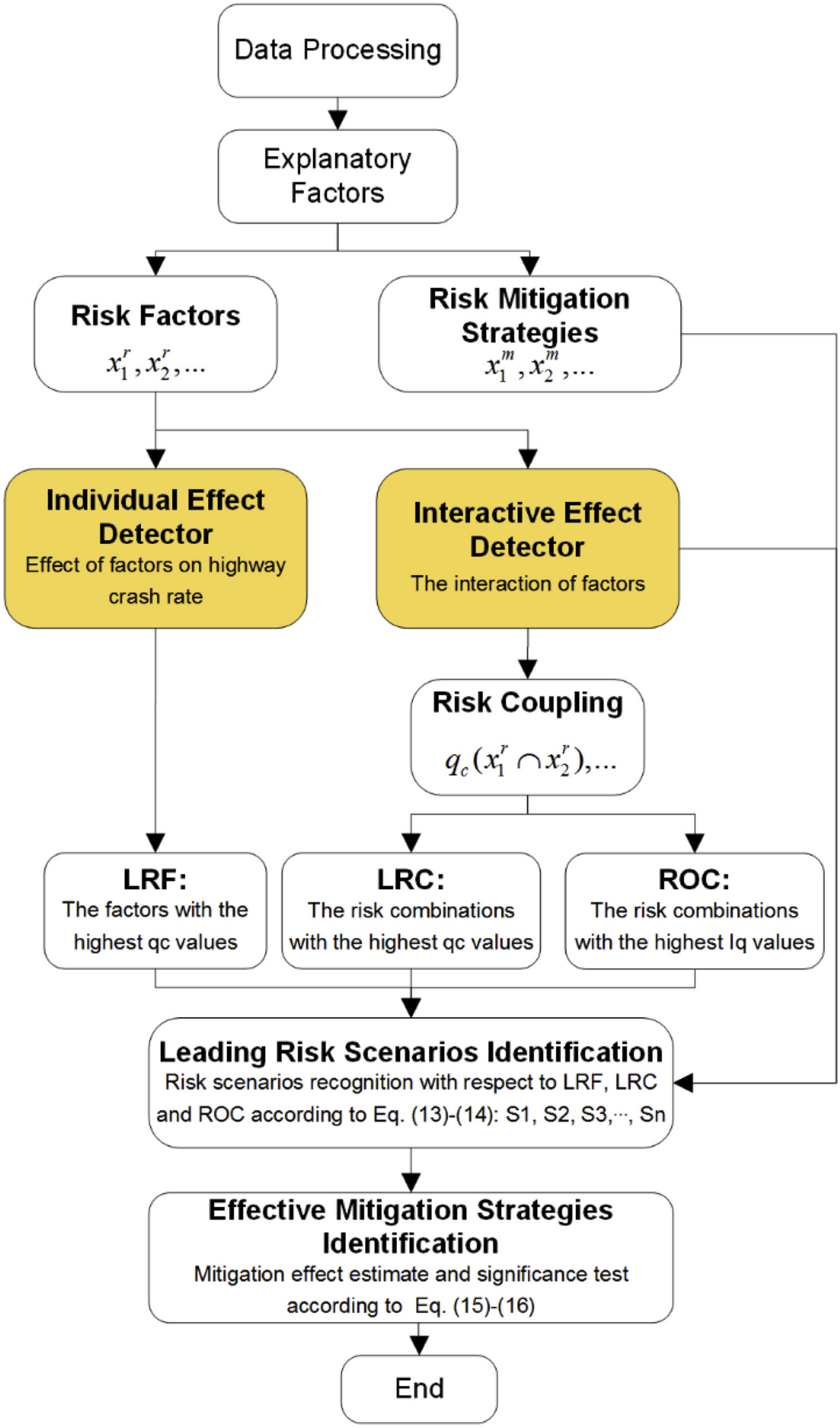


Fig. 3. Technical flow for the mitigation effect estimator.

Step 1: The factors with the highest q_c values can be recognized as the Leading Risk Factors (LRF). Given a risk factor x^r as LRF, a risk scenario S, with respect to LRF, can be defined as the leading impact type of x^r on an accident rate. Suppose x^r is discretized into $h = 1, 2, \ldots, L$ levels, and the h_{th} level is composed of N_h records, the estimated accident rate R_h for the h_{th} level (Luo, 2017) can be computed by:

$$R_h = \frac{\sum_{i=1}^{N_h} y_{hi}}{N_h} \tag{13}$$

given the R_h of all the discretization levels, the leading impact type is defined as the level of x^r with maximized R_h.

$$S = \underset{h}{\operatorname{argmax}}(R_h) \tag{14}$$

Step 2: To identify the Leading Risk Combinations (LRC), the individual effect and interactive effect detectors are employed to compute the q_c values for all the individual and combinations of RF, and those with the highest q_c values can be recognized as the LRC. The risk scenario S with respect to LRC is computed by Eqs. (13) and (14).

Step 3: In particular, the risk factor combinations with the highest I_q values are identified as the Risky but Overlooked Combinations (ROC), since the interactive effects are often hard to observe in practice. The identification of ROC is important for risk management because its potential interaction effect could lead to unexpected accidents. The risk scenario S with respect to ROC is computed by Eqs. (13) and (14).

Step 4: Given a risk scenario$S$ in steps 1–3, the effect of a mitigation strategy x^m is computed. Suppose $h = 1, 2, \ldots, L$ levels for a mitigation strategy x^m and $N_{h,S}$ records for the h_{th} level in scenario S, the mitigation effect from α_{th} level to β_{th} level of x^m can be quantified as:

$$ME_S^{\alpha,\beta} = \frac{\sum_{i=1}^{N_{\alpha,S}} \sum_{j=1}^{N_{\beta,S}} w_{i,j}(y_{i,S} - y_{j,S})}{W_{ij}} \tag{15}$$

where $W_{ij} = \sum_{i=1}^{N_{\alpha,S}} \sum_{j=1}^{N_{\beta,S}} w_{i,j}$, $\alpha, \beta \in \{1,2,\ldots,L\}$ and $y_{i,S}$ refers to the value of y in the i_{th} record for a Scenario S. For example, given a scenario S: $\{x_1^r = 2\}$ and a mitigation strategy x_1^m, there will be $N_{\alpha,S}$ records in which $x_1^r = 2$ and $x_1^m = \alpha$. The calculation of $w_{i,j}$ is the same as Eq. (9).

It is noted that $ME_S^{\alpha,\beta}$ is different from the average treatment effect (ATE) (Li and Graham, 2016) in two ways: first, $ME_S^{\alpha,\beta}$ measures the safety impact of EMS with respect to a certain risk scenario, while ATE estimates an average impact of EMS for all the records; second, there is fewer model assumption when computing the $ME_S^{\alpha,\beta}$ value.

Step 5: To compare the difference in the accident rate from α_{th} level to β_{th} level, the *t*-test (Luo, 2017) is employed to test whether the weighted mean difference, i.e., $\sum_{i=1}^{N_{\alpha,S}} \sum_{j=1}^{N_{\beta,S}} w_{i,j}(y_{i,S} - y_{j,S})/W_{ij}$, is significantly different from 0. Let $\{a_{ij}\}$ and $\{b_{ij}\}$ be two arrays with $N_{i,S}$ rows and $N_{j,S}$ columns, where $a_{ij} = \left\{\frac{w_{i,j}y_{i,S}}{W_{ij}}\right\}$ and $b_{ij} = \left\{\frac{w_{i,j}y_{j,S}}{W_{ij}}\right\}$. The computation formula is as follows:

$$t_{ij} = \frac{\bar{a_{ij}} - \bar{b_{ij}}}{\sqrt{\frac{\sigma_a + \sigma_b}{N_{i,S}N_{j,S}}}} \tag{16}$$

where t_{ij} is the test value of t; $\bar{a_{ij}}$ and $\bar{b_{ij}}$ are the mean of $\{a_{ij}\}$ and $\{b_{ij}\}$, respectively; σ_a and σ_b are the variance of $\{a_{ij}\}$ and $\{b_{ij}\}$, respectively.

4. Case study

The proposed method was applied to the Xishang Highway, an important component of the Shanxi Highway System in China. The leading risk scenarios and effective risk mitigation strategies identified by our method would provide valuable information to improve the safety level of highway design and operation.

4.1. Preprocessing of accident data

To quantify the causal effect of various kinds of risk factors on the highway crash rate, detailed information of crash data and driving environmental information should be collected. The accident records of the Xishang Highway from the Shanxi Traffic Management Bureau were used in this study. Xishang Highway links two cities from Xi'an to Shangluo, with a total length of 118.975 km. Between 2013 and 2016, a total of 3896 traffic accidents were recorded on the highway and 12 of them were severe accidents (fatal or severely injured). Considering that the sample of severe accidents is too small to be analyzed independently, we've converted the fatal accidents into non-fatal accidents according to the equivalent accident factors (Washington et al., 2014) estimated by experts (1:60), which results in a total of 4616 equivalent non-severe traffic accidents. The aerial photographs and GIS map were obtained to extract the highway information, including the highway's geometric and roadside features.

Considering that traffic accidents are distributed in different locations along the highway, the accident data needs to be organized into distinct homogeneous segments. The purpose of highway segmentation is to identify the accident hot spots of highways, which lays the basis for risk analysis and mitigation.

The current segmentation methods can be divided into two categories, the static segmentation method and the dynamic segmentation method. The static segmentation method (Cafiso et al., 2018) splits the highways into segments with a consistent length, while the dynamic

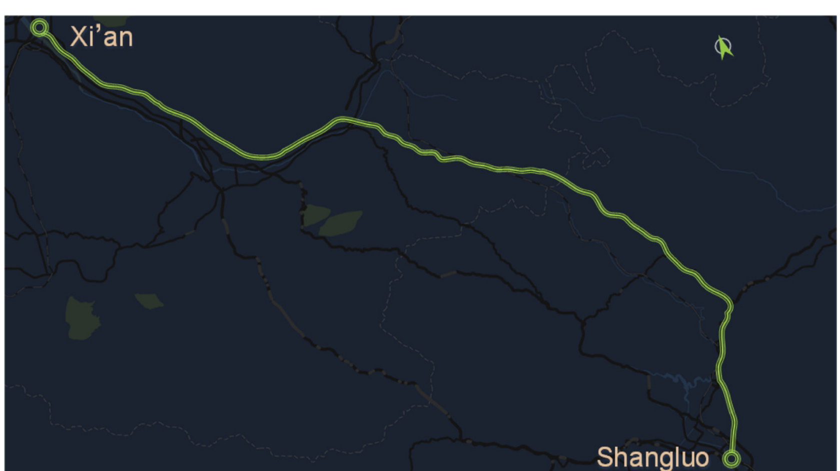


Fig. 4. The beginning and end of the Xishang Highway.

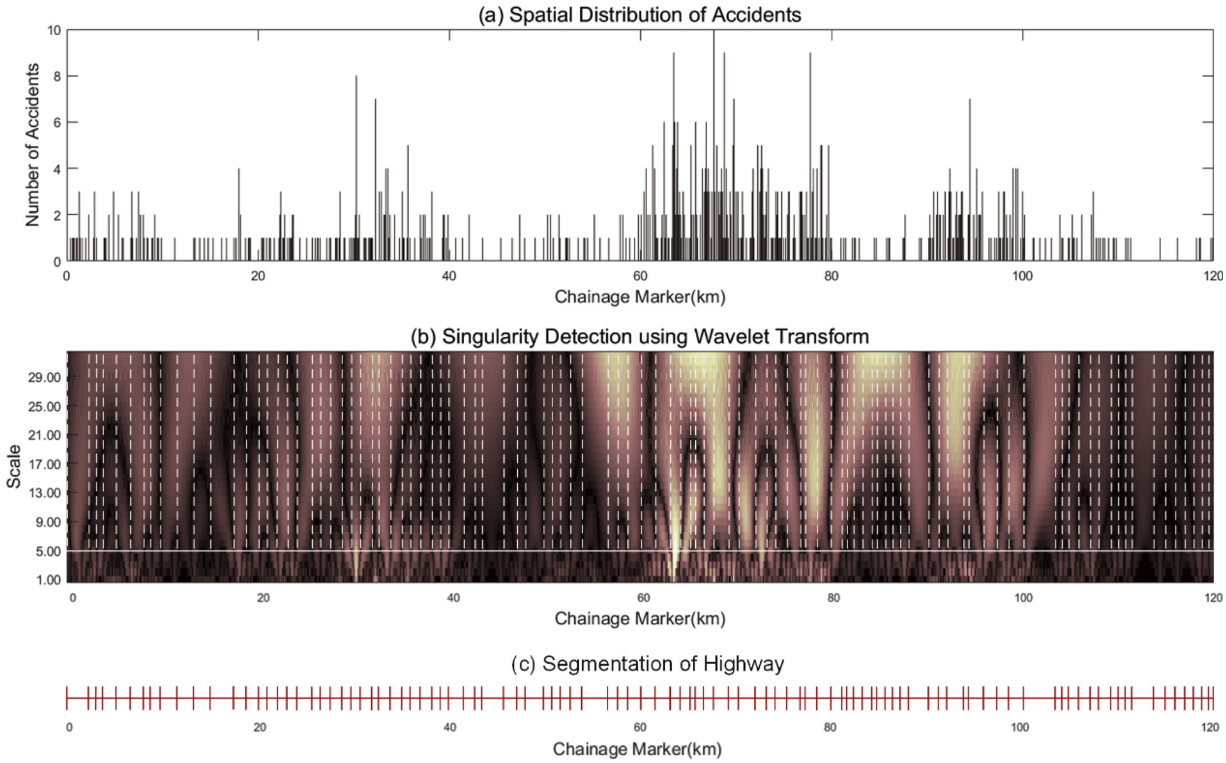


Fig. 5. Segmentation of the Xishang Highway based on the accident data from 2016.

Table 3
Data description.

Variable name	Description	Min.	Max.	Mean	Std. Dev.
Dependent variables					
Accident rate of 2013	Number of crashes	0	126	10.31	7.33
Accident rate of 2014	per km, per segment	0	68	9.89	6.56
Accident rate of 2015		0	59	8.76	6.14
Accident rate of 2016		0	42	8.89	5.27
Independent variables					
Roadway Related Risk Factors					
LSS	Large steep slope indicator (LSS): if the segment is located on a large steep slope defined by CCCC First Highway Consultants Co. (2017), LSS = 1; otherwise, LSS = 0	0	1	0.40	0.49
HCR	Horizontal curve radius indicator (HCR): < 2000m = 2, 2000m˜3000m = 1, > 3000m = 0	0	2	0.37	0.59
RC	Roadside context indicator (RC): open roadways = 0, tunnel inside = 1, tunnel entrance or exit = 2	0	2	0.40	0.72
G	Vertical Gradient (G): > 4% = 2, 2%˜4% = 1, < 2% = 0	0	2	0.63	0.74
Weather Related Risk Factors					
WC	Weather condition-affected (WC) segment: agglomerate fog = 2, black ice = 1, other = 0	0	2	0.34	0.66
Traffic Related Risk Factors					
P_HV	Proportion of heavy vehicles (P_HV): > 45% = 2, 35%˜45% = 1, < 35% = 0	0	2	1.21	0.67
AADT	Annual Average Daily Traffic: > 30,000veh/d = 1, others = 0	0	1	0.46	0.50
SL	Speed limit (SL): 100 km/h = 1, others = 0	0	1	0.74	0.44
Mitigation Strategies					
GB	Guardrail Barrier (GB): Jersey barrier = 1, Two-wave shape guardrail = 0	0	1	0.26	0.44
PM	Pavement Material (PM): Asphalt concrete = 1, Cement concrete = 0	0	1	0.32	0.47
RS	Rumble Strip (RS): Presence of rumble strip = 1, no speed reducing measure = 0	0	1	0.28	0.45
RTR	Runaway truck ramp (RTR): Presence of runaway truck ramp = 1, other = 0	0	1	0.41	0.40

segmentation method (Boroujerdian et al., 2014) splits the highway according to widespread accidents or density. It is obvious that using the static segmentation method may fail to identify the hot spots on the highway (as shown in Fig. 4 of Boroujerdian et al. (2014)), while the dynamic segmentation method adapts the segment length to the accident density. Therefore, this paper uses the wavelet transform technique (Cuhadar et al., 2002), one of the dynamic segmentation methods, to split the highway into homogeneous units. The algorithm of segmentation has been detailed in (Boroujerdian et al., 2014), which will not be covered here. There are 112, 121, 103 and 97 segments for 2013–2016, respectively. The segmentation based on the accident data from 2016 is shown in Fig. 5 (small-scale segmentation, e.g., scale = 5, is chosen in order to get more segments).

The aim of data preprocessing is to distribute the accident data into segments where the accidents occurred. The data on traffic accidents and explanatory variables for each segment of the highway in a year are aggregated into one record, denoted as $\{\boldsymbol{R_{i,j}}, \boldsymbol{M_{i,j}}, Y_{i,j}\}$, where $Y_{i,j}$ denotes the accident rate (per km), $\boldsymbol{R_{i,j}}$ denotes the risk factor vector and $\boldsymbol{M_{i,j}}$ denotes the risk mitigation strategies vector for the i_{th} segment and j_{th} year, $j = \{2013, 2014, 2015, 2016\}$. Considering there are 433 segments for 4 years, the number of records after preprocessing is also 433.

According to the definition of SafetyCube, a European Commission with the objective of developing an innovative road safety Decision Support System (DSS) (SafetyCube, 2016), 'risk factor' refers to any factor that contributes to the occurrence or the consequences of road crashes. Generally, risk factors associated with crash accidents can be classified into four categories (Zhang et al., 2013): human related factors (drivers' gender, age, behavior, etc.), traffic related factors (vehicles' types, density, speed, etc.), weather related factors, roadway related factors (physical parameters of roadway geometry). In principle, all these factors can have significant influence on the accident rate. However, Xishang Highway is a special mountainous highway in which precipitous terrain and adverse weather conditions lead to unfavorable geometric design and always cause a much higher frequency and severity of traffic accidents than on urban highways. Considering the specific characteristic of Xishang Highway, a research project was launched (listed in the acknowledgement of this paper) to investigate the specific parameters of Xishang Highway that worsen the driving environment and cause crash accidents. With this aim, we defined risk factors in this paper as the driving environment parameters that have a direct influence on the risk of a crash occurring, or on the crash severity. A number of risk factors associated with the driving environment were explored in this study, which are described in Table 3. It is noted that the definition and identification of risk factors are based on the judgement and discussion from many experts with rich engineering practices experience and domain knowledge.

It is noted that the risk factors and mitigation strategies in Table 3 were identified by the experts in highway operations management, and the CBDD technique was used to further determine the effect of ranking of risk factors, combinations and mitigation strategies.

All the risk factors are developed as discrete variables, since CBDD can only deal with discrete variables. Explanations of the discretization for risk factors are as follows: The concept of a 'large steep slope' is derived from the Design Specification for Highway Alignment of China (CCCC First Highway Consultants Co., L., 2017), which is defined as the slope whose average gradient and its corresponding segment is located on such a large steep slope length exceed the specification requirement, as shown in Table 4. If the, LSS = 1 for the segment.

The discretization of a horizontal curve radius, vertical gradient are also based on the specification of CCCC First Highway Consultants Co., L. (2017), which is detailed in Table 3. Furthermore, considering the P_HV value for most segments of the Xishang Highway throughout the year is between 35% and 45%, the P_HV and AADT are further discretized according to the judgment of experts. The weather-affected segment is defined as the highway segment affected by agglomerate fog (or black ice) of more than 3 times a year. It is noted that definition criteria of '3 times a year' is proposed by the highway patrol according to their expert judgment. Those segments located nearing to lakes or rivers are more likely subject to agglomerate fog, and the bridge segments located at the tunnel entrance or exit are more likely subject to black ice. Most of the risk factors remain the same throughout the 4 years (except for the WC, AADT and P_HV), while the mitigation strategy variables varied in these years.

Table 4
Definition of large steep slope.

Average Gradient (%)	2.0	2.5	3.0	3.5	4.0	4.5	5.0
Length of the slopes (km)	> 15	. > 7.5	> 3.5	> 3.0	> 2.5	> 2.0	> 2.0

Note: Linear interpolation method is employed to calculate the threshold of slope length for 'large steep slope' if its vertical gradient can not be found in the above Table.

4.2. Results

The individual effect detector tested the influence intensity of risk factors on the highway accident rate. Fig. 6 showed the ranking order of the q_c value for all the risk factors, respectively. The q_c value of the vertical gradient (G) was the maximum among all the risk factors, which meant that G was the leading risk factor (LRF) accounting for the traffic accidents. The individual effect detector ranked the risk factors by their influence on the accident rate in the following order:

G(21.4%) > RC(15.3%) > P_HV(15.5%) > LSS(14.2%) > HCR (14.0%) > AADT(9.4%) > SL(8.6%) > WC(8.5%).

The interactive effect detector classified the interaction types between two risk factors on the highway crash rate. It is noted that the analysis method of the three-factor interactions, or even multi-factor interactions, is the same as that of the two-factor interaction. The interactive impacts measured by the q_c values are shown in Fig. 7. In general, all interaction between risk factors showed varying degrees of enhanced impacts compared to a single factor. The highest q_c values for a two-factor combination came from the interactions of G∩P_HV (0.415), G∩HCR (0.384), which were identified as LRC, as shown in Fig. 7(a). In addition, the highest I_q values came from the interaction of LSS∩WC (0.117) and P_HV∩LSS (0.070), which were identified as ROC, as shown in Fig. 7(b). The lowest I_q value came from the interaction of LSS∩G (-0.088), which indicated that LSS and G were not independent: for example, a steep road segment (G = 2) is very likely a segment in a large steep slope (LSS = 1).

In order to identify the risk scenario for LRF, LRC and ROC, R_h values were computed according to Eqs. (13) and (14). The results were shown in Fig. 6(b) and Fig. 8. The following were the risk scenarios S identified by R_h values: ① $S1$: $\{G = 2\}$, ② $S2$: $\{(G = 2)\&(P\_HV = 2)\}$, ③ $S3$: $\{(G = 2)\&(HCR = 1)\}$, ④ $S4$: $\{(LSS = 1)\&(WC = 1)\}$, ⑤ $S5$: $\{(LSS = 1)\&(P\_HV = 2)\}$.

Given the five risk scenarios above, the effect of various kinds of mitigation strategies was estimated by Eqs. (15) and (16). The computation results were shown in Table 5. It is noted that the variable of a mitigation strategy equals level '0' if the mitigation strategies are not employed. $ME_S^{\alpha,\beta}$ was estimated by setting α ='level 0' and β ='level 1', and the mitigation strategies with p-values less than 0.05 can be identified as effective measures for a risk scenario. The results showed that RS, GB, RTR and GB, were the most effective strategies to mitigate the risk in scenarios S1-S5, respectively. To confirm the reliability of the results, we investigated the opinions of experts in highway operations management and found that the results of the analysis were consistent with the experts' knowledge, which will be shown in the discussion section.

5. Discussion

As previously mentioned, risk evaluation and decision support for mitigation strategies are important concerns in the improvement of highway operations safety.

5.1. Interpretation of risk evaluation results

A persistent problem in risk analysis is that an accident may occur as a consequence of multiple risk factors with interactions instead of an individual factor (Hollnagel, 2002). Different from the existing candidate methods for risk analysis, the CBDD technique focuses on the interaction between different risk factors, which were neglected in previous research. Actually, the interactive effects should be allocated to individual factors to modify the risk evaluation results (Price and MacNicoll, 2015).

Suppose the individual effect and interactive effect of a risk factor x_i^r are denoted as $q_c(x_i^r)$ and $I_q(x_i^r \cap x_j^r)$, (j≠ i), the modified effect

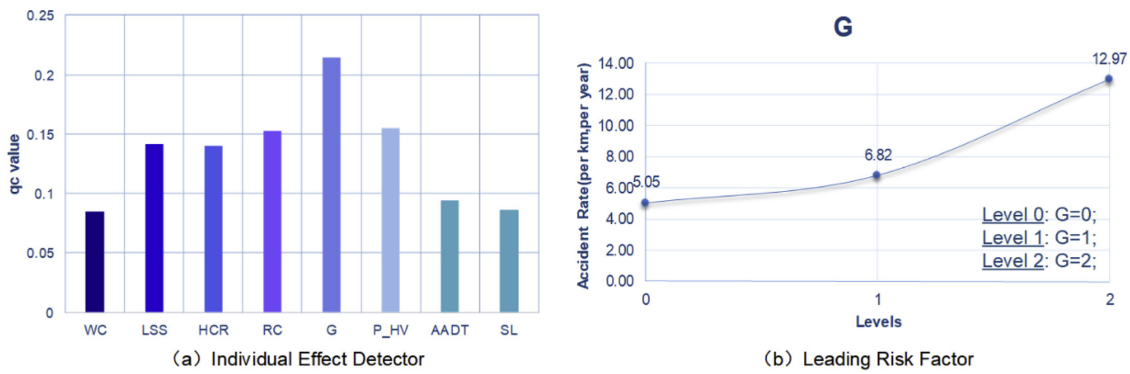


Fig. 6. Effect estimate for individual risk factors.

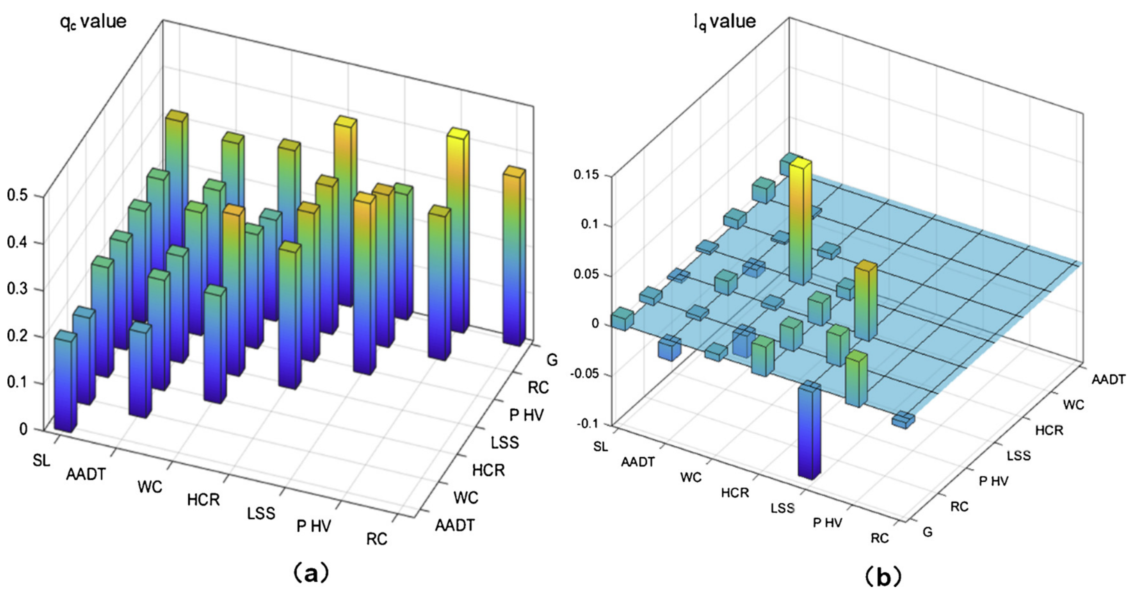


Fig. 7. Interaction between risk factors.

$E_m(x_i^r)$, compensated by the interactive effect, can be computed as

$$E_m(x_i^r) = q_c(x_i^r) + \frac{1}{2}\sum_{j\neq i} I_q(x_i^r \cap x_j^r) + \frac{1}{3}\sum_{j\neq i,k\neq i} I_q(x_i^r \cap x_j^r \cap x_k^r) + \ldots \tag{17}$$

To provide a visual comparison, Fig. 9 shows the difference of risk prioritization results by using a conventional approach and an interaction-compensated approach. Compensated by interactive effect, P_HV, LSS and WC shows a significant difference between the q_c value and E_m value, which indicates that the conventional approach may underestimate the effect of the risk factors with strong interactions. Therefore, the ranking for risk factors is also different for these two approaches.

In addition, Fig. 8 provided a new insight into the interactive effect of risk factors. If factor A and factor B are independent from each other, their effect on the dependent variable should follow the linear superposition principle. In other words, the difference in dependent variable caused by factor A is irrelevant to the value of factor B, which is shown in Fig. 10. Otherwise, it was observed from Fig. 8 that the accident rate corresponding to a certain level of a risk factor becomes more and more unpredictable with the enhancement of the interactive effect. For example, the strong interaction of LSS∩WC results in the most severe scenario (average accident rate = 24.64), which was unexpected by the highway operations managers. The above examples illustrate that factor combinations with strong interactions will more likely lead to unexpected severe consequences than those without.

According to the experts' experience, serious accidents often occur as heavy vehicles are out of control when traveling on a large steep slope. Too frequent braking and black ice on the road surface are the two main reasons that lead to vehicle brake failure and speeding out of control. Moreover, a large proportion of the fatal accidents investigated in this study occurred on the large steep slope with black ice when a heavy vehicle was traveling on it, which accounts for the high interaction of LSS∩WC and LSS∩P_HV.

5.2. Interpretation of mitigation strategies decision results

The results in Table 5 showed that a rumble strip, guardrail barrier and a runaway truck ramp, were recognized as the most effective strategies to mitigate the risk for scenarios S1-S5, respectively. Considering that all scenarios in S1-S5 are related to the vehicle control problem of a steep slope on the highway, all these mitigation strategies are designed either to reduce the speed of vehicles or to provide the last safeguard measurements for runaway vehicles.

However, there are differences between these scenarios. On the one hand, S1-S3 relates to the large gradient of highway segments, which can easily lead to rear-end or head-on collisions due to speeding. The mitigation strategies aimed at these scenarios mainly focus on effective speed control measures, such as the rumble strip. In particular, an enhanced guardrail barrier has a remarkable mitigation effect on preventing fatal accidents at curve segments, such as S3. On the other hand, S4-S5 relates to the large steep slope of the highway, where

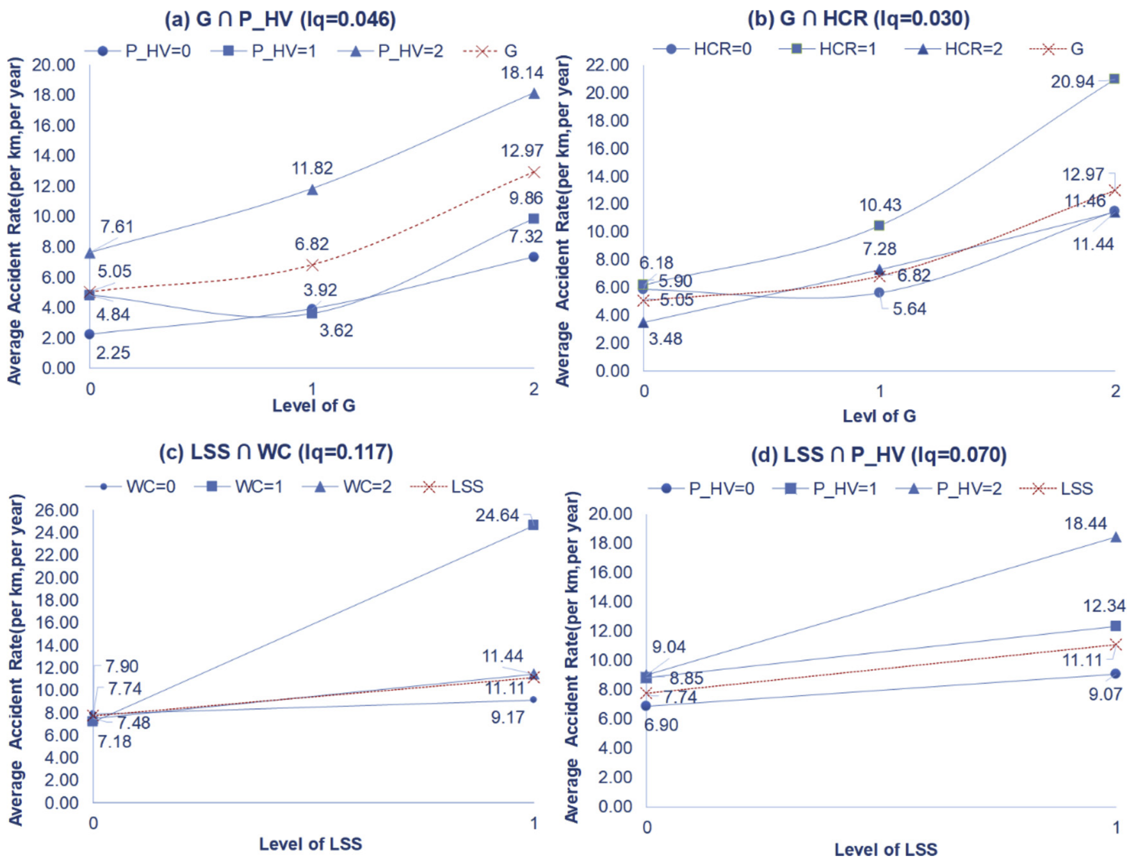


Fig. 8. Risk scenarios identification by R_h values.

vehicles are prone to out-of-control accidents because of frequent braking for a long time. The role of the rumble strip is negligible in the case of a vehicle being out of control. Only the protective measures, such as a runaway truck ramp and guardrail barrier, can reduce the consequences of the accidents.

Another finding in this study is that the misuse of mitigation strategies might worsen the situation and possibly lead to more accidents, which should be avoided. In 2015, many rumble strips were newly installed into the highway sections where accidents frequently occurred, e.g., tunnel sections. As shown in Fig. 11, the mitigation effect of RS on segments of different levels of vertical gradient is very different. The rumble strips installed in tunnel sections with a small gradient (G = 0) not only failed to reduce the number of accidents but led to an increase in the accident rate. According to the opinion of traffic operations managers, the misuse of RS might make vehicle control more difficult and thus cause rear-end or head-on collisions. Therefore, the effect of mitigation strategies on various kinds of risk scenarios should be carefully investigated before being installed.

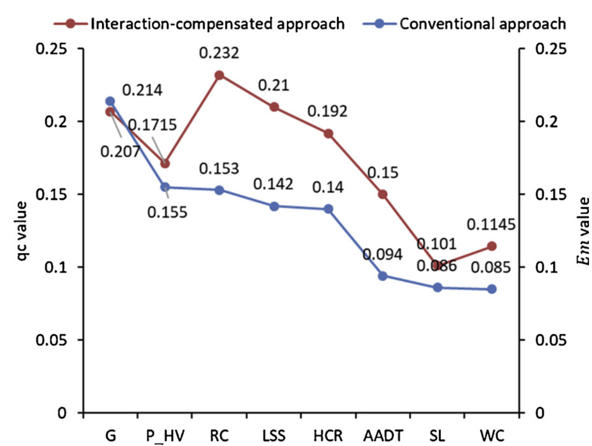


Fig. 9. Risk priority ranking for traffic safety.

Table 5
Mitigation effect of different strategies for various risk scenarios.

Risk Scenario	GB		PM		RS		RTR		Effective Strategies Identification
	$ME_S^{\alpha,\beta}$	p-value	$ME_S^{\alpha,\beta}$	p-value	$ME_S^{\alpha,\beta}$	p-value	$ME_S^{\alpha,\beta}$	p-value	
S1	2.36	0.04	2.48	0.03	5.86	< 0.01	3.16	0.01	RS > RTR > PM > GB
S2	4.65	0.02	5.28	0.02	5.97	0.02	4.86	0.02	RS > PM > RTR > GB
S3	6.82	< 0.01	2.63	0.05	4.74	0.04	0.43	0.82	GB > RS > PM
S4	4.35	0.03	4.49	0.03	0.22	0.92	7.76	< 0.01	RTR > PM > GB
S5	6.58	0.01	4.63	0.03	1.14	0.18	5.42	0.01	GB > RTR > PM

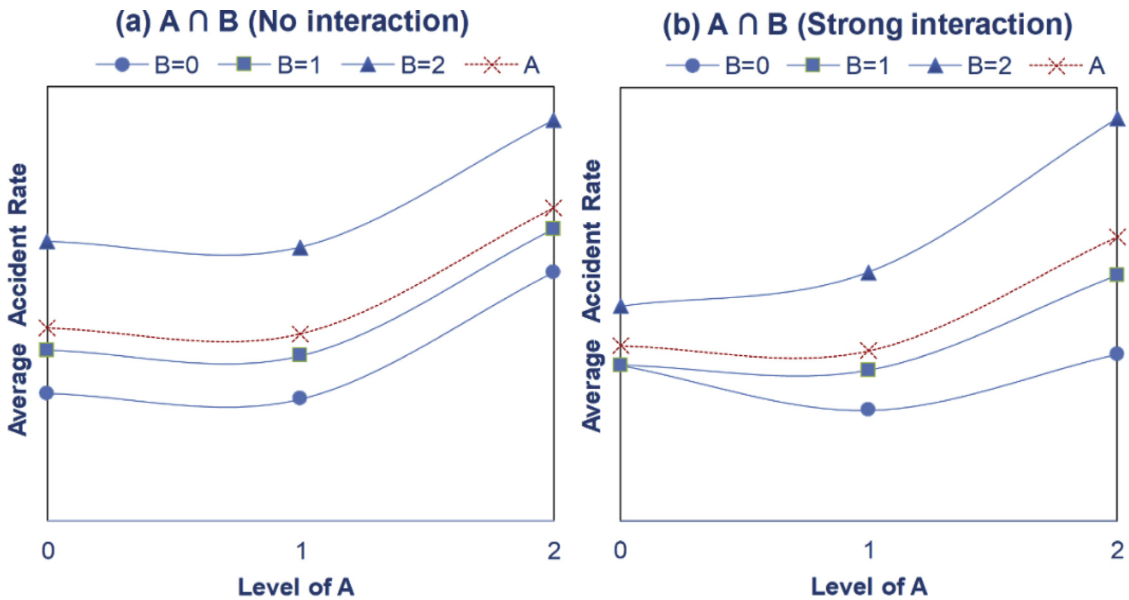


Fig. 10. An illustrative example showing the interaction effect.

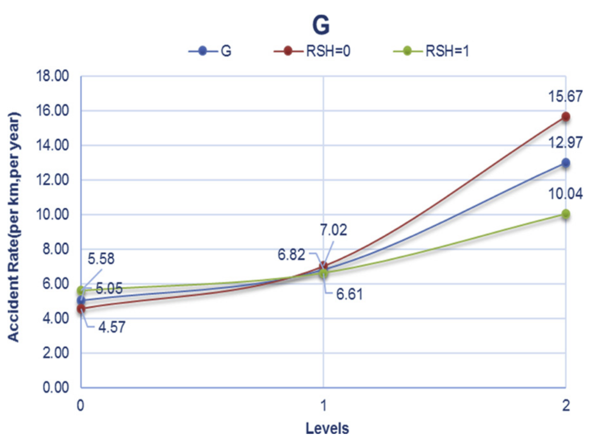


Fig. 11. The mitigation effect of RS on segments of different G values.

5.3. Model characteristics

In comparison with the existing method, the CBDD technique has the following strengths. There is fewer model assumption for the CBDD technique, since it does not specify an explicit function expressions or a causality graph between accident rate and risk factors. Thus, the CBDD technique may avoid the bias resulted from model misspecification. As shown in Fig. 12, different functions expressions lead to different prediction results. Suppose the true model is a quadratic model, severe bias will be caused if a linear model is employed to fit the data, which has been fully displayed in Fig.4 of King and Zeng (2006). Moreover, there is a monotonicity of the generalized linear model (Buddhavarapu et al., 2015; Li and Graham, 2016), which may not always be met in practice, such as the accident data of P_HV = 1 in Fig. 8(a). In addition, the CBDD technique can be used to detect the interactive effect between risk factors as well.

However, this method still has some shortcomings. First, risk factors and mitigation measures have to be identified by expert judgment. Second, the classification of risk factors and mitigation strategies also requires expert knowledge. Sometimes the risk factors and mitigation strategies are not easy to distinguish, and the proposed method may not function if these two types of variables are not properly selected. In summary, although the CBDD technique still requires expert judgment for risk factor and mitigation strategies identification, it can be regarded as a new candidate tool for causal-effect analysis.

6. Conclusion

The frequent occurrence and great damage from traffic accidents, which has become a serious social problem, is worrying the highway

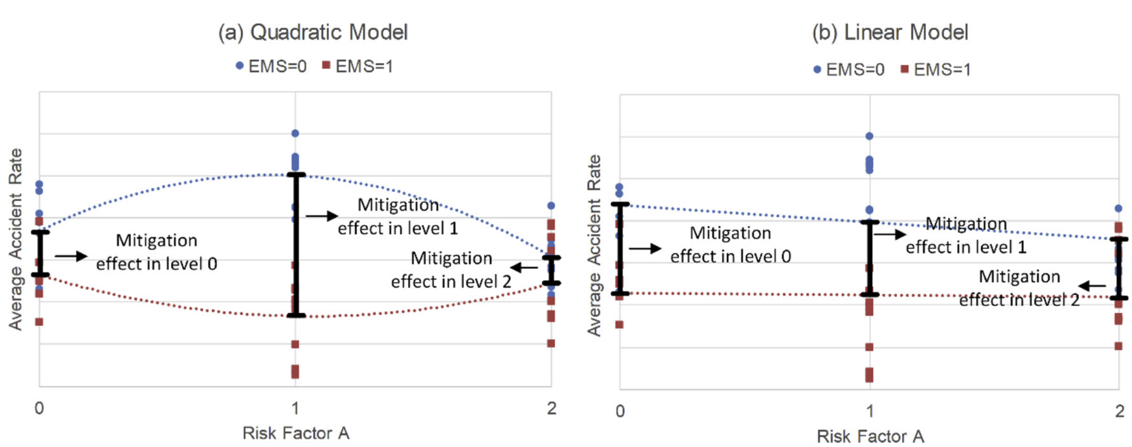


Fig. 12. Model misspecification problem.

operations managers. Risk analysis and mitigation has been a key focus in response to this grim situation. To effectively manage highway safety, it is critical to identify the riskiest scenarios and corresponding effective mitigation strategies. What has not been noticed in previous studies is that nonlinear interaction between risk factors will most likely result in severe highway accidents, which is unexpected by the operations managers.

This paper proposed a new method for highway risk analysis and management. We considered that not only leading individual factors, but also the combined effect of risk factors, deserve our attention. The CBDD technique is employed to quantify the individual and interactive effects of risk factors, and the mitigation effect estimator is used to identify the effect of strategies on a risk scenario. The CBDD technique is applied to the Xishang Highway project to support the mitigation strategies decision for operations safety improvement. In the case study, we found that different strategies have different mitigation effects on the risk scenarios, and the misuse of strategies might not only fail to reduce the accident rate, but lead to a worse situation. Therefore, the mitigation effect of strategies should be analyzed before being put into practice, which also illustrates the application value of this paper.

The biggest limitation in this work which should be noted is that the risk factors were identified based on the expert knowledge. When applying this approach to other roadways, risk factors need to be re-identified and assessed. In future research, a wider variety of risk factors' causal effects (e.g., driving experience, vehicle conditions and drunk driving behavior) on the highway accident rate will be investigated.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We are also thankful for the financial support of: 1) Research project of Chinese Academy of Engineering (no. 2017-XZ-12), 2) National Natural Science Foundation of China (Grant No. 71732001, No.51878311, No. 71821001).

References

Ahmed, M., Huang, H., Abdel-Aty, M., Guevara, B., 2011. Exploring a Bayesian hierarchical approach for developing safety performance functions for a mountainous freeway. Accid. Anal. Prev. 43, 1581–1589.
Al-Bahar, J.F., Crandall, K.C., 1990. Systematic risk management approach for construction projects. J. Constr. Eng. Manag. 116, 533–546.
Barnes, R.J., 1991. The variogram sill and the sample variance. Math. Geol. 23, 673–678.
Boroujerdian, A.M., Saffarzadeh, M., Yousefi, H., Ghassemian, H., 2014. A model to identify high crash road segments with the dynamic segmentation method. Accid. Anal. Prev. 73, 274–287.
Brooks, J.M., Ohsfeldt, R.L., 2013. Squeezing the balloon: propensity scores and unmeasured covariate balance. Health Serv. Res. 48, 1487–1507.
Buddhavarapu, P., Smit, A.F., Prozzi, J.A., 2015. A fully Bayesian before–after analysis of permeable friction course (PFC) pavement wet weather safety. Accid. Anal. Prev. 80, 89–96.
Cafiso, S., D'Agostino, C., Persaud, B., 2018. Investigating the influence of segmentation in estimating safety performance functions for roadway sections. J. Traffic Transp. Eng. 5, 129–136.
Cang, X., Luo, W., 2018. Spatial association detector (SPADE). Int. J. Geogr. Inf. Sci. 32, 2055–2075.
Cao, F., Ge, Y., Wang, J.-F., 2013. Optimal discretization for geographical detectors-based risk assessment. GIsci. Remote Sens. 50, 78–92.
CCCC First Highway Consultants Co., L, 2017. Design Specification for Highway Alignment. Standard Number: JTG D20-2017. China Communication Press, Beijing p. 204.
Chen, S., Saeed, T.U., Alinizzi, M., Lavrenz, S., Labi, S., 2019. Safety sensitivity to roadway characteristics: a comparison across highway classes. Accid. Anal. Prev. 123, 39–50.
Chen, S., Saeed, T.U., Labi, S., 2017. Impact of road-surface condition on rural highway safety: a multivariate random parameters negative binomial approach. Anal. Methods Accid. Res. 16, 75–89.
Chi, C.-F., Lin, S.-Z., Dewi, R.S., 2014. Graphical fault tree analysis for fatal falls in the construction industry. Accid. Anal. Prev. 72, 359–369.
Cuhadar, A., Shalaby, K., Tasdoken, S., 2002. Automatic segmentation of pavement condition data using wavelet transform, IEEE CCECE2002. Canadian Conference on Electrical and Computer Engineering. Conference Proceedings (Cat. No.02CH37373). pp. 1009–1014 vol.1002.
Degn Eskesen, S., Tengborg, P., Kampmann, J., Holst Veicherts, T., 2004. Guidelines for tunnelling risk management: international Tunnelling Association, working group No. 2. Tunn. Undergr. Space Technol. 19, 217–237.
Friedman, N., Geiger, D., Goldszmidt, M., 1997. Bayesian network classifiers. Mach. Learn. 29, 131–163.
Hansen, B.B., Bowers, J., 2008. Covariate balance in simple, stratified and clustered comparative studies. Stat. Sci. 23, 219–236.
Hollnagel, E., 2002. Understanding accidents-from root causes to performance variability. Proceedings of the IEEE 7th Conference on Human Factors and Power Plants. pp. 1-1.
Hu, Y., Wang, J., Li, X., Ren, D., Zhu, J., 2011. Geographical Detector-based risk assessment of the under-five mortality in the 2008 Wenchuan Earthquake, China. PLoS One 6, e21427.
Huang, J., Wang, J., Bo, Y., Xu, C., Hu, M., Huang, D., 2014. Identification of Health Risks of Hand, Foot and Mouth Disease in China Using the Geographical Detector Technique. Int. J. Environ. Res. Public Health 11, 3407.
Imai, K., Ratkovic, M., 2014. Covariate balancing propensity score. J. R. Stat. Soc. Series B Stat. Methodol. 76, 243–263.
Jiang, X., Wang, Q., Li, R., 2018. Investigating factors affecting carbon emission in China and the USA: a perspective of stratified heterogeneity. J. Clean. Prod. 199, 85–92.
King, G., Zeng, L., 2006. The dangers of extreme counterfactuals. Political Anal. 14, 131–159.
Li, C., Ding, L., Zhong, B., 2019. Highway Planning and Design in the Qinghai-Tibet Plateau of China: A Cost-Safety Balance Perspective. Engineering.
Li, H., Graham, D.J., 2016. Heterogeneous treatment effects of speed cameras on road safety. Accid. Anal. Prev. 97, 153–161.
Luo, A.S., 2017. Analysis of groundwater nitrate contamination in the Central Valley: comparison of the geodetector method, principal component analysis and geographically weighted regression. ISPRS Int. J. Geo-Information 6, 297.
Newman, M.E.J., 2005. Power laws, Pareto distributions and Zipf's law. Contemp. Phys. 46, 323–351.
Price, B., MacNicoll, M., 2015. Multiple interacting risk factors: on methods for allocating risk factor interactions. Risk Anal. 35, 931–940.
Qu, X., Yang, Y., Liu, Z., Jin, S., Weng, J., 2014. Potential crash risks of expressway on-ramps and off-ramps: A case study in Beijing, China. Saf. Sci. 70, 58–62.
SafetyCube, 2016. SafetyCube - European Road Safety Decision Support System. URL:. https://www.roadsafety-dss.eu/#/risk-factor-search.
Shankar, V., Mannering, F., Barfield, W., 1995. Effect of roadway geometrics and environmental factors on rural freeway accident frequencies. Accid. Anal. Prev. 27, 371–389.
Zhang, S., H.-q.J, Song, Y., Yu, W., Sun, L., 2016. Genetic study identifies CBLN4 as a novel susceptibility gene for accident proneness. Front. Eng. 3, 30–38.
Wang, H., Mo, X., Wang, Y., Liu, R., Qiu, P., Dai, J., 2016a. Assessing Chinese coach drivers' fitness to drive: the development of a toolkit based on cognition measurements. Accid. Anal. Prev. 95, 395–404.
Wang, J.-F., Zhang, T.-L., Fu, B.-J., 2016b. A measure of spatial stratified heterogeneity. Ecol. Indic. 67, 250–256.
Wang, J., Xu, C., 2017. Geodetector: principle and prospective. Acta Geographica Sinica 72, 116–134.
Wang, J.F., Li, X.H., Christakos, G., Liao, Y.L., Zhang, T., Gu, X., Zheng, X.Y., 2010. Geographical detectors-based health risk assessment and its application in the neural tube defects study of the Heshun Region, China. Int. J. Geogr. Inf. Sci. 24, 107–127.
Washington, S., Haque, M.M., Oh, J., Lee, D., 2014. Applying quantile regression for modeling equivalent property damage only crashes to identify accident blackspots. Accid. Anal. Prev. 66, 136–146.
Wu, X., Jiang, Z., Zhang, L., Skibniewski, M.J., Zhong, J., 2015. Dynamic risk analysis for adjacent buildings in tunneling environments: a Bayesian network based approach. Stoch. Environ. Res. Risk Assess. 29, 1447–1461.
Zhang, G., Yau, K.K.W., Chen, G., 2013. Risk factors associated with traffic violations and accident severity in China. Accid. Anal. Prev. 59, 18–25.
Zhao, Y., Deng, Q., Lin, Q., Cai, C., 2017. Quantitative analysis of the impacts of terrestrial environmental factors on precipitation variation over the Beibu Gulf Economic Zone in Coastal Southwest China. Sci. Rep. 7, 44412.