

doi: 10.3969/j.issn.1002-0268.2022.10.017

# 高速公路合流区冲突极值建模与交通事故预测

郑来<sup>1</sup>, 侯芹忠<sup>2</sup>, 郭延永<sup>3</sup>, 杨晓冬<sup>4</sup>

(1. 哈尔滨工业大学 交通科学与工程学院, 黑龙江 哈尔滨 150090;

2. 哈尔滨工业大学(威海)汽车工程学院, 山东 威海 264209;

3. 东南大学 交通学院, 江苏 南京 210096; 4. 哈尔滨职业技术学院, 黑龙江 哈尔滨 150081)

**摘要:**为了更好地利用交通冲突技术开展高速公路合流区交通安全研究,提出了一种基于贝叶斯层级超阈值理论的冲突极值建模与交通事故预测方法。以辽宁省7个高速公路合流区的交通运行视频和交通事故数据为基础,利用后侵入时间(PET)识别了潜在合流冲突,进而构建了融合超阈值极值理论和贝叶斯层级结构(含数据层、过程层和先验层)的贝叶斯层级超阈值极值模型,提出了基于分位数回归的阈值选取方法,最后利用标定的最优模型对合流区交通事故进行了预测并对事故影响因素进行了分析。结果表明:调查时段内合流区共发生898次合流冲突;在构建的稳态模型、非稳态显著模型和非稳态全模型中,考虑冲突极值非稳态性和异质性的非稳态显著模型最优;最优模型的事故预测精度较高,其预测事故数与观测事故数的平均误差和平均绝对误差分别为1.0次/a和2.1次/a;通过建模发现加速车道长度、合流车辆类型、主线车辆类型、加速车道小时平均交通量对交通事故有显著影响,其中加速车道越长、主线车辆车型越大,交通事故发生的概率越小;合流车辆类型越小、加速车道小时平均交通量越大,交通事故发生的概率越大。所提出的贝叶斯层级超阈值方法解决了冲突极值的稀少性、非稳态性和异质性问题,实现了基于短期观测冲突数据的交通事故可靠预测。

**关键词:**交通安全;交通事故预测;极值统计;高速公路合流区;交通冲突

中图分类号:U268.6

文献标识码:A

文章编号:1002-0268(2022)10-0132-09

## Conflict Extremum Modeling and Traffic Accident Prediction for Expressway Merging Areas

ZHENG Lai<sup>1</sup>, HOU Qin-zhong<sup>2</sup>, GUO Yan-yong<sup>3</sup>, YANG Xiao-dong<sup>4</sup>

(1. School of Transportation Science and Engineering, Harbin Institute of Technology, Harbin Heilongjiang 150090, China;

2. School of Automotive Engineering, Harbin Institute of Technology (Weihai), Weihai Shandong 264209, China;

3. School of Transportation, Southeast University, Nanjing Jiangsu 210096, China;

4. Harbin Vocational & Technical College, Harbin Heilongjiang 150081, China)

**Abstract:** To better use the traffic conflict technique to investigate the safety of expressway merging area, the methods for conflict extremum modeling and traffic accident prediction based on Bayesian hierarchical supra-threshold theory are proposed. On the basis of traffic operation videos and traffic accident data collected from 7 expressway merging areas in Liaoning Province, the potential merging conflicts are identified using the PET. The Bayesian hierarchical supra-threshold extremum model, which integrated hierarchical supra-threshold theory and Bayesian hierarchical structure (including data layer, process layer and prior layer), is constructed. An approach for threshold selection based on quantile regression is developed. Finally, the traffic accidents of the merging areas are predicted and the influencing factors of accidents are analyzed by using the calibrated optimal model. The result shows that (1) There are 898 merging conflicts in the merging areas during the investigation. (2) In the constructed stationary model, non-stationary significant model and non-stationary full model, the non-stationary significant model which accounts for the non-stationarity and

收稿日期:2020-12-08

基金项目:国家自然科学基金面上项目(52072097);国家自然科学基金青年基金项目(71701046)

作者简介:郑来(1985-),男,湖北南漳人,博士,副教授.(zhenglai@hit.edu.cn)

heterogeneity in conflict extremums is the optimal model. (3) The accident prediction accuracy of the optimal model is high, the average error and average absolute error of the predicted accidents and the observed accidents are 1.0 times/a and 2.1 times/a, respectively. (4) It is found through modeling that the length of acceleration lane, type of merging vehicle, type of mainline vehicle, and average hourly traffic volume on acceleration lane have significant influence on the accident occurrence. The longer the acceleration lane and the larger the size of the mainline vehicle, the smaller the probability of accident occurrence, while the smaller the size of merging vehicle and the larger the average hourly traffic volume on acceleration lane, the higher the probability of accident occurrence. The proposed Bayesian hierarchical supra-threshold method overcome the rarity, non-stationarity and heterogeneity issues of traffic conflict extremums, and it also realized the reliable traffic accident prediction based on short-term observed conflict data.

**Key words:** traffic safety; traffic accident prediction; extremum statistics; expressway merging area; traffic conflict

## 0 引言

合流区是高速公路的重要组成部分,同时也是高速公路运行安全的薄弱环节。车辆合流过程中,驾驶员需进行加速、选择可接受间隙、变道等一系列复杂操作,容易出现操作不当从而引发交通事故。统计表明,合流区交通事故占高速公路事故总数的比例一直居高不下,约为高速公路基本路段交通事故的 4~6 倍<sup>[1-2]</sup>。

在对高速公路合流区交通安全的研究中,交通冲突技术得到了越来越多的应用。交通冲突是指车辆在时间和空间上相互接近至一定程度,以至于若不改变其运动状态,就会发生碰撞的交通事件<sup>[3]</sup>。车辆合流过程中会出现较为频繁的侧向冲突和追尾冲突,可为开展高速公路合流区交通安全研究提供比交通事故更为充足的数据支撑。为了识别合流区的交通冲突,学者们多采用后侵入时间(Post Encroachment Time, PET)作为冲突严重程度的衡量指标,通过选取合适的后侵入时间阈值界定交通冲突,进而通过冲突数(率)分析高速公路合流区的安全性<sup>[4-6]</sup>。相关研究发现,交通流量、冲突车辆速度差、加速车道长度、重型车比例等因素对合流区交通冲突有显著影响<sup>[5-9]</sup>。

虽然交通冲突技术为高速公路合流区交通安全研究提供了新的手段,但交通冲突作为合流区安全分析指标的有效性尚未被充分证明<sup>[10]</sup>。本研究旨在通过极值理论建立合流区冲突与交通事故之间的关系模型,在验证交通冲突技术有效性的基础上更好地开展高速公路合流区交通安全研究。极值理论作为一种分析极小概率事件分布的理论,能够通过短时段内常见事件(如降水量、交通冲突)发生的概率外推长时段内罕见极值事件(如洪水、交通事故)发生的概率,已被

应用于交叉口冲突-事故关系研究<sup>[11-13]</sup>、高速公路基本路段冲突-事故关系研究<sup>[14-15]</sup>、高速公路合流区冲突-事故关系研究<sup>[16]</sup>。然而,已有研究主要针对单个研究对象分别建立冲突极值统计模型,虽然证明了该模型在交通事故预测方面的可行性,但都存在单个研究对象冲突极值样本数量较少的问题,在很大程度上影响了事故预测精度。比如,Zheng 等<sup>[16]</sup>针对 15 个高速公路合流区分别构建了冲突极值统计模型,研究发现模型预测事故数与观测事故数的误差最高可达 55.9 次/a。

为了解决单个合流区冲突样本数量有限的问题,也为了更有效地利用合流区交通冲突预测交通事故,本研究将结合贝叶斯层级超阈值理论,构建能融合多个合流区冲突数据的极值统计模型并据此预测合流区交通事故,以期拓展交通冲突极值统计建模的方法体系,也为基于交通冲突的高速公路合流区交通安全研究提供新思路。

## 1 合流区交通冲突判别

### 1.1 数据采集与处理

本研究共采集了 7 个高速公路合流区的交通运行数据,分别为葫芦岛、绥中、兴城、高升、五龙背、通远堡和本溪。前 4 个合流区来自于沈山(沈阳至山海关)高速公路,双向 8 车道,设计速度为 120 km/h;后 3 个合流区来自于沈丹(沈阳至丹东)高速公路,双向 4 车道,设计速度为 100 km/h。

数据采集采用录像法,共采集了白天晴好天气情况下 53 h 的视频。其中,葫芦岛、兴城、本溪的视频时长均为 7 h,其他 4 个合流区的视频时长均为 8 h。基于录制的视频,采用人工观测法进行了分车道、分车型的交通流量统计,具体结果见表 1。

表 1 交通流量与合流冲突统计情况一览表

Tab. 1 Statistic summary of traffic volumes and merging conflicts

高速公路	合流区	长度/m	视频录制时间	总交通量/veh		合流冲突		总事故数/次
				主线(小/中/大型车)	加速车道(小/中/大型车)	总数/次	PET 均值/s	
沈山高速	葫芦岛	243	2013年8月21日 08:00—15:00	1 436/580/1 844	399/7/64	170	3.10	13
沈山高速	绥中	200	2013年8月21日 08:00—16:00	1 291/166/2 392	455/16/214	176	2.66	16
沈山高速	兴城	223	2013年8月21日 08:00—15:00	1 778/369/2 432	258/19/79	88	2.95	7
沈山高速	高升	210	2013年8月21日 08:00—16:00	1 761/230/1 628	606/7/111	144	2.56	5
沈丹高速	五龙背	200	2013年8月22日 08:00—16:00	1 490/81/157	154/8/38	53	3.25	4
沈丹高速	通远堡	200	2013年8月22日 08:00—16:00	1 227/59/278	123/5/14	48	3.59	3
沈丹高速	本溪	200	2013年8月22日 08:00—15:00	1 344/100/191	1 784/49/77	219	3.64	6

交通事故数据来自于沈山高速和沈丹高速的路政管理部门, 时间跨度为 2010 年 1 月至 2012 年 12 月, 该时段与视频采集时间相近, 且在 2010—2013 年间所调查合流区的道路、交通、环境条件无显著变化。基于事故发生时间、地点桩号、上下行方向等信息, 得到了合流区影响范围内白天时段的交通事故, 共 54 次, 具体见表 1。

## 1.2 合流冲突判别

高速公路合流区的合流车辆主要影响主线最外侧两车道的交通运行<sup>[4]</sup>。因此, 本研究重点关注合流车辆与主线最右侧两车道上车辆间的交通冲突。交通冲突严重程度衡量指标选用后侵入时间(PET), 即后车头部到达冲突区域的时间与前车尾

部离开冲突区域的时间差<sup>[3]</sup>。

合流过程中的 PET 是指跟随车辆的头部到达侵入线的时间与先导车辆的尾部离开侵入线的时间之差, 其中一辆车为合流车辆。侵入线指垂直于车道分界线并经过车道分界线与车辆变道轨迹交点的虚拟直线<sup>[17]</sup>。PET 的具体测量如图 1 所示, 合流过程中, 合流车辆(S)可能与每条车道上的前车(L)和后车(F)产生交通冲突, 因此, 每一辆合流车辆最多可测出 4 个 PET 值, 为了保证极值样本的独立性, 本研究选取其中的最小值作为该合流行为的 PET。实际测量利用 VirtualDub 软件逐帧播放视频并计算 PET, 测量精度为 0.04 s (每 25 帧为 1 s)。

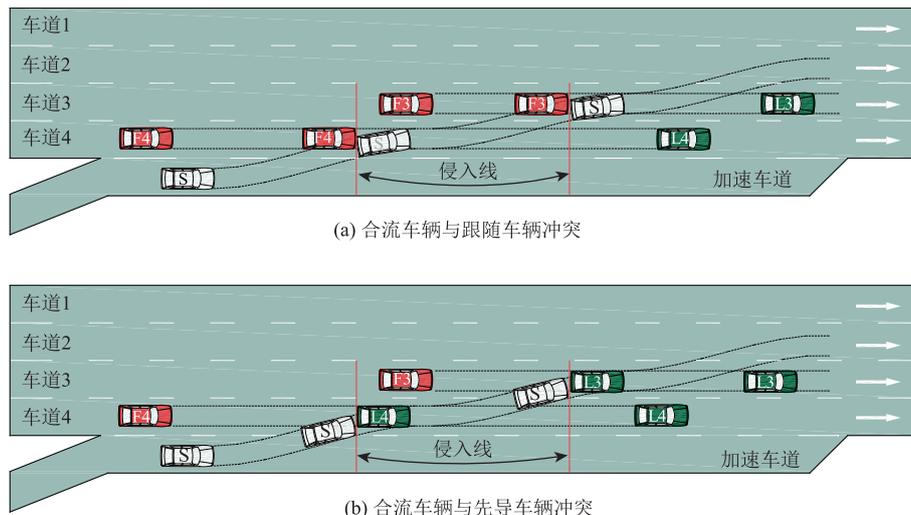


图 1 合流冲突 PET 测量示意图

Fig. 1 Illustration of PET measurement of merging conflicts

由于部分车辆合流时, 视频视野范围内主线上并未出现其他车辆, 故此时不存在合流冲突。最后, 共识别得到合流冲突 898 次, 其中 PET 最小值为 0.20 s、最大值为 11.24 s。各合流区的冲突次数及

对应 PET 均值见表 1。

## 2 基于冲突极值模型的事故预测

交通冲突与交通事故具有过程相似性, 并且交

通事故可视为交通冲突的极值事件, 即时空接近度趋近于 0 的交通冲突为交通事故。因此, 可通过冲突极值建模外推交通事故发生的概率。本研究以超阈值极值理论为基础构建冲突极值模型并预测交通事故。

## 2.1 贝叶斯层级超阈值极值模型

### 2.1.1 超阈值极值模型

假设  $X_1, X_2, \dots, X_n$  是来自分布函数  $F$  的一系列独立样本, 将值大于某一阈值  $u$  的样本视为极值样本的方法即为超阈值方法。当阈值足够大时, 超阈值极值  $y = x - u$  服从广义帕累托分布 (Generalized Pareto Distribution, GPD)<sup>[18]</sup>。在集合  $\{y: y > 0 \text{ 且 } 1 + \frac{\xi y}{\sigma} > 0\}$  中, GPD 的形式如下:

$$G(y) = 1 - \left(1 + \frac{\xi y}{\sigma}\right)^{-1/\xi}, \quad (1)$$

式中,  $G(y)$  为 GPD 分布;  $\sigma > 0$ , 为尺度参数;  $\xi$  为形状参数。

### 2.1.2 贝叶斯层级结构

为了表征冲突极值的非稳态特性, 在超阈值极值模型的基础上引入贝叶斯层级结构, 将影响冲突极值的相关因素融入模型中。层级模型包括数据层、过程层和先验知识层。其中, 数据层采用超阈值模型对冲突极值进行建模, 其形式如下:

$$G(y_{ik} | \phi_{ik}, \xi_{ik}) = 1 - \left(1 + \frac{\xi_{ik} y_{ik}}{\sigma_{ik}}\right)^{-1/\xi_{ik}}, \quad (2)$$

式中,  $G(y_{ik} | \phi_{ik}, \xi_{ik})$  为广义极值分布;  $y_{ik}$  为合流区  $i$  上的第  $k$  个极值 ( $i = 1, \dots, s; k = 1, \dots, n_i$ ; 其中  $s$  为合流区数量,  $n_i$  为第  $i$  个合流区上的冲突极值个数);  $\sigma_{ik}$  和  $\xi_{ik}$  分别为 GPD 的尺度参数和形状参数。过程层则是利用潜高斯过程将 GPD 模型参数与影响冲突极值的变量进行关联, 关联函数选用恒等相关函数, 具体如下:

$$\begin{cases} \phi_{ik} = \alpha_{\phi 0} + \alpha_{\phi} \mathbf{X} + \varepsilon_{\phi i} \\ \xi_{ik} = \alpha_{\xi 0} + \varepsilon_{\xi i} \end{cases}, \quad (3)$$

式中,  $\phi_{ik} = \log \sigma_{ik}$ , 为变换后的尺度参数, 以保证尺度参数的非负性;  $\alpha_{\phi 0}$  和  $\alpha_{\xi 0}$  分别为尺度参数和形状参数的常数项;  $\mathbf{X}$  为影响冲突极值非稳态特性的协变量集合, 为向量;  $\alpha_{\phi}$  为对应的参数向量;  $\varepsilon_{\phi i}$  和  $\varepsilon_{\xi i}$  分别为各合流区尺度参数和形状参数的随机误差项, 以表征各合流区未观测因素导致的冲突极值异质性。需要说明的是, 由于超阈值极值模型对形状参数的敏感性以及形状参数精准估计的难度, 一般不考虑相关变量对形状参数的影响<sup>[18]</sup>。将常数项和

随机误差项结合, 简化得到随机截距过程层模型:

$$\begin{cases} \phi_{ik} = \alpha_{\phi 0i} + \alpha_{\phi} \mathbf{X} \\ \xi_{ik} = \alpha_{\xi 0i} \end{cases}, \quad (4)$$

式中,  $\alpha_{\phi 0i}$  和  $\alpha_{\xi 0i}$  分别为合流区  $i$  的形状参数和尺度参数常数项。

先验层为过程层模型参数分配先验分布。假设所有模型参数相互独立, 由于现阶段没有可靠的先验信息, 所以参数  $\alpha_{\phi 0i}$  和  $\alpha_{\phi}$  均采用无信息先验概率分布, 具体为采用 0 均值大方差的正态分布作为其先验分布, 即  $\alpha_{\phi 0i}, \alpha_{\phi} \sim N(0, 10^6)$ 。同样, 由于超阈值极值模型对其形状参数的敏感性, 不恰当的先验分布容易导致模型不收敛, 所以借鉴了已有冲突极值模型研究中有相关形状参数的先验信息。在已有研究中, 估计得到的形状参数均位于  $(-1, 1)$  之间<sup>[11, 13]</sup>, 因此本研究选用均匀分布作为其先验分布, 即  $\alpha_{\xi 0i} \sim N(0, 0.25)$ 。

## 2.2 阈值选取方法

超阈值模型构建首先需要确定合理的阈值  $u$ , 才能估计参数  $\sigma$  和  $\xi$ 。如果阈值选取不合理, 将会导致错误的模型估计。常用的阈值选取方法包括平均剩余寿命图及阈值稳定性分析图等图解方法和均方差法等数值方法<sup>[18]</sup>, 然而这些方法所确定的均为稳态阈值, 并未考虑外在因素对阈值的影响。为克服这一缺陷, 本研究选用分位数回归方法确定阈值<sup>[19]</sup>。从直观上看, 高分位数对应的阈值一般都能保证所对应的超阈值样本服从 GPD 分布。

分位数回归模型的具体形式如下:

$$Q(\tau | \mathbf{X}) = \mathbf{X}' \boldsymbol{\beta}^{\tau}, \quad (5)$$

式中,  $\tau \in (0, 1)$ , 为分位数;  $Q(\tau | \mathbf{X})$  为分位数  $\tau$  对应的因变量值 (即  $q_{\tau}$ );  $\mathbf{X}_{\tau}$  为影响阈值非稳态特性的协变量集合为向量;  $\boldsymbol{\beta}^{\tau}$  为参数向量。分位数回归模型参数的标定可通过求解式 (6) 所示的优化问题解决:

$$\min_{\boldsymbol{\beta} \in R^p} \left[ \sum_{\tau: q_{\tau} \geq x'_{\tau} \boldsymbol{\beta}^{\tau}} \tau | q_{\tau} - x'_{\tau} \boldsymbol{\beta}^{\tau} | + \sum_{\tau: q_{\tau} < x'_{\tau} \boldsymbol{\beta}^{\tau}} (1 - \tau) | q_{\tau} - x'_{\tau} \boldsymbol{\beta}^{\tau} | \right], \quad (6)$$

式中,  $x'_{\tau} \boldsymbol{\beta}^{\tau}$  为式 (5) 向量集合中的一个变量。

在分位数回归确定的阈值基础上, 进一步采用阈值稳定性分析图验证所选取分位数的合理性。该方法的基本原理为: 如果超出阈值  $u_0$  的样本服从  $GPD(\sigma_0, \xi)$ , 则对任意阈值  $u > u_0$ , 其阈值  $u$  超出值同样服从  $GPD(\sigma_u, \xi)$ , 且 GPD 对应的修订后的尺度参数  $\sigma_u$  和形状参数  $\xi$  在理论上是恒定不变的, 其中:

$$\sigma_u = \sigma_0 + \xi(u - u_0), \quad (7)$$

式中,  $\sigma_0$  为超出阈值  $u_0$  的 GPD 分布的尺度参数。

### 2.3 模型求解与比选

采用贝叶斯方法对贝叶斯层级超阈值极值模型进行估计。在确定阈值、模型结构和先验分布的基础上,模型参数集合  $\theta$  (向量) 的后验概率密度函数为:

$$p(\theta | Y) \propto \underbrace{p_1(Y | \theta_1)}_{\text{数据}} \underbrace{p_2(\theta_1 | \theta_2)}_{\text{过程}} \underbrace{p_3(\theta_2)}_{\text{先验}}, \quad (8)$$

式中,  $Y$  为输入数据集合, 为向量;  $p_1, p_2, p_3$  分别为数据层, 过程层和先验层的概率密度函数, 由参数  $\theta_j(j=1, 2)$  决定, 其中  $\theta_1 = [\phi, \xi]^T, \theta_2 = [\alpha_{\phi 0}, \alpha_{\phi}, \alpha_{\xi 0}]^T$ 。  $p_1, p_2, p_3$  的具体计算公式如下:

$$p_1(Y | \theta_1) = \prod_{i=1}^s \prod_{k=1}^{n_i} \left( \frac{1}{\exp \phi_{ik}} \right) \left( 1 + \xi_{ik} \frac{y_{ik}}{\exp \phi_{ik}} \right)^{-1-1/\xi_{ik}}, \quad (9)$$

$$p_2(\theta_1 | \theta_2) = \frac{1}{\sqrt{2\pi\delta_{\phi}^2}} \exp \left[ -\frac{(\phi - \phi_i)^2}{2\delta_{\phi}^2} \right] \times \frac{1}{\sqrt{2\pi\delta_{\xi}^2}} \exp \left[ -\frac{(\xi - \xi_i)^2}{2\delta_{\xi}^2} \right], \quad (10)$$

$$p_3(\theta_2) = p_{\alpha_{\phi 0}}(\alpha_{\phi 0}) \times p_{\alpha_{\phi}}(\alpha_{\phi}) \times p_{\alpha_{\xi 0}}(\alpha_{\xi 0}), \quad (11)$$

式中,  $\phi, \phi_i, \xi, \xi_i$  分别为尺度参数和形状参数的样本值和假设的先验分布的均值,  $\phi_i$  和  $\xi_i$  来自于式 (4);  $\delta_{\phi}^2$  和  $\delta_{\xi}^2$  为对应尺度参数和形状参数先验分布的方差;  $p_{\alpha_{\phi 0}}(\alpha_{\phi 0}), p_{\alpha_{\phi}}(\alpha_{\phi}), p_{\alpha_{\xi 0}}(\alpha_{\xi 0})$  分别为参数  $\alpha_{\phi 0}, \alpha_{\phi}, \alpha_{\xi 0}$  的后验概率密度函数。结合式 (9)~式(11) 即可推导出模型参数的后验概率分布, 通常通过马尔可夫链蒙特卡罗 (Markov Chain Monte Carlo, MCMC) 仿真完成。

在具体构建贝叶斯层级冲突极值模型时, 考虑不同变量对冲突极值的影响, 即将不同变量引入式 (4) 可形成一系列模型。最优模型的选取采用方差信息准则 (Deviance Information Criterion, DIC)<sup>[20]</sup>。方差信息准则权衡了贝叶斯模型的拟合优度和复杂度, 计算公式为:

$$DIC = \bar{D} + p_D, \quad (12)$$

式中,  $\bar{D}$  为模型贝叶斯方差的后验均值;  $p_D$  为模型中有效参数的数量。一般而言,  $DIC$  值越小, 模型越好。当两个模型的  $DIC$  值之差大于 5 时,  $DIC$  值较小的模型明显优于  $DIC$  值较大的模型<sup>[21]</sup>。

### 2.4 交通事故预测方法

$PET$  衡量的交通冲突严重程度表征了事故发生的风险或交通冲突与交通事故的接近程度。 $PET$  值越小, 交通冲突与交通事故越接近, 当  $PET \leq 0$  时意味着交通事故必然发生。考虑到超阈值极值模型仅适用于大于某一阈值的极值事件, 对  $PET$  取负变换为  $NPET$  (Negated  $PET$ ), 此时  $NPET \geq 0$  (即  $PET \leq 0$ ) 即意味着会发生交通事故。根据上述方法标定得

到基于  $NPET$  的 GPD 分布, 进而可推导出交通事故的发生概率, 即:

$$R = \Pr(NPET \geq 0) = 1 - \Pr(NPET < 0) = \left( 1 - \frac{u}{\sigma} \right)^{-1/\xi}, \quad (13)$$

式中,  $R$  为对应冲突观测时段  $t$  的交通事故发生概率。假设冲突观测时段能够代表更长时段  $T$  的交通运行状况, 那么该时段内的交通事故数  $N_T$  为:

$$N_T = \frac{T}{t} R. \quad (14)$$

## 3 建模结果分析

利用高速公路合流区  $PET$  数据进行贝叶斯层级冲突极值模型构建, 其中, 考虑的冲突极值非稳态特性影响因素包括加速车道长度 ( $LEN$ , m)、合流车辆类型 ( $MVT$ , 1, 2, 3 分别代表小、中、大型车)、主线车辆类型 ( $TVT$ , 1, 2, 3 分别代表小、中、大型车)、加速车道小时平均交通量 ( $MVV$ , veh/h)、主线小时平均交通量 ( $TVV$ , veh/h)。模型估计时, 分位数回归通过 R 软件包 `quantreg` 实现, 贝叶斯估计通过 WinBUGS 软件实现。

### 3.1 阈值选取结果分析

为保证阈值足够大从而使超阈值样本服从 GPD 分布, 本研究首先分别选取了 80%, 85% 和 90% 这 3 个分位数, 并将  $LEN, MVT, TVT, MVV, TVV$  这 5 个协变量代入分位数回归模型中进行参数估计, 结果表明: 3 个分位数模型中的显著变量均为  $LEN, MVT, TVT, MVV$ , 对应的参数估计结果见表 2。

表 2 分位数回归结果

Tab. 2 Result of quantile regression							
分位数	变量	估计值	标准差	$t$	$p >  t $	95% 置信区间	
80%	常数项	-0.903	0.278	-3.25	0.001	-1.448	-0.357
	$LEN^{**}$	-2.410	0.792	-3.04	0.002	-3.965	-0.855
	$MVT^{**}$	0.204	0.059	3.47	0.001	0.089	0.319
	$TVT^*$	-0.144	0.089	-1.62	0.095	-0.318	0.030
	$MVV^{**}$	-0.003	0.001	-3.63	0	-0.004	-0.001
85%	常数项	-0.518	0.305	-1.7	0.09	-1.117	0.081
	$LEN^{**}$	-2.815	0.846	-3.33	0.001	-4.474	-1.155
	$MVT^{**}$	0.148	0.051	2.88	0.004	0.047	0.248
	$TVT^{**}$	-0.150	0.069	-2.18	0.029	-0.286	-0.015
	$MVV^{**}$	-0.003	0.001	-3.44	0.001	-0.005	-0.001
90%	常数项	-0.389	0.332	-1.17	0.242	-1.041	0.263
	$LEN^{**}$	-2.262	0.980	-2.31	0.021	-4.186	-0.337
	$MVT^{**}$	0.114	0.053	2.16	0.031	0.010	0.218
	$TVT^{**}$	-0.208	0.065	-3.2	0.001	-0.335	-0.080
	$MVV^{**}$	-0.002	0.001	-1.98	0.048	-0.004	0.000

注: “\*\*” 和 “\*” 分别表示该变量在 0.05 和 0.1 水平上显著。

针对每一个分位数进行了 GPD 分布的参数估计, 结果表明: 在 85% 和 90% 分位数上满足阈值稳定性要求, 即估计得到的修正后的尺度参数和形状参数基本不变, 但在 90% 分位数上对应的估计值方差较大。考虑到模型估计方差和偏差的平衡, 本研究选取 85% 分位数对应的值为最终阈值, 即  $u = -0.518 - 2.815 \times LEN + 0.148 \times MVT - 0.150 \times TVT - 0.003 \times MVV$ 。

### 3.2 模型标定结果分析

通过向后逐步回归法构建贝叶斯层级冲突极值模型, 即首先将 5 个变量均通过关联函数与 GPD 的尺度参数连接, 对应的模型称为非稳态全模型; 进而逐步减少变量的数量直至变量数量为 0, 此时对应的模型为稳态模型; 期间构建的仅包括显著变量 ( $p < 0.05$ ) 的模型称为非稳态显著模型。

采用 Winbugs14 软件对上述模型进行标定时, 构造 2 条包含 100 000 次迭代的 MCMC 链进行贝叶斯推理。通过检查 2 条链的迭代历史图可确定 MCMC 算法是否达到收敛; 此外, 也可以通过 BGR (Brooks-Gelman-Rubin) 值进行收敛判断, 一般认为 BGR 值小于 1.2 时为收敛<sup>[22]</sup>。将前 50 000 次迭代视为燃烧样本, 不参与模型参数的估计, 利用后 50 000 次迭代值进行模型参数估计。稳态模型、非稳态显著模型、非稳态全模型的估计结果见表 3。3 个模型均达到收敛, 其中非稳态显著模型中仅有 MVT, MVV 为显著变量, 对应的迭代历史和 BGR 值如图 2 和图 3 所示。

由表 3 可知, 稳态模型、非稳态显著模型和非稳态全模型的 DIC 值分别为 6 567.1, 6 549.3 和 6 557.3。相较于稳态模型, 非稳态模型的 DIC 值明显更小, 这说明考虑协变量对冲突极值的影响明显提高了贝叶斯层级极值模型的拟合优度。此外, 两个非稳态模型中, 仅包含显著变量的非稳态显著模型的 DIC 值更小, 并且其与非稳态全模型的 DIC 值之差为 8.0。因此, 非稳态显著模型为最优模型。

### 3.3 事故预测结果分析

#### 3.3.1 事故预测精度分析

根据标定的最优模型并结合式 (13)、式 (14), 计算得到各合流区的预测年均事故数。式 (14) 中,  $T = 365 \text{ d} \times 12 \text{ h/d} = 4\,380 \text{ h}$ ,  $t = 8 \text{ h}$  (绥中、高升、五龙背和通远堡合流区) 或  $7 \text{ h}$  (葫芦岛、兴城和本溪合流区)。预测事故数与观测事故数之间的关系如图 4 所示。

由图 4 可以看出, 预测事故数与观测事故数间具有明显的正相关关系 (计算得出相关性系数  $\rho = 0.820\,1$ , 显著性水平  $sig = 0.023\,9$ ), 线性拟合优度值  $R^2 = 0.670\,4$ 。为进一步量化事故预测精度, 计算

表 3 贝叶斯层级极值模型估计结果

Tab. 3 Estimation result obtained by Bayesian hierarchical extremum model

变量	稳态模型		非稳态显著模型		非稳态全模型	
	估计值	标准差	估计值	标准差	估计值	标准差
$\alpha_{\phi 01}$	-1.051	0.244	-0.795	0.332	-0.721	0.873
$\alpha_{\phi 02}$	-0.811	0.215	-0.610	0.277	-0.533	1.030
$\alpha_{\phi 03}$	-0.751	0.180	-0.546	0.269	-0.406	0.824
$\alpha_{\phi 04}$	-0.900	0.285	-0.666	0.283	-0.552	0.868
$\alpha_{\phi 05}$	-0.702	0.226	-0.559	0.249	-0.477	0.786
$\alpha_{\phi 06}$	-0.817	0.315	-0.605	0.304	-0.566	0.835
$\alpha_{\phi 07}$	-0.158	0.218	-0.708	0.389	-0.648	0.966
LEN	—	—	—	—	0.382	3.592
MVT	—	—	-0.243	0.063	-0.278	0.090
TVT	—	—	—	—	-0.099	0.093
MVV	—	—	0.005	0.002	0.005	0.002
TVV	—	—	—	—	0.000	0.001
$\alpha_{\xi 01}$	-0.269	0.169	-0.373	0.189	-0.272	0.204
$\alpha_{\xi 02}$	-0.321	0.155	-0.292	0.143	-0.388	0.170
$\alpha_{\xi 03}$	-0.380	0.122	-0.509	0.152	-0.479	0.153
$\alpha_{\xi 04}$	-0.382	0.216	-0.587	0.273	-0.514	0.270
$\alpha_{\xi 05}$	-0.434	0.215	-0.528	0.225	-0.538	0.265
$\alpha_{\xi 06}$	-0.364	0.226	-0.400	0.270	-0.453	0.283
$\alpha_{\xi 07}$	-0.367	0.186	-0.720	0.233	-0.651	0.207
$\bar{D}$	6 561		6 544		6 549	
$p_D$	6.1		5.3		8.3	
DIC	6 567.1		6 549.3		6 557.3	

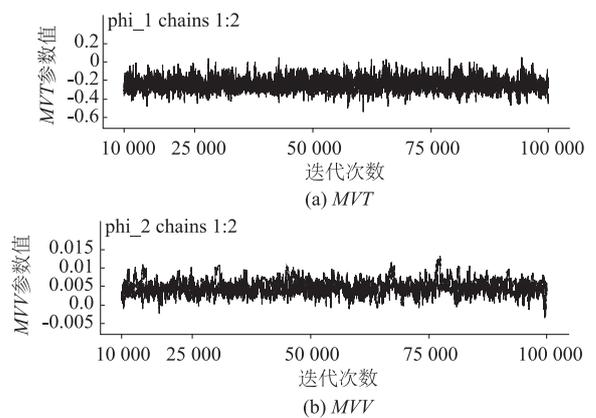


图 2 非稳态显著模型迭代历史图  
Fig. 2 Iterative history diagrams of non-stationary significant model

了预测事故数与观测事故数间的平均误差 (Mean Error, ME) 和平均绝对误差 (Mean Absolute Error, MAE), 计算公式为:

$$ME = \sum_{i=1}^7 \frac{(P_i - O_i)}{7}, \quad (15)$$

$$MAE = \sum_{i=1}^7 \frac{|P_i - O_i|}{7}, \quad (16)$$

式中,  $P_i$  为第  $i$  个合流区的预测事故数;  $O_i$  为第  $i$  合流区的观测事故数。计算结果见表 4, 模型预测事故数与观测事故数之间的平均误差和平均绝对误差均较小, 分别为 1.0 次/a 和 2.1 次/a, 这说明基于贝叶斯层级冲突极值统计的事故预测方法具有较高的预测精度。此外, 平均误差为正, 表明模型在总体上高估了事故发生次数。

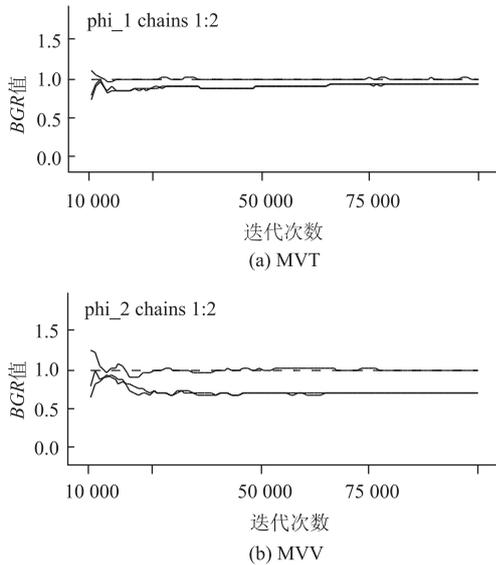


图 3 非稳态显著模型 BGR 图

Fig. 3 BGR diagrams of non-stationary significant model

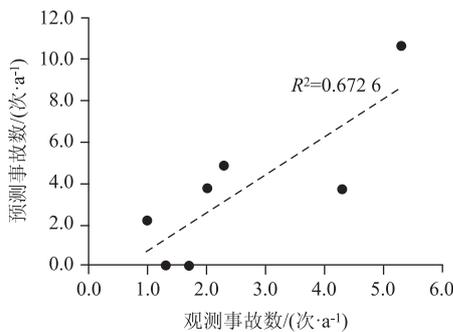


图 4 观测事故数与预测事故数关系

Fig. 4 Relationship between observed accidents and estimated accidents

### 3.3.2 事故影响因素分析

建模发现, 加速车道长度 ( $LEN$ )、合流车辆类型 ( $MVT$ )、主线车辆类型 ( $TVT$ )、加速车道小时平均交通量 ( $MVV$ ) 为显著变量, 但这些变量并不直接影响交通事故的风险, 而是影响 GPD 分布的阈值和尺度参数。阈值和尺度参数决定了大部分极值事件的集中度, 不同阈值和尺度参数对应的 GPD 分

表 4 事故预测精度

Tab. 4 Accuracy of accident estimation

合流区	观测事故数/ (次·a <sup>-1</sup> )	预测事故数/ (次·a <sup>-1</sup> )	平均误差/ (次·a <sup>-1</sup> )	平均绝对误差/ (次·a <sup>-1</sup> )
葫芦岛	4.3	3.7		
绥中	5.3	10.6		
兴城	2.3	4.8		
高升	1.7	0.0	1.0	2.1
五龙背	1.3	0.0		
通远堡	1.0	2.2		
本溪	2.0	3.8		

布示意如图 5 所示。当尺度参数和形状参数保持不变时, 阈值越小, GPD 分布整体上越偏离事故风险区域 (即越安全); 但阈值和形状参数不变时, 尺度参数越大, GPD 分布跨度越大, 分布曲线的右尾部越可能进入事故风险区域 (即越危险)。

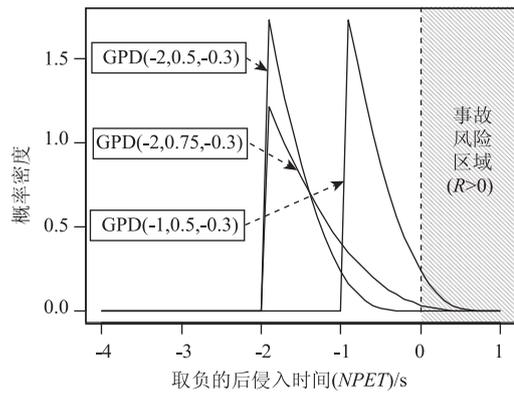


图 5 不同阈值和尺度参数对应的 GPD 分布示意

Fig. 5 Illustration of GPD distribution with different thresholds and scale parameters

上述 4 个显著变量中,  $LEN$  和  $TVT$  仅对阈值有显著影响, 其系数均为负值 (分别为 -2.185 和 -0.150)。为了量化分析其对合流安全性的影响, 假设其他变量均取其均值, 计算得到某一变量值逐渐增大时的事故发生概率, 如图 6 (a) 和图 6 (b) 所示。由图可知,  $LEN$  和  $TVT$  值越大, 阈值越小, 对应的合流区越安全。即加速车道长度较长时, 发生交通事故的风险更低, 因为加速车道较长时, 合流车辆驾驶员有更为充足的时间完成加速和可接受间隙选择, 因而降低了合流事故风险, 但当加速车道长度增加至一定值时, 其继续增加的安全边际效益则会降低, 这也说明加速车道长度并不是越长越好。此外, 当主线上有大型车时, 合流车辆驾驶员往往更加谨慎, 因而合流事故风险会降低。

变量  $MVT$  和  $MVV$  对阈值和尺度参数均有显著影响, 且  $MVT$  值越大, 阈值越大, 尺度参数值越小;

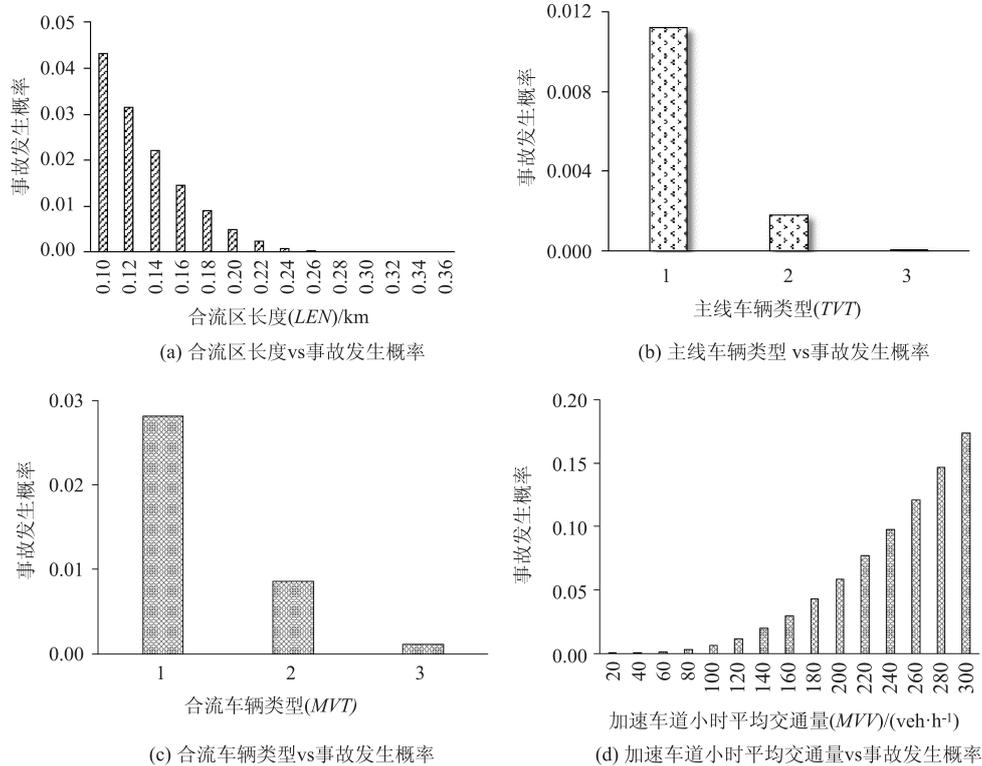


图 6 变量对合流事故风险的影响

Fig. 6 Influence of variables on merging accident risk

MVV 值越大, 阈值越小, 尺度参数值越大。为了综合分析这些变量对合流安全性的影响, 同样假设其他变量均取其均值, 计算得到某一变量值逐渐增大时的事故发生概率。对 MVT 而言, 保持 LEN, MVV, TVT 不变时, 不同合流车型对应的事故发生概率变化如图 6 (c) 所示, 随着车型的增大交通事故发生概率逐渐降低。由于大型车的尺寸和性能与小型车差异较大, 一般认为大型车比例增高会导致交通事故数增加, 这与本研究发现存在相悖之处。潜在原因是本研究使用的冲突样本中车辆类型分布不均衡, 合流车辆主要是小型车。对 MVV 而言, 保持 LEN, MVT, TVT 不变时, 合流车流量增加时对应的事故发生概率变化如图 6 (d) 所示, 随着合流车流量增加, 事故发生概率增大。

4 结论

(1) 提出了贝叶斯层级冲突极值建模方法, 一方面融合了不同地点的冲突极值进行联合建模, 解决了冲突极值的稀少性问题; 另一方面引入了协变量和随机误差项, 解决了冲突极值的非稳态性和异质性问题, 显著提升了模型拟合优度。该方法可实现基于短期观测冲突数据的交通事故预测, 且相较于传统的基于冲突极值统计的交通事故预测方法更

为可靠。

(2) 采用贝叶斯层级冲突极值建模方法构建了高速公路合流区交通事故预测模型, 结果表明该模型具有较高的预测精度, 模型预测事故数和观测事故数的平均误差仅为 1.0 次/a, 平均绝对误差为 2.1 次/a。

(3) 加速车道长度、合流车辆类型、主线车辆类型、加速车道小时平均交通量对冲突极值有显著影响, 进而影响交通事故发生概率。其中, 加速车道越长、主线车辆车型越大, 交通事故发生概率越小; 合流车辆类型越小、加速车道小时平均交通量越大, 交通事故发生概率越大。

(4) 本研究主要从模型的角度改进了传统的冲突极值建模方法, 但受冲突数据采集时间较短的限制, 个别合流区事故预测结果并不理想, 下一步将增加采集时间以提升冲突样本的代表性, 收集合流区线形数据以更全面揭示冲突极值的非稳态性, 从而进一步验证贝叶斯层级冲突极值建模方法的优越性。

参考文献:

References:

[1] 胡江碧, 何禄诚, 王荣华. 高速公路互通立交安全性评价研究综述 [J]. 中国公路学报, 2020, 33 (7): 17-28.  
 HU Jiang-bi, HE Lu-cheng, WANG Rong-hua. Review of

- Safety Evaluation of Freeway Interchange [J]. *China Journal of Highway and Transport*, 2020, 33 (7): 17-28.
- [2] 王春娥. 高速公路立交变速车道长度研究 [J]. *公路交通科技*, 2013, 30 (1): 120-125.  
WANG Chun-e. Length of Speed-change Lane at Highway Interchange [J]. *Journal of Highway and Transportation Research and Development*, 2013, 30 (1): 120-125.
- [3] 朱顺应, 蒋若曦, 王红, 等. 机动车交通冲突技术研究综述 [J]. *中国公路学报*, 2020, 33 (2): 15-33.  
ZHU Shun-ying, JIANG Ruo-xi, WANG Hong, et al. Review of Research on Traffic Conflict Techniques [J]. *China Journal of Highway and Transport*, 2020, 33 (2): 15-33.
- [4] 孟祥海, 林兰平. 高速公路分合流区潜在事故风险研究 [J]. *中国安全科学学报*, 2015, 25 (8): 164-170.  
MENG Xiang-hai, LIN Lan-ping. Research on Potential Crash Risk in Freeway Merging and Diverging Areas [J]. *China Safety Science Journal*, 2015, 25 (8): 164-170.
- [5] 张南, 秦兵杰. 基于交通冲突技术的高速公路合流区段安全性研究 [J]. *综合运输*, 2016, 38 (10): 55-59.  
ZHANG Nan, QIN Bing-jie. Safety Evaluation of Freeway Merge Segment Based on Traffic Conflict Technique [J]. *China Transportation Review*, 2016, 38 (10): 55-59.
- [6] 马艳丽, 祁首铭, 吴昊天, 等. 基于PET算法的匝道合流区交通冲突识别模型 [J]. *交通运输系统工程与信息*, 2018, 18 (2): 142-148.  
MA Yan-li, QI Shou-ming, WU Hao-tian, et al. Traffic Conflict Identification Model Based on Post Encroachment Time Algorithm in Ramp Merging Area [J]. *Journal of Transportation Systems Engineering and Information Technology*, 2018, 18 (2): 142-148.
- [7] LI S, XIANG Q J, MA Y F, et al. Crash Risk Prediction Modeling Based on the Traffic Conflict Technique and a Microscopic Simulation for Freeway Interchange Merging Areas [J]. *International Journal of Environmental Research and Public Health*, 2016, 13 (11): 1157.
- [8] CAI Y W, XIANG Q J, LI Y, et al. Collision Prediction Model for Interchange Merging Area Based on Traffic Conflict Technique [C] // *Proceedings of 2018 3rd IEEE International Conference on Intelligent Transportation Engineering (ICITE)*, Singapore: IEEE, 2018: 324-329.
- [9] GUO Y F, XIANG Q J, LI S, et al. Impacts of Large Vehicles on Traffic Safety in Freeway Interchange Merging Areas and Improvement Measures [J]. *MATEC Web of Conferences*, 2017, 124: 04004.
- [10] ZHENG L, ISMAIL K, MENG X H. Traffic Conflict Techniques for Road Safety Analysis: Open Questions and Some Insights [J]. *Canadian Journal of Civil Engineering*, 2014, 41 (7): 633-641.
- [11] SONGCHITRUKSA P, TARKO A P. The Extreme Value Theory Approach to Safety Estimation [J]. *Accident Analysis & Prevention*, 2006, 38 (4): 811-822.
- [12] 王晨, 夏井新, 陆振波, 等. 基于微观仿真与极值理论的城市交叉口安全评价方法 [J]. *中国公路学报*, 2018, 31 (4): 288-295, 303.  
WANG Chen, XIA Jing-xin, LU Zhen-bo, et al. Safety Evaluation Method Based on Traffic Simulation and Extreme Value Theory [J]. *China Journal of Highway and Transport*, 2018, 31 (4): 288-295, 303.
- [13] FU C, SAYED T. Bayesian Hierarchical Modeling of the Non-stationary Traffic Conflict Extremes for Crash Estimation [J]. *Analytic Methods in Accident Research*, 2019, 23: 100100.
- [14] ZHENG L, ISMAIL K, MENG X H. Freeway Safety Estimation Using Extreme Value Theory Approaches: A Comparative Study [J]. *Accident Analysis & Prevention*, 2014, 62: 32-41.
- [15] 郑来, 邓晓庆, 孟祥海. 基于PET极值统计的高速公路车道变换行为安全性研究 [J]. *公路交通科技*, 2016, 33 (8): 120-126.  
ZHENG Lai, DENG Xiao-qing, MENG Xiang-hai. Study on Safety of Lane Changing Behaviours on Expressway Based on PET Extreme Statistics [J]. *Journal of Highway and Transportation Research and Development*, 2016, 33 (8): 120-126.
- [16] ZHENG L, ISMAIL K, SAYED T, et al. Bivariate Extreme Value Modeling for Road Safety Estimation [J]. *Accident Analysis & Prevention*, 2018, 120: 83-91.
- [17] 郑来. 基于交通冲突极值统计的安全分析模型研究 [D]. 哈尔滨: 哈尔滨工业大学, 2014.  
ZHENG Lai. Research of Safety Analysis Models Based on Extreme Statistics of Traffic Conflicts [D]. Harbin: Harbin Institute of Technology, 2014.
- [18] COLES S. An Introduction to Statistical Modeling of Extreme Values [M]. London: Springer, 2001.
- [19] KOENKER R, BASSETT G. Regression Quantiles [J]. *Econometrica*, 1978, 46 (1): 33-50.
- [20] SPIEGELHALTER D J, BEST N G, CARLIN B P, et al. Bayesian Measures of Model Complexity and Fit [J]. *Journal of the Royal Statistical Society B (Statistical Methodology)*, 2002, 64 (4): 583-639.
- [21] ZHENG L, ISMAIL K. A Generalized Exponential Link Function to Map a Conflict Indicator into Severity Index within Safety Continuum Framework [J]. *Accident Analysis & Prevention*, 2017, 102: 23-30.
- [22] 郭延永, 刘攀, 吴瑶, 等. 基于贝叶斯多元泊松-对数正态分布的交通冲突模型 [J]. *中国公路学报*, 2018, 31 (1): 101-109.  
GUO Yan-yong, LIU Pan, WU Yao, et al. Traffic Conflict Model Based on Bayesian Multivariate Poisson-lognormal Normal Distribution [J]. *China Journal of Highway and Transport*, 2018, 31 (1): 101-109.