

DOI:10.20035/j.issn.1671-2668.2023.06.009

## 山区高速公路交通事故特征分析与动态风险评估\*

高银钧<sup>1</sup>, 刘德敬<sup>1</sup>, 熊昌安<sup>2,3</sup>, 孔令智<sup>2,3</sup>, 杨文臣<sup>2,3</sup>(1.云南南景高速公路有限公司, 云南 普洱 676200; 2.云南省交通规划设计研究院 陆地交通气象灾害防治技术  
国家工程实验室, 云南 昆明 650200; 3.云南省数字交通重点实验室, 云南 昆明 650200)

**摘要:**为实现山区高速公路交通事故风险动态评估,选择云南省某典型山区高速公路的道路条件数据、交通事故数据和交通流数据,采用配对式病例一对照法对事故数据和非事故数据进行匹配处理,通过描述性统计分析和二元 Logistic 回归分析,研究山区高速公路交通事故时空分布特征,并构建动态风险评估模型。结果表明,季节变化带来的客观环境差异并不足以导致交通事故数明显的季度波动;交通事故主要集中在雨季和节假日出行高峰的 4 月、5 月、8 月、10 月,主要发生在连续下坡、连续下坡+隧道路段等道路线形组合不良区域;采用显著变量建模、阈值为 0.5、事故与非事故比例为 1:1 时,构建的二元 Logistic 回归模型的预测效果最佳。

**关键词:**交通安全;山区高速公路;交通事故;时空分布;动态风险评估

中图分类号:U491.31

文献标志码:A

文章编号:1671-2668(2023)06-0044-04

随着国家公路网的规划和实施,中国高速公路建设里程不断延伸,机动化交通需求迅速增长,机动车保有量及驾驶证申领人数逐年增加,在“人-车-路-环”四要素的综合作用下,山区高速公路事故频发,交通安全形势严峻<sup>[1]</sup>。受复杂地形地貌、自然环境等因素限制,山区高速公路的平、纵、横线形指标较平原区高速公路低,且平纵、直曲等特殊组合路段多,为了克服高差、避开地质灾害区域,连续下坡、桥隧群、隧道群等特殊结构物较普遍,山区高速公路行车条件不连续<sup>[2-4]</sup>。在安全防护设施、监管技术与应急管理不到位的情况下,重特大交通事故频繁。

当前关于山区高速公路交通安全管理的研究多停留在传统的静态、被动、后评估层面,如改善道路几何线形设计<sup>[5]</sup>、在事故高发路段上安装控速设施<sup>[6]</sup>、实施速度执法<sup>[7]</sup>等,缺乏主动性、预见性和对风险的响应及时性,且存在成本高、见效慢、干扰交通正常运营等问题,无法对高速公路交通安全问题做出迅速响应。在大幅改善山区高速公路交通安全现状的现实需求下,有必要以特殊的山区高速公路道路条件为基础,创新研究山区高速公路交通事故形成机理及主动调控策略,切实提高动态交通安全管理水平,以尽可能地减少交通事故发生及降低事

故严重程度。鉴于此,本文采集云南省某典型山区高速公路的道路条件数据、交通事故数据和交通流数据,采用配对式病例一对照法对事故数据和非事故数据进行匹配处理,构建二元 Logistic 回归模型,实现山区高速公路交通事故风险动态评估。

## 1 数据来源与处理

### 1.1 数据来源

(1) 事故数据。统计滇西某典型山区高速公路的道路条件数据、2018—2020 年交通事故数据,根据事故发生时间和桩号范围确定对应交调站,匹配车速、交通流量数据,剔除信息完整度小于 80% 的数据,共获得 2 833 起有效交通事故数据样本。

(2) 非事故数据。根据事故发生时间和地点提取事故发生前 5~10 min 的车速、交通流量数据,提取规则如下:1) 非事故组的日期与对应事故所在日期不同;2) 非事故组的星期与对应事故所在星期相同;3) 非事故组的时间段与对应事故所在时间段相同;4) 非事故组的地点与对应事故所在地点相同;5) 非事故组所对应事故前后各 5 h 内在相同地点无事故。采用配对式病例一对照法<sup>[8-9]</sup>按照事故:非事故=1:4 的原则选择非事故案例,将其填充进模型样本库,共获得 11 332 起非事故数据样本。

\* 基金项目:云南省交通运输厅科技创新及示范项目(云交科教便[2021]6号;云交科教便 2021-90-2);云南省数字交通重点实验室(202205AG070008)

## 1.2 数据预处理

### 1.2.1 车道时间占有率计算

车道时间占有率描述的是某一时间内所有车辆通过某断面的累计时间与该段时间的百分比<sup>[10]</sup>。交通流量和车速可从交调系统中获取,统计时间为5 min。车道时间占有率按下式计算:

$$R_t = \frac{1}{t_L} \sum_{i=1}^n t_i \quad (1)$$

式中: $R_t$ 为车道时间占有率; $t_L$ 为总观测时间; $n$ 为该路段的车辆数; $t_i$ 为第*i*辆车的占用时间。

### 1.2.2 交通运行变量获取

将获取的交通基本数据转换为表1所示14个字段。为保证上下游数据一致,对数据进行以下预处理:若上游或下游的交通数据均为零,则删除该条交通数据。选择 Avg\_Cnt<sub>u</sub>、Avg\_Occ<sub>u</sub>、Avg\_Spd<sub>u</sub>、Avg\_Cnt<sub>d</sub>、Avg\_Occ<sub>d</sub>、Avg\_Spd<sub>d</sub>共6个字段为显著自变量,定义事故的发生与否为因变量,其中“1”代表事故,“0”代表非事故。为验证模型的性能,按7:3的比例将采集的数据分为训练集和测试集。

## 2 基本特征

### 2.1 时间分布特征

分析2018—2020年该山区高速公路交通事故

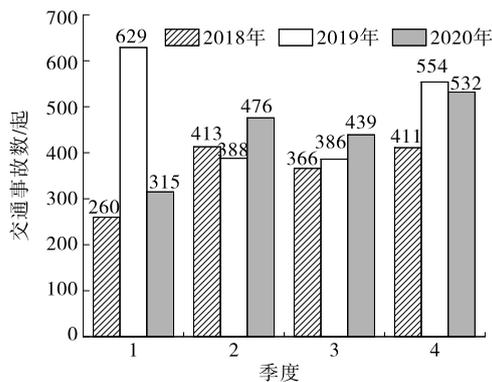


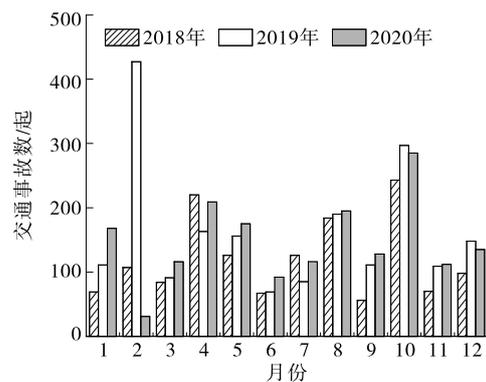
图1 2018—2020年某山区高速公路交通事故时间分布特征

由图1可知:1) 二季度和第三季度的交通事故数较接近,四季度的事故数较高。总体来看,导致季度差异的主要原因是出行量,季节变化带来的客观环境差异并不足以导致事故数明显的季度波动。2) 4月、5月、8月和10月是交通事故高峰期,其余月份的交通事故分布较均衡。究其原因,4月和5月为雨季,受降雨影响路面附着系数下降、能见度降低,车辆制动性能变差,事故频发;8月和10月是节假日人们外出活动最集中的月份,受安全设施不足、

表1 字段表

序号	字段	说明
1	Avg_Cnt <sub>u</sub>	上游交调站的5 min分类型车流量
2	Avg_Occ <sub>u</sub>	上游交调站的5 min分类型占用率
3	Avg_Spd <sub>u</sub>	上游交调站的5 min分类型车速
4	Avg_Cnt <sub>d</sub>	下游交调站的5 min分类型车流量
5	Avg_Occ <sub>d</sub>	下游交调站的5 min分类型占用率
6	Avg_Spd <sub>d</sub>	下游交调站的5 min分类型车速
7	Road_Width	路面宽度
8	Lane_Width	车道宽度
9	Inner_Width	内肩宽
10	Outer_Width	外肩宽
11	I_Me_Width	中央分隔带宽度
12	lane_num	车道数
13	terrain	地形
14	weather	天气条件

的时间分布特征,结果见图1。考虑到2020年上半年受新型冠状病毒感染的影响,2020年出行量较低,记录的总事故数偏少,将2020年交通事故数据设为对照组。



交通流相互干扰等影响,事故高发。除天气因素外,2019年2月由于沿线某高速铁路开通,集散客流量骤增,导致该月交通事故数远高于2018年2月和2020年2月。

### 2.2 空间分布特征

以2018—2020年交通事故数据为基础,分上、下行统计研究路段内逐桩的年事故次数及3年事故次数总和,分析交通事故空间分布特征。上、下行交通事故空间分布见图2。

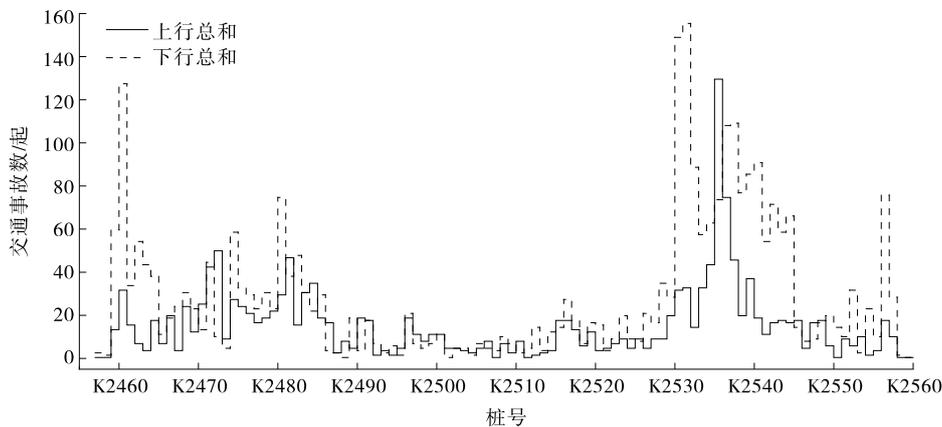


图2 上下行逐桩交通事故数对比

由图2可知:交通事故数整体表现为下行方向高于上行方向;上行方向 K2470—K2485、K2530—K2545 路段,下行方向 K2460—K2485、K2530—K2545 路段交通事故表现出明显的聚集高发特点。究其原因,上行 K2470—K2485 路段和下行 K2460—K2485 路段为典型连续下坡路段,车辆行驶时容易因频繁制动减速使轮毂温度过高,导致刹车失灵;K2530—K2545 路段包含一段隧道,车辆在进入隧道前会产生减速行为,导致后车与隧道入口过渡段的慢行车辆或停止车辆发生冲突,引发追尾等交通事故。

### 3 动态风险评估建模

#### 3.1 二元 Logistic 回归模型构建

Logistic 回归是应用非常广泛的一种分类机器学习算法,它将数据拟合到 Logistic 函数中,对事件发生概率进行预测,分析事件的影响因素<sup>[11-12]</sup>。逻辑回归是一种对数概率模型,是离散选择法模型之一,其对数概率分布公式如下:

$$P(Y=1 | X=x) = \frac{e^{x\beta}}{1 + e^{x\beta}} \quad (2)$$

式中: $Y$  为响应变量(因变量); $X$  为解释变量(自变量); $\beta$  为参数(常采用最大似然估计)。

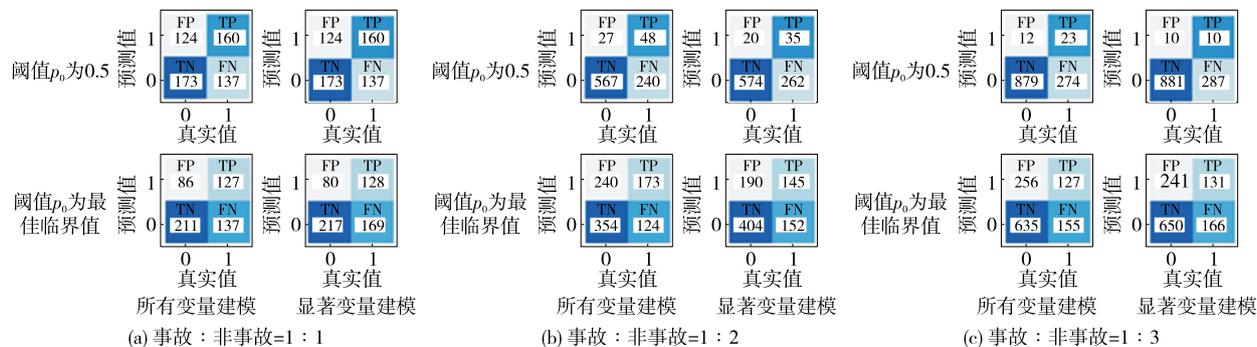
简而言之,Logistic 回归假设数据服从伯努利分布,采用极大化似然函数的方法,运用梯度下降来求解参数,达到将数据二分类的目的。本文的因变量是交通事故是否发生,故采用二元 Logistic 回归分析法。模型如下:

$$P(Y=1 | x_1, x_2, \dots, x_n) = \frac{1}{1 + e^{-(\alpha + \sum_{i=1}^n \beta_i x_i)}} \quad (3)$$

式中: $P$  为发生交通事故的概率; $Y$  为因变量; $x_i$  为自变量( $i=1, 2, \dots, n$ ); $\alpha, \beta$  分别为回归截距和回归系数。

#### 3.2 结果分析

分别用所有变量和显著变量建立二元 Logistic 回归模型,分别取事故与非事故的比例为 1:1、1:2、1:3 进行调整,根据模型结果的混淆矩阵(见图3)计算模型的正确率、特异度、准确率和灵敏度,并将4个指标绘制在雷达图中进行对比分析(见图4)。



FP 表示将负样本预测为正样本,预测错误;TP 表示将正样本预测为正样本,预测正确;TN 表示将负样本预测为负样本,预测正确;FN 表示将正样本预测为负样本,预测错误

图3 不同事故与非事故比例下的混淆矩阵

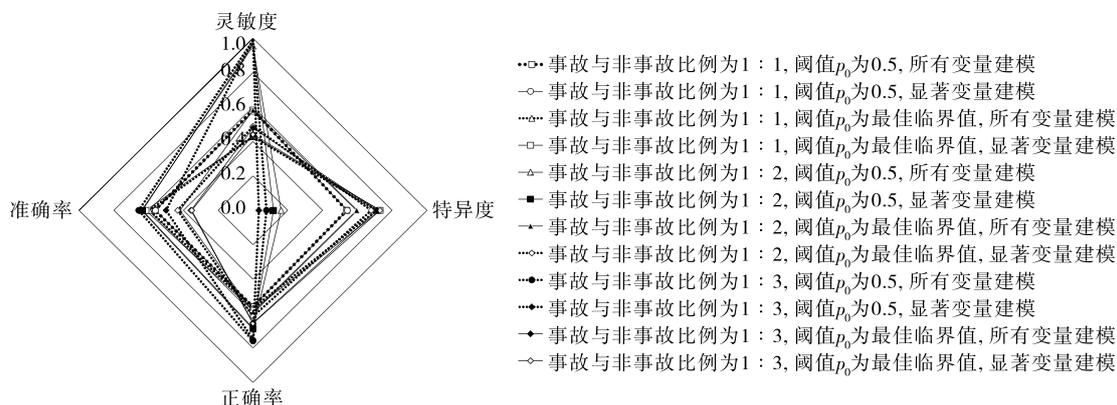


图 4 不同事故与非事故比例数据集建模结果

由图 4 可知:在事故与非事故比例为 1:1、阈值为 0.5 时,4 个指标较平均;事故与非事故比例为 1:2 时,4 个指标分布不均匀,使用显著变量建模、阈值为 0.5 时,模型的灵敏度、准确率、正确率最大,特异度最小;事故与非事故比例为 1:3 时,4 个指标分布不均匀,使用显著变量建模、阈值为 0.5 时,模型的灵敏度、正确率最大,特异度最小。综上,使用显著变量建模、阈值为 0.5、事故与非事故比例为 1:1 时,模型的预测效果最佳。

#### 4 结论

(1) 季节变化带来的客观环境差异并不足以导致交通事故数明显的季度波动;交通事故季度分布特征为四季度事故高发,主要原因是出行量骤增;交通事故月分布特征为 4 月、5 月、8 月和 10 月交通事故高发,主要受雨季和节假日出行高峰的影响。

(2) 交通事故空间分布特征为在连续下坡、连续下坡+隧道路段等道路线形组合不良区域因车辆刹车失灵、制动不及时而引发追尾、碰撞等交通事故。

(3) 提取事故前兆特征,采用配对式病例一对照法对事故数据和非事故数据进行匹配处理,基于实时交通流数据、事故数据构建二元 Logistic 回归模型,结果表明使用显著变量建模、阈值为 0.5、事故与非事故比例为 1:1 时,模型的预测效果最佳。

本文构建的山区高速公路交通事故动态风险评估模型参数仅包含交通流指标,后续可考虑加入高精度的气象数据,进一步提升模型预测准确率。

#### 参考文献:

[1] 席建锋,王晓燕,王双维,等.基于粗糙集的道路交通事

故成因层次分析方法[J].长春理工大学学报(自然科学版),2009,32(2):257-259.

[2] 段志宏,文元勇,苏宇,等.不利天气下山区高速公路交通流特性分析[J].交通节能与环保,2023,19(2):136-140.

[3] 马筱栋,樊博.降雨对山区高速公路运行车速的影响研究[J].公路与汽运,2020(5):33-36.

[4] 田毕江,梁超,鲍彦莅,等.山区高速公路交通事故时空分布特性与安全改善对策[J].武汉理工大学学报(交通科学与工程版),2018,42(6):1014-1018.

[5] 何洋.山区高速公路施工便道路线平纵设计[J].工程建设与设计,2023(3):86-88.

[6] 马瑾,李薇,罗忠祥,等.暴雨团雾多发公路交通工程设施设置技术[J].交通节能与环保,2022,18(4):148-151.

[7] 邓星月.山区高速公路弯道路段行驶稳定性分析及速度管理策略[D].重庆:重庆交通大学,2020.

[8] HOSSAIN M, ABDEL-ATY M, QUDDUS M A, et al. Real-time crash prediction models: State-of-the-art, design pathways and ubiquitous requirements[J].Accident Analysis & Prevention,2019,124:66-84.

[9] JIMÉNEZ-GAMERO M D, ALBA-FERNÁNDEZ M V. A test for the geometric distribution based on linear regression of order statistics[J].Mathematics and Computers in Simulation,2021,186:103-123.

[10] 张园笛.车流量与占有率算法模块及其在“平交柔管系统”的应用[D].南京:南京大学,2019.

[11] 孟云伟,张熙衍,青光焱,等.基于 Logistic 回归的高速公路交通事故后果的影响因素分析[J].武汉理工大学学报(交通科学与工程版),2022,46(1):12-16.

[12] 褚端峰,吴超仲,李顺喜,等.基于 Logistic 回归的高速公路车-车碰撞事故深度分析[J].中国安全科学学报,2014,24(3):103-108.

收稿日期:2023-06-12