第12卷 第2期 交通运输工程学报 Vol.12 No.2
2012年4月 Journal of Traffic and Transportation Engineering Apr. 2012

文章编号:1671-1637(2012)02-0093-07

高速公路交通事故起数时空分析模型

马壮林[1],邵春福[2],胡大伟[1],马社强[3]

(1. 长安大学 汽车学院,陕西 西安 710064;2. 北京交通大学 城市交通复杂系统理论与技术教育部重点实验室,北京 100044;3. 中国人民公安大学 交通管理工程系,北京 102623)

摘 要:为了分析交通事故起数与时间、道路空间结构及交通运行环境等潜在影响因素之间的关系,从时间和空间角度选择9个自变量,分别从路段长度一致和路段坡度一致2个角度,构建交通事故起数时段、周日和月分布模型。以某典型交通事故多发段为例,分别运用泊松回归模型、负二项回归模型、零堆积泊松回归模型和零堆积负二项回归模型拟合交通事故起数时段、周日和月分布模型,根据模型的拟合优度检验,分别确定3个模型的最佳形式,从而构建交通事故起数时空分析模型。研究结果表明:从AIC准则和BIC准则来看,基于路段长度一致的交通事故起数时段、月分布模型采用负二项回归模型拟合效果较好,其他模型选择泊松回归模型拟合效果较好;基于路段长度一致的交通事故起数时段、周日、月分布模型的预测误差小于基于路段坡度一致的交通事故起数时段、周日、月分布模型。

关键词:交通工程;交通事故起数;时空分析模型;泊松回归模型;负二项回归模型;零堆积泊松回归模型;零堆积负二项回归模型

中图分类号:U491.31 **文献标志码:**A

Temporal-spatial analysis model of traffic accident frequency on expressway

MA Zhuang-lin[1], SHAO Chun-fu[2], HU Da-wei[1], MA She-qiang[3]

(1. School of Automobile, Chang'an University, Xi'an 710064, Shaanxi, China; 2. MOE Key Laboratory for Urban Transportation Complex Systems Theory and Technology, Beijing Jiaotong University, Beijing 100044, China; 3. Department of Traffic Management Engineering, Chinese People's Public Security University, Beijing 102623, China)

Abstract: In order to analyze the relationships among traffic accident frequency and potential influencing factors such as time, road space structure and traffic running environment, nine independent variables were selected from the aspects of time and space, two kinds of section divided methods were adopted, which were fixed-length consistent segment and longitudinal grade consistent segment, and the hourly, weekly and monthly distribution models of traffic accident frequency were constructed. A typical accident-prone section was selected, and Poisson regression model, negative binomial regression model, zero-inflated Poisson regression model and zero-inflated negative binomial regression model were used to fit hourly, weekly and monthly distribution models respectively. The best forms of three models were determined, and the temporal-spatial analysis model of traffic accident frequency was established based on the goodness of fit test. Analysis result shows that the fitting effect of negative binomial regression model is better for traffic accident hourly and monthly distribution models based on fixed-length

收稿日期:2011-11-18

基金项目:"十一五"国家科技支撑计划项目(2007BAK35B06);高等学校博士学科点专项科研基金项目(20090009110011);中央高校基本科研业务费专项资金项目(CHD2011JC112,CHD2011JC002,CHD2011JC129)

作者简介:马壮林(1980-),男,辽宁鞍山人,长安大学讲师,工学博士,从事交通安全研究。

consistent segment from the views of AIC and BIC, and the fitting effect of Poisson regression model is better for other models. The prediction errors of traffic accident hourly, weekly and monthly distribution model based on fixed-length consistent segment are less than those of longitudinal grade consistent segment. 4 tabs, 15 refs.

Key words: traffic engineering; traffic accident frequency; temporal-spatial analysis model; Poisson regression model; negative binomial regression model; zero-inflated Poisson regression model; zero-inflated negative binomial regression model

Author resume: MA Zhuang-lin(1980-), male, lecturer, PhD, +86-29-82334425, mazhuanglin@126.com.

0 引 言

中国高速公路的快速发展提高了中国公路网的整体技术水平,优化了交通运输结构,对缓解交通运输的"瓶颈"制约发挥了重要作用,有力地促进了中国经济发展和社会进步,但同时也以它的高事故率、高伤亡率困扰着中国高速公路的管理者和使用者。2010年,中国高速公路里程仅占公路总里程的1.85%,而高速公路交通事故起数、死亡人数和受伤人数分别占总的公路交通事故相应数据的7.76%、13.54%和9.22%;高速公路百公里事故率、百公里死亡率和百公里受伤率分别为13.09、8.50和18.54,分别是普通公路的4.47、8.31、5.39倍[1]。国外研究表明,发达国家高速公路交通事故率约为普通公路的1/10,事故死亡率约为普通公路的1/3[2]。由此可见,中国高速公路交通安全形势不容乐观,高速公路交通事故预防研究具有很大潜力。

随着西部大开发战略的实施,山区高速公路的里程越来越多,所占比例越来越大。由于山岭重丘区地形地貌的限制,山区高速公路部分路段的线形设计不可避免地采用极限值或高于极限值,导致部分路段事故频繁发生,进而形成事故多发段。但是,目前缺乏对事故多发段进行系统、深入的理论研究,不仅没有对事故多发段的定义形成统一标准,而且没有对事故多发段的产生机理和改善研究进行详细剖析,因此,以典型事故多发段为研究对象,分析交通事故特征与道路空间结构、交通运行环境等因素之间的关系,剖析交通事故的形成机理,基于事故产生机理提出交通事故预防措施并预测其预防效果是目前交通安全领域研究迫切需要解决的问题之一。

李铁洪等分析了小半径曲线路段的交通事故成因,并提出了相应的预防措施[3];张生瑞等结合高速公路隧道交通事故的成因,提出了高速公路隧道群交通事故预防对策[4];陈斌等分析了急弯陡坡、连续陡坡和连续缓坡3种线形组合形式路段的交通事故特征和事故原因[5]。但是,这些研究没有具体分析交通事故与时间、空间因素之间的量化关系,无法分析某一影响因素的变动引起交通事故的变化程度,进而无法分析交通事故预防措施的预防效果。

由于交通事故起数具有离散、独立和稀少的特性,因此,一些学者提出运用泊松回归模型来分析交通事故与影响因素之间的关系[6-7]。泊松回归模型要求均值和方差相等,事实上,样本的方差往往大于均值,为了消除和减少这种不利影响,一些学者提出采用负二项回归模型来代替泊松回归模型[8-9]。当统计间隔较小或者道路使用率较低时,导致某些路段的交通事故起数为0,为了处理过多的零值现象,一些学者提出采用零堆积概率模型来代替泊松回归模型和负二项回归模型[10-12]。

本文以某典型交通事故多发段为研究对象,分别运用泊松回归模型(Poisson)、负二项回归模型(NB)、零堆积泊松回归模型(ZIP)和零堆积负二项回归模型(ZINB)建立交通事故起数时段、周日和月分布模型,根据模型的拟合优度检验,确定3个模型的最佳形式,从而构建交通事故起数时空分析模型,为制定交通事故预防措施提供理论依据。

1 交通安全影响因素确定

从理论上讲,建立交通事故起数时空分析模型考虑的影响因素越多越好,考虑的因素越多,模型的拟合效果可能越好。但是,实际上任何一种模型都不可能涵盖所有的影响因素,并且考虑的因素越多,用来建模的样本量就越少,建立的模型精度也大大降低;此外,如果考虑的因素过多,就会出现模型可移植性差,不利于实际应用等问题,因此,根据一般性、简单性和实用性原则,本文初步考虑了可能影响交通事故发生的9个因素,即曲线比例、曲度、曲率变化率、坡度、坡度累积效应、弯坡组合、特殊路段、交通

量和大型车混入率。其中，交通量和大型车混入率体现时间特性，其余 7 个影响因素体现空间特性。

2　时空分析模型

2.1　模型变量选择

选择交通事故起数作为因变量，从时间角度和空间角度选择 9 个自变量，其中，时间角度的自变量包括交通量和大型车混入率，空间角度的自变量包括曲线比例、曲度、曲率变化率、坡度、坡度累积效应、弯坡组合、特殊路段。

2.1.1　曲线比例

曲线比例 C_i 为选定路段平面曲线所占比例，即

$$C_i = \frac{\sum_{j=1}^{k'} L_{cij}}{L_i} \tag{1}$$

式中：L_{cij} 为第 i 个路段第 j 个曲线的长度(m)；k' 为第 i 个路段的曲线数量；L_i 为第 i 个路段长度(m)。

2.1.2　曲度

第 j 个曲线的曲度 D_j 为

$$D_j = 1\ 748/R_j \tag{2}$$

式中：R_j 为第 j 个曲线的半径(m)。

2.1.3　曲率变化率

曲率变化率 r 为平面线形中角度改变的绝对值与路线长度的比值[13]，即

$$r = \frac{90\ 000(L_{cl1} + 2L_{cr} + L_{cl2})}{\pi R(L_{cl1} + L_{cr} + L_{cl2})} \tag{3}$$

式中：L_{cr} 为圆曲线长度(m)；L_{cl1}、L_{cl2} 分别为第 1、2 条缓和曲线长度(m)；R 为圆曲线半径(m)。

如果第 i 个路段不只包含 1 条曲线，那么第 i 个路段的曲率变化率 r_i 为

$$r_i = \frac{\sum_{j=1}^{k'} r_{ij} L_{cij}}{\sum_{j=1}^{k'} L_{cij}} \tag{4}$$

式中：r_{ij} 为第 i 个路段第 j 个曲线的曲率变化率[(°)·km^{-1}]。

2.1.4　坡度

当路段的坡度一致时，用该路段的坡度作为自变量；当路段的坡度不一致时，用该路段的平均坡度作为自变量。

平均坡度指路段不同坡度的加权平均值，第 i 个路段的平均坡度 A_{Gi} 为

$$A_{Gi} = \frac{\sum_{g=1}^{q} G_{ig} L_{ig}}{L_i} \tag{5}$$

式中：G_{ig}、L_{ig} 分别为第 i 个路段第 g 个纵坡的坡度(%)和坡长(m)；q 为第 i 个路段的纵坡数量。

2.1.5　坡度累积效应

为了更好地描述连续下坡路段对交通事故的影响，本文提出连续下坡的坡度累积效应，其定义如下。

首先，计算第 l 个下坡的坡长和坡度的综合作用 E_l 为

$$E_l = \frac{G_l}{L_l} + G_l L_l \tag{6}$$

式中：G_l、L_l 分别为第 l 个下坡的坡度(%)和坡长(km)。

其次，假设相邻纵坡路段的影响呈指数递增关系，则第 $l-1$ 个下坡对第 l 个下坡的坡度影响系数 F_l 为

$$F_l = (1 + |G_l - G_{l-1}|)^{E_l} \tag{7}$$

对于第 1 个下坡路段而言，其坡度影响系数 F_1 为

$$F_1 = (1 + |G_1|)^{E_1} \tag{8}$$

假设连续下坡路段长度超过 5 km 将产生坡度累积效应，则第 l 个下坡的坡度累积效应 C_{Dl} 为

$$C_{Dl} = \begin{cases} F_l & \sum L_l \leqslant 5\ \text{km} \\ F_l F_{l-1} & \sum L_l > 5\ \text{km} \end{cases} \tag{9}$$

同理，第 i 个路段的坡度累积效应 C_{Di} 为

$$C_{Di} = \begin{cases} C_{D(l-1)} & K_{l-1} \leqslant K_i < K_{i+1} \leqslant K_l \\ \dfrac{1}{L_i}[(K_l - K_{l-1})C_{D(l-1)} + (K_{l+1} - K_l)C_{Dl} + \cdots + \\ \quad (K_{l+t-1} - K_{l+t-2})C_{D(l+t-2)} + (K_{l+t} - K_{l+t-1})C_{D(l+t-1)}] & K_{l-1} \leqslant K_i < K_l < K_{l+t-1} < K_{i+1} \leqslant K_{l+t} \end{cases} \tag{10}$$

式中：K_i 为第 i 个路段的起点桩号，即第 $i-1$ 个路段的终点桩号；K_l 为第 l 个下坡的起点桩号，即第 $l-1$ 个下坡的终点桩号；t 为第 i 个路段的下坡数量。

2.1.6　弯坡组合

弯坡组合综合考虑了纵坡坡度和平曲线半径对交通安全的综合影响，第 i 个路段的弯坡组合 C_{Bi} 为

$$C_{Bi} = \max\left(\left| \frac{G_{iju}}{R_{ij}} \right| \right) \quad u = 1, 2, \cdots, s \tag{11}$$

式中：G_{iju} 为第 i 个路段第 j 个曲线第 u 个纵坡的坡度(%)；R_{ij} 为第 i 个路段第 j 个曲线的平曲线半径(km)；s 为第 i 个路段第 j 个曲线的纵坡数量。

2.1.7 特殊路段

特殊路段是一个二元变量，如果第 i 个路段包括桥梁、隧道等特殊构造物，特殊路段的值取 1，反之，特殊路段的值取 0。

2.1.8 交通量

交通量指单位时间内通过道路指定地点或断面的车辆数，根据单位时间的不同，可以分为小时交通量、日交通量、月交通量等。

2.1.9 大型车混入率

大型车混入率指单位时间内，在交通组成中大型车所占的比例，根据单位时间的不同，可分为小时大型车混入率、日大型车混入率、月大型车混入率。

2.2 路段长度划分

路段长度划分是建立高速公路交通事故起数时空分析模型的先决条件，直接影响到模型拟合精度以及模型的应用效果，通常路段长度划分有定长法和不定长法[14-15]。

鉴于定长法和不定长法都有一定程度的不足，暂时无法判断哪种路段划分方法更适合本文的研究，因此，本文运用这两种路段划分方法分别建立交通事故起数时空分析模型，然后根据构建的分析模型比较这 2 种方法的适用性。

2.3 模型构建

由于交通事故具有随时间而变化的特征，本文分别从交通事故起数的时段变化、周日变化和月变化进行分析，分别建立交通事故起数时段分布模型、周日分布模型和月分布模型。

2.3.1 交通事故起数时段分布模型

受人们生活规律的影响，交通量在 1 d 的 24 h 中各不相同，存在高峰小时与非高峰小时之分，同时交通事故的分布随时段不同有着明显的差异，因此，建立的交通事故起数时段分布模型为

$$\lambda_{ih} = x_1 \exp\Big[\sum_{k=1}^{n}(a_0 + a_k x_{ihk})\Big] \tag{12}$$

$$x_1 = \frac{L_i H_{ih} T}{10^6} \tag{13}$$

式中：λ_{ih} 为第 i 个路段第 h 个时段的交通事故起数；x_1 为时段道路使用度；x_{ihk} 为第 i 个路段第 h 个时段的第 k 个自变量；a_0、a_k 为模型参数；n 为自变量个数；H_{ih} 为第 i 个路段第 h 个时段的交通量；T 为统计时间。

2.3.2 交通事故起数周日分布模型

受人们生活规律的影响，道路上的交通量在一个星期内是不相同的，相应地交通事故周日分布的情况受出行次数的影响而有所不同，因此，建立的交通事故起数周日分布模型为

$$\beta_{iw} = x_2 \exp\Big[\sum_{k=1}^{n}(a_0 + a_k x_{iwk})\Big] \tag{14}$$

$$x_2 = \frac{L_i W_{iw} T}{10^6} \tag{15}$$

式中：β_{iw} 为第 i 个路段第 w 个周日的交通事故起数；x_2 为周日道路使用度；x_{iwk} 为第 i 个路段第 w 个周日的第 k 个自变量；W_{iw} 为第 i 个路段第 w 个周日的交通量。

2.3.3 交通事故起数月分布模型

交通事故月分布的情况受气候条件、不同月份交通情况等因素的影响，因此，建立的交通事故起数月分布模型为

$$\gamma_{im} = x_3 \exp\Big[\sum_{k=1}^{n}(a_0 + a_k x_{imk})\Big] \tag{16}$$

$$x_3 = \frac{L_i M_{im} T}{10^6} \tag{17}$$

式中：γ_{im} 为第 i 个路段第 m 个月的交通事故起数；x_3 为月道路使用度；x_{imk} 为第 i 个路段第 m 个月的第 k 个自变量；M_{im} 为第 i 个路段第 m 个月的交通量。

2.4 模型检验

2.4.1 回归系数的显著性检验

回归系数的显著性检验指检验每个自变量对因变量的影响是否显著，剔除对因变量影响不显著的自变量，保留对因变量影响显著的自变量。其基本步骤为：给定一个显著性水平 α，根据 Wald 统计量判断自变量的显著程度，如果显著性概率 p 值小于显著性水平 α，表明该自变量与因变量显著相关，选择该自变量，反之，表明该自变量与因变量显著不相关，剔除该自变量。

2.4.2 模型的拟合优度检验

本文选择似然比检验、AIC 准则和 BIC 准则 3 个拟合优度检验指标来评价模型的拟合效果。

似然比检验统计量 d 是 2 个带有不同参数模型的对数似然函数值差的 2 倍，其计算式为

$$d = -2[\ln(\hat{L}_s) - \ln(\hat{L}_f)] \tag{18}$$

式中：$\hat{L}_s$ 为提出模型的最大似然估计值；$\hat{L}_f$ 为完全模型或饱和模型的最大似然估计值。

AIC(Akaike's Information Criterion)准则由 Akaike 提出，其计算值 A 为

$$A = -2\lg(L) + 2h' \tag{19}$$

式中：L 为拟合模型的最大似然值；h' 为模型中参数的数量。在回归分析建模过程中，A 的值越小，其模

型拟合的越好。

BIC（Bayesian Information Criterion）准则由 Schwardz 提出，其计算值 B 为

$$B = -2\ln(L) + h'\ln(N) \tag{20}$$

式中：N 为模型中数据的数量。在回归分析建模过程中，B 的值越小，其模型拟合的越好。

3　实证分析

3.1　数据来源

本文以某典型交通事故多发段为研究对象，自 2003 年4 月 3 日通车以来，交通事故频繁发生，截至 2004 年 12 月 31 日，该路段共发生 332 起交通事故，其中死亡事故 34 起，受伤事故 60 起，其他事故 238 起。

本文分别采用定长法和不定长法划分路段长度。对于定长法，13 km 长的路段被划分成 13 个定长为 1 km 的路段，其中单个路段平均最大纵坡为 −4.02%，平均最小纵坡为 −1.25%，整个路段平均纵坡为 −2.93%。对于不定长法，13 km 长的路段被划分成 16 个坡度一致的路段，其中最大纵坡和坡长分别为 −5% 和 1.2 km，最小纵坡和坡长分别为 −0.65% 和 0.5 km，平均纵坡和坡长分别为 −3.06%和0.81 km。

3.2　交通事故起数模型

由于交通事故的偶然性和稀少性，以 1 h 为统计周期，交通事故起数的离散性很大，因此，根据交通事故起数时段分布特征，将 24 h 分成 6 个时段，即0～4、4～8、8～12、12～16、16～20 和 20～24。将事故数据按长度一致和时段进行统计，得到 78 个不同路段和时段的交通事故起数统计样本（表 1），从而根据不同路段和时段交通事故起数与时间、道路空间结构和交通运行环境等因素之间的关系，构建基于路段长度一致的交通事故起数时段分布模型。

首先，分别采用 Poisson、NB、ZIP 和 ZINB 模型对候选自变量进行回归分析，建立完全模型；然后，采用混合逐步选择法，取显著性水平 α 为 0.05，针对显著性水平 α 大于 0.05 的变量，每次剔除 1 个最不符合的变量，同时每次检查已经剔除的变量，看是否仍存在显著性水平 α 小于 0.05 的变量没有被纳入模型中，直到模型中不再包含不符合判别条件的自变量，且已经剔除的变量中不存在著性水平 α 小于 0.05 的变量为止，从而得到 4 种分布模型的交通事故起数时空分析模型，其拟合优度指标见表 2。由表 2 可以得出如下结论。

表 1　不同路段和时段的交通事故起数

Tab. 1　Traffic accident frequencies of different segments and hours

路段	0～4	4～8	8～12	12～16	16～20	20～24
K39＋180～K40＋180	0	1	0	0	2	0
K40＋180～K41＋180	0	1	0	2	3	0
K41＋180～K42＋180	0	0	1	1	0	2
K42＋180～K43＋180	1	0	0	0	8	1
K43＋180～K44＋180	3	4	1	1	0	4
K44＋180～K45＋180	1	0	1	0	1	2
K45＋180～K46＋180	0	0	2	0	0	0
K46＋180～K47＋180	0	1	1	2	2	1
K47＋180～K48＋180	1	2	4	5	2	2
K48＋180～K49＋180	4	4	12	7	3	4
K49＋180～K50＋180	6	8	7	8	12	6
K50＋180～K51＋180	15	16	13	13	10	9
K51＋180～K52＋180	10	26	22	22	16	13
合计	41	63	64	61	59	44

表 2　路段长度一致的时段分布模型拟合优度

Tab. 2　Goodness of fit for hourly distribution model based on fixed-length segment

模型	Vuong 检验值	A	B	似然比检验值
Poisson		4.00	−18.19	
NB		3.93	−21.46	7.63
ZIP	0.00	4.03	−13.83	
ZINB	−0.01	3.96	−17.10	

（1）ZIP 模型和 ZINB 模型的 Vuong 检验值分别为 0、−0.01，介于 −1.96 和 1.96 之间，因此，不能判断 Poisson 模型和 ZIP 模型哪个更优，以及 NB 模型和 ZINB 模型哪个更优。

（2）从 A、B 的指标值来看，NB 模型的值最小，因此，NB 模型的回归拟合效果较好。

（3）NB 模型的似然比检验值为 7.63，远远大于 NB 模型中过离散系数 K 为 0 假设的临界值，即表示 K 值不可能为 0，反映出 NB 模型要优于 Poisson 模型。

综上所述，选择 NB 形式的回归模型最优。在

路段长度一致的条件,第 i 个路段第 h 个时段的交通事故起数时段分布模型为

$$\lambda_{ih}^{1} = x_{1}^{1}\exp(-0.46 + 0.62C_{\mathrm{D}ih} + 0.18C_{\mathrm{B}ih} - 1.01L_{\mathrm{P}ih}) \quad (21)$$

式中:λ_{ih}^{1} 为路段长度一致的第 i 个路段第 h 个时段交通事故起数;x_{1}^{1} 为路段长度一致的时段道路使用度;$C_{\mathrm{D}ih}$ 为第 i 个路段第 h 个时段的坡度累积效应;$C_{\mathrm{B}ih}$ 为第 i 个路段第 h 个时段的弯坡组合;$L_{\mathrm{P}ih}$ 为第 i 个路段第 h 个时段的大型车混入率。

由式(21)可以看出,路段长度一致的交通事故起数时段分布与坡度累积效应和弯坡组合呈正相关,与大型车混入率呈负相关。大型车混入率对交通事故起数时段分布的影响最大,其次是坡度累积效应和弯坡组合。

同理,可以得到 2 种路段长度划分方法下的交通事故起数模型,见表 3,表中上标带 1、2 变量分别为长度一致、坡度一致下对应变量,由表 3 可得出如下结论。

表 3 交通事故起数模型

Tab. 3 Traffic accident frequency models

模型类型	路段划分	拟合模型	模型形式
时段分布模型	长度一致	NB	$\lambda_{ih}^{1}=x_{1}^{1}\exp(-0.46+0.62C_{\mathrm{D}ih}+0.18C_{\mathrm{B}ih}-1.01L_{\mathrm{P}ih})$
	坡度一致	Poisson	$\lambda_{ih}^{2}=x_{1}^{2}\exp(0.16-0.88C_{ih}+0.36D_{ih}+0.01r_{ih}+0.47C_{\mathrm{D}ih}-0.92L_{\mathrm{P}ih})$
周日分布模型	长度一致	Poisson	$\beta_{iw}^{1}=x_{2}^{1}\exp(5.52+0.63C_{\mathrm{D}iw}+0.18C_{\mathrm{B}iw}-8.52L_{\mathrm{P}iw})$
	坡度一致	Poisson	$\beta_{iw}^{2}=x_{2}^{2}\exp(5.8-C_{iw}+0.42D_{iw}+0.01r_{iw}+0.43C_{\mathrm{D}iw}-7.9L_{\mathrm{P}iw})$
月分布模型	长度一致	NB	$\gamma_{im}^{1}=x_{3}^{1}\exp(-4.05+0.62C_{\mathrm{D}im}+0.19C_{\mathrm{B}im}+3.83L_{\mathrm{P}im})$
	坡度一致	Poisson	$\gamma_{im}^{2}=x_{3}^{2}\exp(-2.96-0.92C_{im}+0.39D_{im}+0.01r_{im}+0.45C_{\mathrm{D}im}+3.34L_{\mathrm{P}im})$

(1)路段坡度一致的交通事故起数时段分布模型选择 Poisson 形式的回归模型最优;路段坡度一致的交通事故起数时段分布 λ_{ih}^{2} 与曲度 D_{ih}、曲率变化率 r_{ih} 和坡度累积效应 $C_{\mathrm{D}ih}$ 呈正相关,与曲线比例 C_{ih} 和大型车混入率 $L_{\mathrm{P}ih}$ 呈负相关;大型车混入率对交通事故起数时段分布的影响最大,其次是曲线比例。

(2)路段长度一致的交通事故起数周日分布模型选择 Poisson 形式的回归模型最优;路段长度一致的交通事故起数周日分布 β_{iw}^{1} 与坡度累积效应 $C_{\mathrm{D}iw}$ 和弯坡组合 $C_{\mathrm{B}iw}$ 呈正相关,与大型车混入率 $L_{\mathrm{P}iw}$ 呈负相关;大型车混入率对交通事故起数周日分布的影响最大,其次是坡度累积效应。

(3)路段坡度一致的交通事故起数周日分布模型选择 Poisson 形式的回归模型最优;路段坡度一致的交通事故起数周日分布 β_{iw}^{2} 与曲度 D_{iw}、曲率变化率 r_{iw} 和坡度累积效应 $C_{\mathrm{D}iw}$ 呈正相关,与曲线比例 C_{iw} 和大型车混入率 $L_{\mathrm{P}iw}$ 呈负相关;大型车混入率对交通事故起数周日分布的影响最大,其次是曲线比例。

(4)路段长度一致的交通事故起数月分布模型选择 NB 形式的回归模型最优;路段长度一致的交通事故起数月分布 γ_{im}^{1} 与坡度累积效应 $C_{\mathrm{D}im}$、弯坡组合 $C_{\mathrm{B}im}$ 和大型车混入率 $L_{\mathrm{P}im}$ 呈正相关;大型车混入率对交通事故起数周日分布的影响最大,其次是坡度累积效应。

(5)路段坡度一致的交通事故起数月分布模型选择 Poisson 形式的回归模型最优;路段坡度一致的交通事故起数月分布 γ_{im}^{2} 与曲度 D_{im}、曲率变化率 r_{im}、坡度累积效应 $C_{\mathrm{D}im}$ 和大型车混入率 $L_{\mathrm{P}im}$ 呈正相关,与曲线比例 C_{im} 呈负相关;大型车混入率对交通事故起数周日分布的影响最大,其次是曲线比例。

3.3 对比分析

不同路段长度划分方法、不同时间分布特征的交通事故起数模型的预测效果见表 4,从表 4 可以看出:基于路段长度一致的交通事故起数时段、周日、月分布模型的预测误差小于基于路段坡度一致的交通事故起数时段、周日、月分布模型的误差;基于路段长度一致的交通事故起数时段、周日分布模型的预测值低于实际值,其他模型的预测值高于实际值。由此可见,长度一致的路段划分方法更适合于本文的研究。

表 4 交通事故起数模型对比

Tab. 4 Comparison of traffic accident frequency models

类型	路段长度划分	预测误差/%
时段分布模型	长度一致	−3.01
	坡度一致	6.63
周日分布模型	长度一致	−1.20
	坡度一致	5.72
月分布模型	长度一致	3.31
	坡度一致	6.63

4 结　语

(1)以交通事故起数为因变量，以曲线比例、曲度、曲率变化率、纵坡、坡度累积效应、弯坡组合、特殊路段、交通量和大型车混入率为侯选自变量，从路段长度一致和路段坡度一致 2 个角度，分别运用泊松回归模型、负二项回归模型、零堆积泊松回归模型和零堆积负二项回归模型构建了交通事故起数时段、周日和月分布模型。

(2)路段长度一致的交通事故起数时段、月分布模型选择 NB 形式的回归模型最优；路段坡度一致的交通事故起数时段、月分布模型选择 Poisson 形式的回归模型最优；路段长度一致、坡度一致的周日分布模型选择 Poisson 形式的回归模型最优。

(3)基于路段长度一致的交通事故起数时段、周日、月分布模型与坡度累积效应、弯坡组合和大型车混入率显著相关，基于路段坡度一致的交通事故起数时段、周日、月分布模型与曲线比例、曲度、曲率变化率、坡度累积效应和大型车混入率显著相关。

(4)基于路段长度一致的交通事故起数时段、周日、月分布模型的预测误差小于基于路段坡度一致的交通事故起数时段、周日、月分布模型，因此，长度一致的路段划分方法更适合于本文的研究。

参 考 文 献：

References：

[1] 公安部交通管理局．中华人民共和国道路交通事故统计年报(2010 年度)[R]．无锡：公安部交通管理科学研究所，2011．
Traffic Management Bureau of the Ministry of Public Security. Road traffic accident statistics annual report of the People's Republic of China(2010)[R]. Wuxi: Traffic Management Research Institute of the Ministry of Public Security, 2011. (in Chinese)

[2] 李相勇，张　南，张学尽，等．高速公路交通事故致因的人机工程学分析[J]．公路，2003(8)：108-112．
LI Xiang-yong, ZHANG Nan, ZHANG Xue-jin, et al. Analysis of factors affecting expressway traffic accidents based on ergonomics[J]. Highway, 2003(8): 108-112. (in Chinese)

[3] 李铁洪，吴华金．长直线接小半径曲线公路交通事故成因及预防对策[J]．中国公路学报，2007，20(1)：35-40．
LI Tie-hong, WU Hua-jin. Causes and countermeasures of highway traffic accidents in long straight line combined with sharp curve[J]. China Journal of Highway and Transport, 2007, 20(1): 35-40. (in Chinese)

[4] 张生瑞，马壮林，石　强．高速公路隧道群交通事故分布特点及预防对策[J]．长安大学学报：自然科学版，2007，27(1)：63-66．
ZHANG Sheng-rui, MA Zhuang-lin, SHI Qiang. Distribution characteristics and countermeasures of traffic accidents in expressway tunnel group[J]. Journal of Chang'an University: Natural Science Edition, 2007, 27(1): 63-66. (in Chinese)

[5] 陈　斌，袁　伟，付　锐，等．连续长大下坡路段交通事故特征分析[J]．交通运输工程学报，2009，9(4)：75-78，84．
CHEN Bin, YUAN Wei, FU Rui, et al. Analysis of traffic accident characteristic on continuous long downgrade section[J]. Journal of Traffic and Transportation Engineering, 2009, 9(4): 75-78, 84. (in Chinese)

[6] MIAOU S P, LUM H. Modeling vehicle accidents and highway geometric design relationships[J]. Accident Analysis and Prevention, 1993, 25(6): 689-709.

[7] KRAUS J F, ANDERSON C L, ARZEMANIAN S, et al. Epidemiological aspects of fatal and severe injury urban freeway crashes[J]. Accident Analysis and Prevention, 1993, 25(3): 229-239.

[8] HAUER E, NG J C N, LOVELL J. Estimation of safety at signalized intersections (with discussion and closure)[J]. Transportation Research Record, 1988(1185): 48-61.

[9] HINDE J, DEMETRIO C G B. Overdispersion: models and estimation[J]. Computational Statistics and Data Analysis, 1998, 27(2): 151-170.

[10] LEE J, MANNERING F. Impact of roadside features on the frequency and severity of run-off-roadway accidents: an empirical analysis[J]. Accident Analysis and Prevention, 2002, 34(2): 149-161.

[11] KUMARA S S P, CHIN H C. Modeling accident occurrence at signalized tee intersections with special emphasis on excess zeros[J]. Traffic Injury Prevention, 2003, 4(1): 53-57.

[12] LORD D, WASHINGTON S P, IVAN J N. Poisson, Poisson-gamma and zero-inflated regression models of motor vehicle crashes: balancing statistical fit and theory[J]. Accident Analysis and Prevention, 2005, 37(1): 35-46.

[13] PERCO P. Influence of the general character of horizontal alignment on operating speed of two-lane rural roads[J]. Transportation Research Record, 2008(2075): 16-23.

[14] 罗石贵，周　伟．道路交通事故多发点鉴定方法探讨[J]．西安公路交通大学学报，1999，19(1)：30-33．
LUO Shi-gui, ZHOU Wei. Research on the identification of road accident blackspots[J]. Journal of Xi'an Highway University, 1999, 19(1): 30-33. (in Chinese)

[15] HAUER E, KONONOV J, ALLERY B, et al. Screening the road network for sites with promise[J]. Transportation Research Record, 2002(1784): 27-32.