

文章编号:1673-2049(2011)02-0008-06

极值分析方法在车辆荷载评估中的应用与比较

韩大建, 潘 玲

(华南理工大学 土木与交通学院, 广东 广州 510640)

摘要:针对现有车辆荷载安全评估中采用拟合得到的近似分布作为极值估计的底分布而导致的极值估计结果误差较大的问题,引入了极值统计的新方法,即阈值模型、次序统计量模型、点过程模型,并分别采用这些模型对广州北环高速公路的 5 型车车质量样本进行统计分析,得到重车荷载的概率分布尾部及最大值分布,将新方法的结果与现行方法得到的结果进行了比较。结果表明:阈值模型和点过程模型与样本的一致性较好,现行方法得到的结果与样本点相差较大,且现行方法在一定程度上低估了车辆荷载的极值。

关键词:车辆荷载;极值估计;阈值模型;次序统计量模型;点过程模型

中图分类号:U441.2

文献标志码:A

Application and Comparison of Extreme Value Analysis Methods in Vehicle Load Assessment

HAN Da-jian, PAN Ling

(School of Civil Engineering and Transportation, South China University of Technology,
Guangzhou 510640, Guangdong, China)

Abstract: In accordance with the existing problems, that current analysis methods for vehicle loads using an approximate distribution obtained by fitting as a base distribution of extreme value estimation, which might cause a significant error for extreme values, new methods for extreme value statistics were introduced, including threshold model, order statistic measurement model and point process model. In order to obtain the probability distribution tail and the maximum value distribution of heavy truck loads, these new models were respectively adopted to analyze the samples of type 5 vehicle weight on Guangzhou North-ring Expressway. The comparisons between results from the new extreme value statistic models and those from the current method were carried out. The results show that the estimations from the threshold model and the point process model are more consistent with the measured samples, whereas the results from the current method show errors. Moreover, to a certain extent, the value of estimation load from the current method undervalues the extreme value of vehicle loads.

Key words: vehicle load; extreme value estimation; threshold model; order statistic measurement model; point process model

0 引 言

服役结构的安全性评估应该考虑实际的运营荷

载,如车辆荷载、风荷载等。现有的大多研究^[1-8]对包括车辆荷载在内的各种作用,均采用分析基准期内的某个极值分位数作为评估参考值,但是这些研

收稿日期:2011-03-01

基金项目:“十一五”国家科技支撑计划项目(2006BAJ01B07-05)

作者简介:韩大建(1940-),女,广西北海人,教授,博士研究生导师,工学博士,E-mail:djhan@scut.edu.cn。

究基本上都对荷载样本进行常用的几种分布类型的拟合并做假设检验,得到一个不被拒绝的分布作为底分布,再用极值统计的精确模型得到荷载最大值的高分位数,作为评估的指标。

这些研究都忽略了极值精确模型的前提条件,即底分布已知,而采用主观拟合得到的近似分布代替随机变量的真实分布,因此不能对荷载的极值进行较准确的分析。本文中采用 3 种新的极值统计模型,即阈值模型、次序统计量模型、点过程模型,利用广东广州北环高速公路的 5 型车车质量样本,对车辆荷载极值分析中存在的问题进行验证,同时为车辆荷载的极值分析选择较便捷、精确的模型。

1 精确模型

极值统计模型分为精确模型和渐进模型两类。如果随机变量 X 的底分布(或称为截口分布) $F(x)$ 已知,那么 X 的 N 次独立观测的最大值 M_N 的分布 $F_{M_N}(x)$ 可用式(1)精确求出

$$F_{M_N}(x) = [F(x)]^N \quad (1)$$

式(1)即为极值分布的精确模型^[9]。但在大多数的应用中,底分布是未知的,因此精确模型很难直接用于统计分析^[10]。

文献[2]中分别针对车辆一般运行状态和密集运行状态建立车辆荷载随机过程模型,推导出车辆荷载在分析基准期 T 内的最大值分布函数 $F_{QT}(x)$

$$F_{QT}(x) = \begin{cases} e^{-\lambda T[1-F(x)]} & \text{一般运行状态} \\ e^{-\lambda T^\beta[1-F(x)]} & \text{密集运行状态} \end{cases} \quad (2)$$

式中: λ 、 β 为随机过程参数; λT 、 λT^β 分别为分析基准期 T 内车辆一般运行和密集运行的平均出现次数,即 N 的期望值。

由此可见,采用现行方法对车辆荷载进行评估的关键在于其底分布 $F(x)$ 的确定。

2 极值分析的新方法

实际应用中,底分布 $F(x)$ 往往是未知的,式(1)很难直接用于统计分析,因此,研究极值的极限分布具有很重要的理论和实际意义。

基本定理:设 X_1, X_2, \dots, X_n 是独立同分布的随机变量序列, $M_n = \max\{X_{n,n}, X_{n-1,n}, \dots, X_{1,n}\}$, 如果存在常数序列 $\{a_n > 0\}$ 和 $\{b_n\}$, 使得

$$\lim_{n \rightarrow +\infty} P[(M_n - b_n)/a_n \leq x] = H(x) \quad x \in \mathbf{R} \quad (3)$$

成立,其中 $H(x)$ 为非退化的分布函数, $P(\cdot)$ 为概

率函数,那么 $H(x)$ 必是广义极值(GEV)分布函数,即有

$$H(x; \mu, \sigma, \xi) = \exp\left\{-[1 + \xi(x - \mu)/\sigma]^{-1/\xi}\right\} \quad \begin{cases} 1 + \xi(x - \mu)/\sigma > 0 \end{cases} \quad (4)$$

式中: μ 为位置参数, $\mu \in \mathbf{R}$; σ 为尺度参数, $\sigma > 0$; ξ 为形状参数, $\xi \in \mathbf{R}$ 。

式(4)就是极值的渐进模型。常数序列 $\{a_n > 0\}$ 和 $\{b_n\}$ 称作规范化常数, \tilde{M}_n 为规范化最大值变量, $\tilde{M}_n = (M_n - b_n)/a_n$ 。

由此可见,只有当样本为区组最大值时,即 $M_n = \max\{X_{n,n}, X_{n-1,n}, \dots, X_{1,n}\}$ 时,才可用 GEV 分布直接拟合样本,这需要大量的极值样本,在实际中是难以实现的。为了避免数据浪费,尽可能地利用样本中其他较大值所蕴含的极值信息^[11-12], 阈值模型、次序统计量模型、点过程模型等极值统计方法被陆续提出。

2.1 阈值模型

对基本定理中的 X_1, X_2, \dots, X_n , 如果式(3)成立,则对于足够大的阈值 u , 超出量 $Y = X - u$ 近似服从二参数的广义 Pareto(GP)分布

$$\begin{cases} G(y; \bar{\sigma}, \xi) = 1 - (1 + \xi y / \bar{\sigma})^{-1/\xi} \\ \bar{\sigma} = \sigma + \xi(u - \mu) \end{cases} \quad (5)$$

用 $\bar{F}(\cdot)$ 表示分布函数 $F(\cdot)$ 的尾部,即 $\bar{F}(\cdot) = 1 - F(\cdot)$ 。根据超出量分布的定义,有

$$\begin{aligned} \bar{F}_u(y) &= P(X - u \geq y \mid X > u) = \\ &= \frac{P(X \geq y + u)}{P(X > u)} = \frac{\bar{F}(y + u)}{\bar{F}(u)} \quad y \geq 0 \end{aligned} \quad (6)$$

即有

$$\bar{F}(y + u) = \bar{F}_u(y) \bar{F}(u)$$

$\bar{F}(u)$ 可以由 X 的经验分布函数估计得到,即 $\bar{F}(u) = N_u/n$, N_u 为超阈值样本的个数。 $\bar{F}_u(y)$ 可由式(5)中的 GP 分布近似表示,则有

$$F(x) = F(y + u) = 1 - \bar{F}(y + u) = 1 -$$

$$\bar{F}_u(y) \bar{F}(u) = 1 - \frac{N_u}{n} (1 + \xi \frac{y}{\sigma})^{-1/\xi} \quad (7)$$

X 的 p 分位数 x_p 为

$$x_p = F^{-1}(p) = u + \frac{\sigma}{\xi} \{[\frac{N_u}{n}(1 - p)]^{-\xi} - 1\} \quad (8)$$

2.2 次序统计量模型

将序列 X_1, X_2, \dots, X_n 按由小到大的次序排列成 $X_{n,n} \leq X_{n-1,n} \leq \dots \leq X_{1,n}$, 则称 $X_{r,n}$ 为第 r 大次序统计量。

对基本定理中的 X_1, X_2, \dots, X_n , 如果式(3)成

立,则规范化次序统计量 $X_{r,n}$ 收敛于 $H_r(x)$,即

$$\left. \begin{aligned} H_r(x) &= \exp[-\tau(x)] \sum_{s=0}^{r-1} \frac{[\tau(x)]^s}{s!} \\ \tau(x) &= (1 + \xi \frac{x - \mu}{\sigma})^{-1/\xi} \end{aligned} \right\} \quad (9)$$

进一步推导出前 r 个最大次序统计量的联合分布,则可以根据样本中的前 r 个最大值,采用各种参数估计方法(如极大似然估计法),估计最大值分布的参数。

2.3 点过程模型

对基本定理中的 X_1, X_2, \dots, X_n 及 \widetilde{M}_n ,如果式(3)成立,令 $\widetilde{X}_i = (X_i - b_n)/a_n, x^*$ 与 x_* 分别表示规范化最大值变量 \widetilde{M}_n 的极值分布 $H(x)$ 的上界与下界,则对于任意的 $u > x_*$,点过程序列 $N_n(\cdot)$ 为

$$N_n(\cdot) = \sum_{i=1}^n \epsilon_{i/n, \widetilde{X}_i}(\cdot) \quad n = 1, 2, 3, \dots \quad (10)$$

式(10)收敛到 $(0, 1] \times [u, +\infty)$ 上的 Poisson 过程,其在 $A = [t_1, t_2] \times [x, x^*]$ 上的强度测度 $\Lambda(A)$ 为

$$\Lambda(A) = (t_2 - t_1)[1 + \xi(x - \mu)/\sigma]^{-1/\xi}$$

据此可推导该 Poisson 过程的极大似然函数,并选取大于阈值 u 的样本,对最大值变量 M_n 的 GEV 分布参数进行估计。

3 极值估计

利用全概率方法对结构进行安全性评估,需要获取荷载效应的最大值分布函数,但实际中多采用荷载及其效应的底分布或最大值分布的某个高分位数来表示可能发生的极值。由式(1)可知,最大值变量 M_n 分布的 p 分位数等于底分布 $F(x)$ 的 $p^{1/n}$ 分位数,即

$$F_{M_n}(x_p) = F^n(x_p) = p^n \quad (11)$$

重现期及重现水平也常用于表述极值,对于分析基准期 T ,称 $u(T)$ 为荷载(效应)在 T 内的重现水平

$$u(T) = F^{-1}(1 - 1/T) \quad (12)$$

$u(T)$ 表示 T 年内荷载(效应)超过 $u(T)$ 的平均次数为 1。对于荷载(效应)水平 u ,称 T_u 为 u 的重现期

$$T_u = 1/[1 - F(u)] \quad (13)$$

在式(12)、(13)中,如果 F 是年最大值的截口分布,则 T 的单位为年;如果 F 是任意一辆重车车辆荷载的截口分布,则 T 的单位为辆。

4 重车车辆荷载的概率分布

根据连续 5 d 内采集的广州北环高速公路的 5 型车(4 轴及以上车辆,本文中亦称为重车)车质量样本,共计 1 383 个,统计得到车质量分布的直方图,并用非参数估计方法得到其密度曲线,见图 1。由图 1 可以看出,样本的分布呈现出多峰特征。

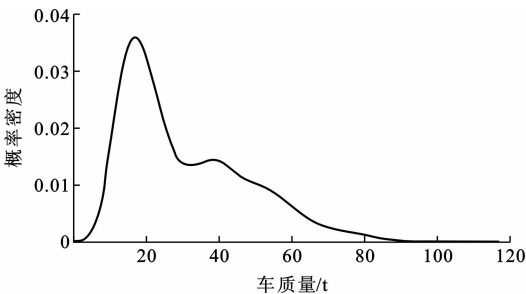


图 1 5 型车车质量样本的概率密度曲线
Fig. 1 Probability Density Curves of Weight of Type 5 Vehicles

4.1 现行方法

采用加权和函数对车质量样本拟合概率分布,函数形式表示如下

$$F(x) = \sum_{i=1}^k a_i \Phi_i(x; \mu_i, \sigma_i) \quad (14)$$

式中: k 为模态数量; a_i 为各模态权重, $\sum_{i=1}^k a_i = 1$; Φ_i 为第 i 个峰的累积分布函数; μ_i, σ_i 为分布参数。

选用正态分布的加权和函数作为样本分布函数,模型参数的极大似然估计值见表 1,对拟合结果做假设检验,得到 p 值为 0.994,说明样本不拒绝四峰正态分布。

表 1 四峰正态分布的参数估计

Tab. 1 Parameter Estimations of Four-modal Normal Distribution				
i	1	2	3	4
a	0.252	0.239	0.284	0.225
μ	14.7	21.1	38.9	51.8
σ	2.1	3.4	10.3	17.9

注: p 值为 0.994。

类似地,采用对数正态分布的加权和函数拟合样本数据,结果见表 2。图 2 为 2 种多峰正态分布的概率密度曲线。

4.2 阈值模型

用 GP 分布拟合 5 型车车质量的概率分布尾部,首先需要确定合理的阈值。图 3 为尺度参数和

表 2 双峰对数正态分布的参数估计
Tab.2 Parameter Estimations of Bimodal Lognormal Distribution

<i>i</i>	1	2
<i>a</i>	0.560	0.440
<i>μ</i>	2.89	3.82
<i>σ</i>	0.28	0.28

注:*p* 值为 0.526。

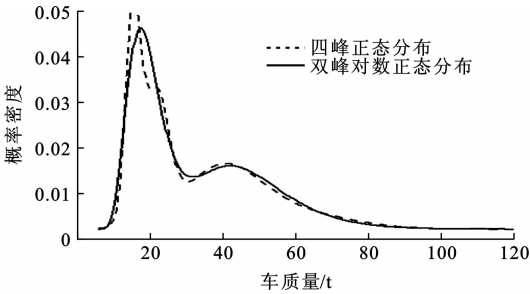


图 2 2 种多峰正态分布的概率密度曲线

Fig.2 Probability Density Curves of Two Types of Multimodal Normal Distribution

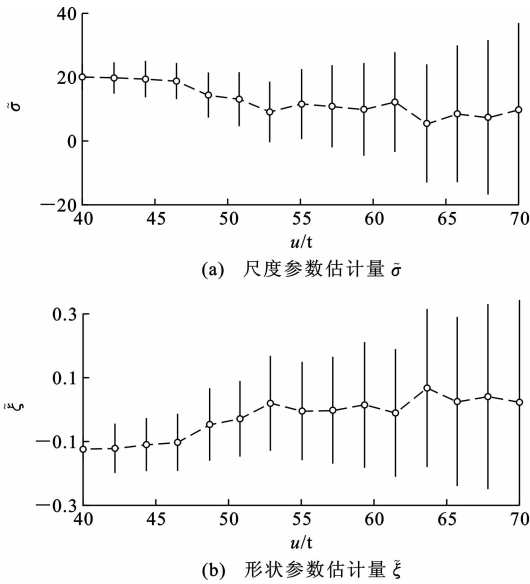


图 3 GP 分布参数估计及相应的 95%置信区间

Fig.3 Parameter Estimations of GP Distribution and Its Corresponding 95% Confidence Interval

形状参数的估计量 $\hat{\sigma}$ 、 $\hat{\xi}$ 与阈值 u 变化的关系。根据阈值选取的原则,即所选的阈值可以获得尽量多的超出量以减小估计值的方差,且该阈值超出量的 GP 分布的形状参数应保持不变,取阈值 $u=51.68\text{ t}$ 。由图 3 可以看出,该取值不仅能使估计量 $\hat{\sigma}$ 、 $\hat{\xi}$ 变化较小,且此时估计量有较小的偏差。

对样本的阈值超出量拟合 GP 分布,得到各参数的估计,见表 3。对拟合的分布做假设检验,得到 p 值

表 3 超出量 GP 分布的参数估计
Tab.3 Parameter Estimations of GP Distribution of Vehicle Weight Excess

参数	σ	ξ	样本数	N_{un}^{-1}	<i>p</i> 值
估计值	11.5	0.003 4	208	0.15	0.998
标准误差	1.1	0.068 0			

为 0.998,因此样本超出量不拒绝 GP 分布。

根据式(7),重车车辆荷载的概率分布函数为

$$F(x) = 1 - 0.15(1 + \xi \frac{x - 51.68}{\sigma})^{-\frac{1}{\xi}} \quad x \geq 51.68 \tag{15}$$

对于小于阈值的部分,可以用经验分布函数代替或另行拟合,但与极值的估计无关。

4.3 阈值模型与现行方法的比较

将阈值模型与现行方法得到的重车概率分布函数绘制在对数坐标下,观察其尾部分布,即 $\bar{F}(x) = 1 - F(x)$, $x > 51.68$,则差别立刻显现(图 4)。由图 4 可以看出:阈值模型得到的分布与样本数据较为一致;四峰正态分布与双峰对数正态分布在车质量大于 90 t 后逐渐偏离样本,且这 2 种分布的尾部比阈值模型得到的 GP 分布及经验分布收敛得快。由此可知,采用现行方法得到的多峰概率分布对车质量最大值分布进行估计,结果小于实际值,也小于用阈值模型得到的 GP 分布对极值的估计。

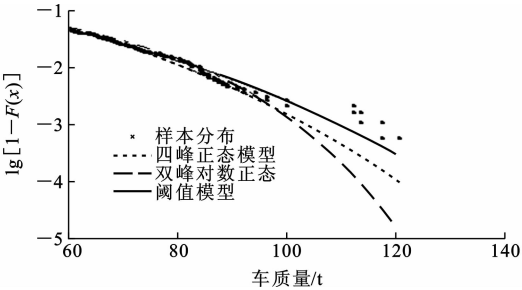


图 4 车质量概率分布的尾部比较

Fig.4 Tail Comparisons of Probability Distributions of Vehicle Weights

5 次序统计量模型

将 1 383 个 5 型车车质量样本,按每天分组,取每组的前 r 个最大值作为极值样本,又根据广州北环高速公路公司提供的车流量,统计出广州广园至沙河路段的日均 5 型车车流量约为 2 613 辆。采用次序统计量模型拟合 1 d 的车质量最大值分布,结果见表 4、5,表 4、5 中括号内数值为对应参数估计值的标准误差。图 5 为不同 r 时模型的重现水平诊断图,由图 5 可以看出, $r=10$ 时模型对样本的拟合

表 4 不同 r 时 5 型车 1 d 的车质量最大值的
GEV 分布 (按天分组)

Tab. 4 GEV Distributions of Maximum Weight of Type 5
Vehicles in 1 d with Different r (Grouped by Day)

r	10	30	50	70
μ	93.8(4.80)	94.9(3.90)	94.9(3.10)	95.2(2.97)
σ	13.2(4.20)	11.4(2.30)	9.9(1.39)	9.6(1.23)
ξ	0.187 3 (0.179 3)	-0.011 5 (0.074 4)	-0.094 1 (0.045 6)	-0.115 6 (0.038 6)
$x_{0.95}/t$	146.2	128.3	120.6	119.4

注： $x_{0.95}$ 表示概率为 95% 时对应的分位数。

表 5 不同 r 时 5 型车 1 d 的车质量最大值的
GEV 分布 (按小时分组)

Tab. 5 GEV Distributions of Maximum Weight of Type 5
Vehicles in 1 d with Different r (Grouped by Hour)

r	1	2	5
μ	58.3(1.7)	60.5(1.4)	64.7(1.4)
σ	15.5(1.2)	15.4(0.8)	16.7(0.5)
ξ	-0.133 4 (0.049 5)	-0.171 7 (0.033 7)	-0.094 1 (0.045 6)
$x_{0.95}/t$	96.33	96.31	99.19

效果较好。

表 4、5 的结果表明,GEV 分布的估计对 r 的选取具有较强的敏感性。此外,不同的分组对 GEV 分布的估计也有较大影响,因此,不建议采用次序统计量模型对荷载极值进行估计。

6 点过程模型

选取与阈值模型相同的阈值,即 $u=51.68\text{ t}$,采用点过程模型,估计 100 辆 5 型车车质量最大值的 GEV 分布;并将阈值模型得到的 5 型车车质量分布尾部[式(15)],按式(1)计算 100 辆 5 型车车质量最大值的 GEV 分布;二者得到的结果绘制于图 6。由图 6 可以看出,2 条概率密度曲线几乎重合,因此,2 种模型所得结果是一致的。

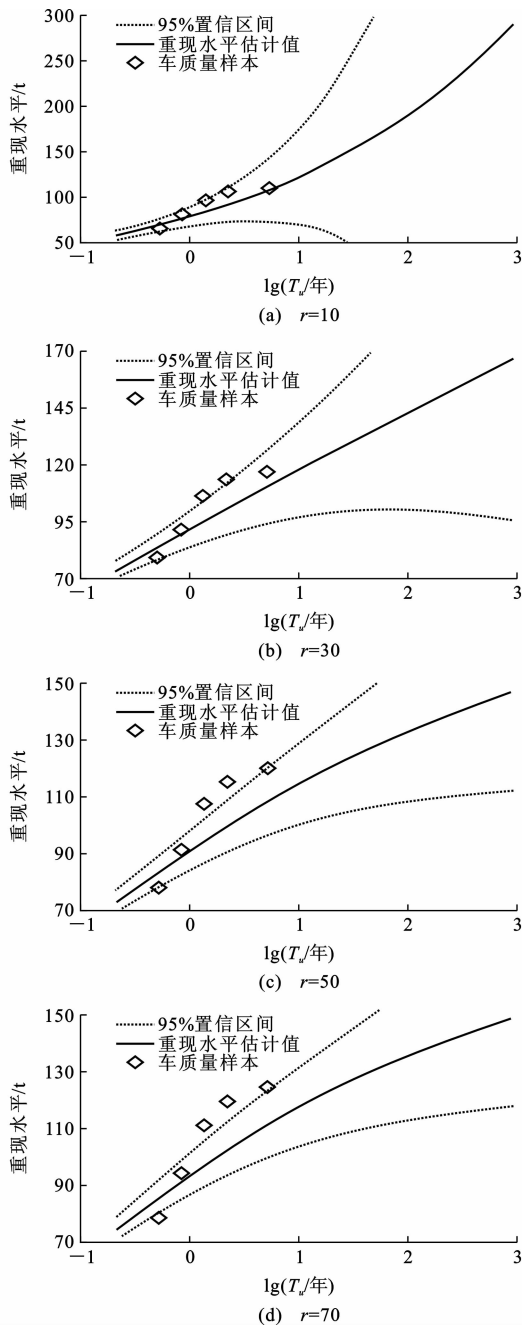


图 5 不同 r 时模型的重现水平诊断图

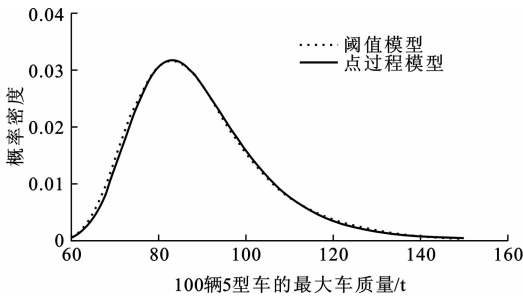


图 6 100 辆 5 型车的最大车质量概率密度分布
Fig. 6 Probability Density Distributions of Maximum
Weight of 100 Type 5 Vehicles

7 车辆荷载最大值估计的比较

表 6 为各种方法得到的 5 型车 1 d 的车质量最大值分布的结果。由表 6 可以看出,现行方法得到的结果小于用新的极值方法得到的结果。原因是多峰正态分布的尾部偏离样本数据较大,阈值模型得到的重车概率分布尾部与样本数据较为接近(图 4),在此基础上,得到的极值也能反映实际的车质量极值。此外,点过程模型得到的最大值分布与阈值

Fig. 5 Return Level Plots of Models with Different r

表 6 各种方法对极值估计结果的比较
Tab. 6 Comparisons of Extreme Value Estimations with Different Methods

方法	得到的分布	分布类型	分布参数	$x_{0.95}/t$
现行方法	底分布	四峰正态		118.7
		双峰对数正态		136.0
		GP		156.1
次序统计量模型(按天分组, $r=10$)	最大值分布	GEV	$\mu=94.9,$ $\sigma=9.9,$ $\xi=-0.094\ 1$	146.2
点过程模型	最大值分布	GEV	$\mu=121.2,$ $\sigma=11.7,$ $\xi=-0.003\ 4$	156.2

模型基础上得到的结果是一致的,而次序统计量模型的拟合结果与阈值模型和点过程模型相差较大,因为其样本受分组及 r 的影响很大。

8 结 语

- (1)对车辆荷载进行极值分析,不建议采用类似参数统计的方法,即对其底分布进行主观拟合及检验。因为常用方法只满足出现概率较大的中间样本的拟合,对出现概率很小的极值则会造成较大的统计误差,从而导致车辆荷载极值的低估。
- (2)3 种极值统计模型中,阈值模型与点过程模型对样本的拟合结果较好,建议车辆荷载的极值分析中采用这 2 种模型;而次序统计量模型受样本分组及 r 的影响较大,不建议采用。

参考文献:

References:

[1] GB 50009—2001,建筑结构荷载规范[S].
GB 50009—2001, Load Code for Design of Building Structures[S].

[2] GB/T 50283—1999,公路工程结构可靠度设计统一标准[S].
GB/T 50283—1999, Unified Standard for Reliability Design of Highway Engineering Structures[S].

[3] “公路桥梁车辆荷载研究”课题组.公路桥梁车辆荷载研究[J].公路,1997(3):8-12.
“The Research of Vehicle Load of Highway and Bridge” Research Team. The Research of Vehicle Load of Highway and Bridge[J]. Highway, 1997(3): 8-12.

[4] 梅刚,秦权,林道锦.公路桥梁车辆荷载的双峰分布概率模型[J].清华大学学报:自然科学版,2003,43(10):1394-1396,1404.

MEI Gang, QIN Quan, LIN Dao-jin. Bi-modal Probabilistic Model of Highway and Bridge Vehicle Loads [J]. Journal of Tsinghua University: Science & Technology, 2003, 43(10): 1394-1396, 1404.

[5] MIAO T J, CHAN T H T. Bridge Live Load Models from WIM Data[J]. Engineering Structures, 2002, 24(8): 1071-1084.

[6] 贡金鑫,李文杰,赵君黎,等.公路桥梁车辆荷载概率模型研究(一)——非治超地区[J].公路交通科技, 2010, 27(6): 40-45, 57.
GONG Jin-xin, LI Wen-jie, ZHAO Jun-li, et al. Research on Probabilistic Model of Highway Bridge Vehicle Loads (1) — Non-controlling Area[J]. Journal of Highway and Transportation Research and Development, 2010, 27(6): 40-45, 57.

[7] 贡金鑫,李文杰,赵君黎,等.公路桥梁车辆荷载概率模型研究(二)——计重收费地区和强制治超地区[J].公路交通科技, 2010, 27(7): 56-60.
GONG Jin-xin, LI Wen-jie, ZHAO Jun-li, et al. Research on Probabilistic Model of Highway Bridge Vehicle Loads (2) — Toll-by-weight Area and Overload Controlling Area[J]. Journal of Highway and Transportation Research and Development, 2010, 27(7): 56-60.

[8] NOWAK A S. Load Model for Highway Bridges[J]. Computer Applications in Technology, 1993, 12(2): 17-30.

[9] 史道济.实用极值统计方法[M].天津:天津科学技术出版社,2006.
SHI Dao-ji. Practical Method for Extremal Statistics [M]. Tianjin: Tianjin Science and Technology Press, 2006.

[10] 朱国庆,张维,张小薇.极值理论应用进展评析[J].系统工程学报,2001,16(1):72-77.
ZHU Guo-qing, ZHANG Wei, ZHANG Xiao-wei. A Review on the Process of Applications Research of the Extreme Value Theory [J]. Journal of Systems Engineering, 2001, 16(1): 72-77.

[11] 杜江.桥梁管理系统中安全性评估与维护决策分析的方法探讨[D].广州:华南理工大学,2008.
DU Jiang. Safety Assessment and Maintenance Decision in the Bridge Management System [D]. Guangzhou: South China University of Technology, 2008.

[12] 韩大建,杜江.工程结构作用极值分析方法研究[J].建筑科学与工程学报,2008,25(2):68-71,126.
HAN Da-jian, DU Jiang. Research on Extreme Value Analysis Method of Actions on Engineering Structures[J]. Journal of Architecture and Civil Engineering, 2008, 25(2): 68-71, 126.