

广东省风速极值 I 型分布参数估计方法的比较

黄浩辉 宋丽莉 植石群 刘爱君

(广东省气候中心,广州 510080)

提 要: 利用广东省 86 个气象站 1951—2003 年的年最大风速序列资料,应用极值 I 型分布函数对各站的年极值风速进行了概率计算,计算过程中极值 I 型分布函数分别采用了 3 种参数估计方法:矩法、耿贝尔法和极大似然法,结合 3 种表征参数估计优良性的指标,对不同的参数估计方法进行了比较,结果表明:大多数情况下采用耿贝尔法时拟合效果最好,即使在矩法或极大似然法为优的情况下,其拟合优良性指标与耿贝尔法也较为接近。

关键词: 风速 极值 I 型 参数估计

Comparison of Estimation of Wind Speed Extreme-I Distribution Parameters in Guangdong Province

Huang Haohui Song Lili Zhi Shiqun Liu Aijun

(Climate Center of Guangdong Province, Guangzhou 510080)

Abstract: On the basis of the series of annual maximum wind velocity at 86 national weather stations in Guangdong Province from 1951 to 2003, the return periods of annual maximum wind velocity of every station are calculated with the function of extreme-I using three methods of estimation of parameters, i. e., squares method, Gumbel's method and the method of maximum likelihood. The fitting results show that Gumbel's method is the best in most cases, even in the case that the squares method or the method of maximum likelihood seems better and its choiceness index is closer to that of the best method.

Key Words: wind speed function of extreme-I estimation of parameter

引 言

随着国民经济的高速发展,很多大型基

建项目需要进行抗风设计,如大型桥梁、高耸建筑物、港口、码头、仓储等。在抗风设计过程中,目前国内外普遍应用极值 I 型分布函数进行极值风速的概率计算。本文主要讨

论极值 I 型分布的 3 种常用参数估计方法: 矩法、耿贝尔法和极大似然法。根据我们在港珠澳大桥、广州新中轴线电视塔和深圳 LNG 等项目的实践经验, 不同的参数估计方法得到的分布参数有一定差别, 从而计算得到的重现期风速也有一定的差别。有时候, 对较长的重现期(如 100 年), 计算得到的最大风速差距较大, 由于工程抗风设计所需的重现期最大风速值直接影响到工程的安全性和投资造价, 因此, 在实际工作中如何选用合适的参数估计方法, 使计算得到的结果更为合理、准确, 是一个具有重要实际意义的问题。孟庆珍等应用极值 I 型分布函数对成都地面年最大风速进行了概率计算, 参数估计方法采用矩法和耿贝尔法, 结果表明耿贝尔法效果较好^[1]; 段忠东等比较了极值 I 型分布的几种参数估计方法: 矩法、耿贝尔法、极大似然法和概率权值法, 结果表明, 当样本容量较小时, 耿贝尔法的估计结果较差^[2]。那么, 广东省的具体情况又是如何呢? 本文应用极值 I 型分布函数对广东省 86 个气象站的年最大风速序列进行了计算, 对不同的参数估计方法进行了比较。

1 极值 I 型分布函数

极值 I 型分布函数为:

$$F(x) = \exp(-\exp(-a(x-u)))$$

$$(a > 0, -\infty < u < \infty)$$

式中, a 为分布的尺度参数, u 为分布的位置参数, 只要利用已有的最大风速序列 x_1, x_2, \dots, x_n 合理估计出参数 a, u 的数值, 则 $F(x)$ 被唯一确定。

重现期为 R (概率为 $1/R$) 时最大风速为:

$$R_x = u - \frac{1}{a} \ln \left[\ln \left(\frac{R}{R-1} \right) \right]$$

2 极值 I 型分布的 3 种常用参数估计方法

2.1 矩法

矩法估计在数学计算上最为简单。参数 a, u 与矩的关系为:

$$\text{一阶矩(数学期望): } E(x) = \frac{\gamma}{a} + u, \text{ 式}$$

中 γ 为欧拉常数, $\gamma \approx 0.57722$

$$\text{二阶矩(方差): } \sigma^2 = \frac{\pi^2}{6a^2}$$

由此得到:

$$a = \frac{1.28255}{\sigma} \quad u = E(x) - \frac{0.57722}{a}$$

在实际计算中一般用有限样本容量的均值和标准差作为理论值 $E(x)$ 和 σ 的近似估计。

2.2 耿贝尔法

耿贝尔法是一种直接与经验频率相结合的参数估计方法。假定最大风速有序序列: $x_1 \leq x_2 \leq \dots \leq x_n$, 则经验分布函数为:

$$F^*(x_i) = \frac{i}{n+1} \quad i=1, 2, \dots, n$$

取如下序列:

$$y_i = -\ln(-\ln(F^*(x_i))) \quad i=1, 2, \dots, n$$

$$\text{可得: } a = \frac{\sigma(y)}{\sigma(x)} \quad u = E(x) - \frac{E(y)}{a}$$

在实际计算中可用有限样本容量的均值和标准差作为 $E(x)$ 和 $\sigma(x)$ 的估计值。

2.3 极大似然法

在统计学理论上, 极大似然估计是一种较优的参数估计方法。极值 I 型分布函数的概率密度函数为:

$$f(x) = F'(x) = a \exp[-a(x-u) - \exp(-a(x-u))]$$

当观测资料 x_1, x_2, \dots, x_n 给定时, 作极大似然函数并取对数, 得:

$$\ln L = n \ln a - \sum_{i=1}^n a(x_i - u) - \sum_{i=1}^n \exp(-a(x_i - u))$$

将 a, u 看作变量, 将上式分别对 a, u 求导并令其为零, 得:

$$n \exp(-au) = \sum_{i=1}^n \exp(-ax_i)$$

$$n(\bar{x} - \frac{1}{a} \exp(-au)) = \sum_{i=1}^n x_i \exp(-ax_i)$$

参数 a, u 可用迭代法求解。

2.4 样本序列出现特大值时参数的修改方法

我国气象台站基本在 1949 年以后才有年最大风速的记录, 至今也就 50 多年的资料, 在不长的年最大风速序列中, 有时会发生这样的现象: 有一个数据特别大(特大值), 按序列的大小顺序来看, 很易发觉这个特大值和其它数值之间有着显著脱节的现象, 例如广东省湛江站 1951—2003 年的年最大风速序列, 样本容量为 53, 其中 1954 年的年最大风速达 $47.2 \text{ m} \cdot \text{s}^{-1}$, 而其余 52 个数据中最大才是 $26.3 \text{ m} \cdot \text{s}^{-1}$, 两者相差 $20.9 \text{ m} \cdot \text{s}^{-1}$, 这种现象说明了这个特大值的重现期是比现有的观测年数(资料序列长度)要长得多, 如果对它不加以处理, 风速分布曲线的数学期望和方差均会显著提高, 而使所求的风速偏高^[3]。

设年最大风速序列 x_1, x_2, \dots, x_n , 另有一特大风速 x_N , 并估计其重现期为 N 年($N > n$), 设 \bar{x}_n 和 σ_n^2 分别表示 n 年资料的均值与方差, 则 N 年资料的均值为:

$$\bar{x}_N = \frac{1}{N} [x_N + (N-1)\bar{x}_n]$$

N 年资料的方差为:

$$\sigma_N^2 = \frac{1}{N-1} [(x_N - \bar{x}_N)^2 + (N-1)\sigma_n^2]$$

故出现特大值时, 分布曲线参数 a 和 u 应当修改为:

(1) 矩法估计

$$a_N = \frac{1.28255}{\sigma_N} \quad u_N = \bar{x}_N - \frac{0.57722}{a_N}$$

(2) 耿贝尔法估计

$$a_N = \frac{\sigma_N(x)}{\sigma_N(y)} \quad u_N = \bar{x}_N - \frac{\bar{y}_N}{a_N}$$

3 参数估计方法的比较

3.1 表征参数估计优良性的指标

为了找出有效性最高的估计值和比较不同方法所求得的估计值的优良性, 计算了以下 3 种表征参数估计优良性的指标。

$$(1) \text{ 剩余方差 } S_f^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \hat{x}_i)^2$$

式中 n 为样本容量, x_i 为有序样本: $x_1 \leq x_2 \leq \dots \leq x_n$; \hat{x}_i 为拟合值, 按下式计算:

$$\hat{x}_i = u - \frac{1}{a} \ln(-\ln(F^*(x_i)))$$

(2) 拟合相对偏差

$$R_f = 100 * \frac{1}{n} \sum_{i=1}^n \frac{|x_i - \hat{x}_i|}{x_i}$$

(3) 柯尔莫哥洛夫拟合适度检验指标

$$K_f = D_n \sqrt{n}$$

这里 n 为样本容量, D_n 如下式所示:

$$D_n = \max \{ |F^*(x_i) - F(x_i)| \}$$

这里 x_i 为有序样本, D_n 表示在所有各点上, 经验分布与假设的理论分布之差的最大值。取信度为 0.05, 查表得只要 $K_f < 1.35$, 则认为年最大风速服从极值 I 型分布。

显然, 以上 3 种指标均是越小表示拟合效果越好。

3.2 不同参数估计方法的比较

对广东省 86 个气象站 1951—2003 年的年最大风速序列根据其历史沿革, 依据有关规范进行了次时和高度订正, 统一订正为距地面 10m 高度的 10 分钟平均年最大风速,

这 86 个站均匀地分布在广东省境内,测站地面海拔高度在 2.5~302.8m 之间,较好地代表了沿海地区、珠江三角洲平原地区以及内陆丘陵和山地的情况。

对 3 种表征参数估计优良性的指标 S_f^2 、 R_f 、 K_f 的计算结果统一取 3 位小数,根据统计学理论,比较不同参数估计方法的优良性主要比较剩余方差 S_f^2 ,若两种方法的 S_f^2 指标持平,则再比较 R_f 、 K_f 指标^[4]。

计算结果表明,86 个站中,有 4 个站用 3 种参数估计方法计算得到的 K_f 指标均大于 1.35,即年最大风速不服从极值 I 型分布;其余 82 个站用 3 种参数估计方法计算得到的 K_f 指标均小于 1.35,即 3 种参数估计方法均能得到有效的分布参数,年最大风速服从极值 I 型分布。

上述 82 个站中,有 6 个站出现特大值(此时实质只有矩法和耿贝尔法估计),另有 6 个站极大似然法无解,为便于比较,列出余下 70 个站 3 种参数估计方法的 S_f^2 、 R_f 、 K_f 指标的平均值,矩法估计为 $0.518\text{m}^2 \cdot \text{s}^{-2}$ 、

4.062% 和 0.730,耿贝尔法估计为 $0.440\text{m}^2 \cdot \text{s}^{-2}$ 、3.616% 和 0.670,极大似然法估计为 $0.537\text{m}^2 \cdot \text{s}^{-2}$ 、3.588% 和 0.676。根据上述的指标比较办法,这 70 个站的计算实例中,有 4 例以矩法估计为优,有 8 例以极大似然法估计为优,其余 58 例以耿贝尔法估计为优;在 6 个站极大似然法无解的计算实例中,有 1 例以矩法估计为优,5 例以耿贝尔法估计为优;在 6 个出现特大值的站的计算实例中,均是耿贝尔法估计为优。因此,总体而言,耿贝尔法最优。

另一方面,上述 70 个站的计算实例中,有 38 例以矩法估计效果最差,有 31 例以极大似然法估计效果最差,仅有 1 例以耿贝尔法估计效果最差。

将矩法、耿贝尔法和极大似然法估计分别为优的 3 种情况的计算结果列于表 1、2、3,表中列出不同参数估计方法计算得到的极值 I 型分布参数 a 、 u 值及其优良性指标 S_f^2 、 R_f 和 K_f ,同时列出 100 年重现期的年最大风速 V_{100} 。

表 1 矩法估计为优的计算实例

站名	参数估计方法	$S_f^2/\text{m}^2 \cdot \text{s}^{-2}$	$R_f/\%$	K_f	a	$u/\text{m} \cdot \text{s}^{-1}$	$V_{100}/\text{m} \cdot \text{s}^{-1}$
阳山	矩法	0.740	7.356	1.179	0.493	10.432	19.8
	耿贝尔法	0.750	6.833	1.114	0.449	10.384	20.6
	极大似然法	1.073	6.334	1.037	0.384	10.272	22.2
翁源	矩法	1.270	10.73	1.304	0.422	8.958	19.8
	耿贝尔法	1.305	9.69	1.171	0.382	8.899	20.9
	极大似然法	1.651	8.79	1.036	0.337	8.775	22.4
四会	矩法	1.292	4.824	0.903	0.366	12.517	25.1
	耿贝尔法	1.301	6.111	0.927	0.333	12.451	26.3
	极大似然法	2.079	3.813	0.884	0.476	12.724	22.4

在矩法估计为优的计算实例中,矩法和耿贝尔法的 3 项优良性指标较为接近,计算得到的 100 年重现期的年最大风速 V_{100} 值也相差不大。从表 1 看出,在四会站的计算实例中,矩法估计效果最好,极大似然法估计效果最差,极大似然法计算得到的 V_{100} 值比矩法小 $2.7\text{m} \cdot \text{s}^{-1}$,比耿贝尔法小 $3.9\text{m} \cdot \text{s}^{-1}$,

虽然其 R_f 、 K_f 指标均为最小,但其 S_f^2 指标最大,显然,有理由认为矩法估计的结果更为合理。

表 2 为耿贝尔法估计为优的计算实例,在徐闻站的计算实例中,耿贝尔法计算得到的 V_{100} 值比矩法大 $1.8\text{m} \cdot \text{s}^{-1}$,比极大似然法大 $2.0\text{m} \cdot \text{s}^{-1}$,但其 3 项优良性指标均为最

表 2 耿贝尔法估计为优的计算实例

站名	参数估计方法	$S_f^2/m^2 \cdot s^{-2}$	$R_f/\%$	K_f	a	$u/m \cdot s^{-1}$	$V_{100}/m \cdot s^{-1}$
惠来	矩法	0.543	4.152	0.709	0.261	14.271	31.9
	耿贝尔法	0.375	3.101	0.610	0.238	14.181	33.5
	极大似然法	0.385	3.248	0.648	0.243	14.159	33.1
珠海	矩法	0.354	2.489	0.445	0.278	16.645	33.2
	耿贝尔法	0.160	1.726	0.314	0.252	16.554	34.8
	极大似然法	0.229	2.082	0.376	0.267	16.587	33.8
徐闻	矩法	0.555	3.105	0.410	0.231	13.269	33.2
	耿贝尔法	0.320	2.748	0.316	0.210	13.166	35.0
	极大似然法	0.606	3.222	0.412	0.233	13.247	33.0
湛江*	矩法	2.379	4.612	0.597	0.270	14.205	31.3
	耿贝尔法	2.246	4.460	0.574	0.265	14.186	31.5
	极大似然法	—	—	—	—	—	—

* : 湛江站含有特大值

表 3 极大似然法估计为优的计算实例

站名	参数估计方法	$S_f^2/m^2 \cdot s^{-2}$	$R_f/\%$	K_f	a	$u/m \cdot s^{-1}$	$V_{100}/m \cdot s^{-1}$
曲江	矩法	0.099	1.443	0.694	0.793	12.131	17.9
	耿贝尔法	0.085	1.581	0.675	0.723	12.102	18.5
	极大似然法	0.085	1.456	0.663	0.753	12.126	18.2
博罗	矩法	0.824	5.33	0.893	0.344	11.977	25.4
	耿贝尔法	0.769	5.347	0.852	0.312	11.907	26.6
	极大似然法	0.761	5.318	0.853	0.326	11.961	26.1
新会	矩法	0.818	3.604	0.742	0.344	13.906	27.3
	耿贝尔法	0.768	3.008	0.608	0.313	13.838	28.5
	极大似然法	0.762	3.013	0.632	0.317	13.817	28.3

小,因此,有理由认为耿贝尔法估计的结果最为合理,徐闻站的年最大风速概率分布拟合曲线如图 1 所示,从图中看出,当频率较小(即极值风速较大)时,耿贝尔法的拟合曲线最接近实测曲线。在惠来站的计算实例中,矩法估计效果最差,其计算得到的 V_{100} 值分别较耿贝尔法和极大似然法小 $1.6m \cdot s^{-1}$ 和 $1.2m \cdot s^{-1}$ 。在 6 个出现特大值的站的计算实例中,均是耿贝尔法估计为优,但其优良性指标与矩法均较为接近,计算得到的 V_{100} 值均是以耿贝尔法的结果稍大一点。此外,如对湛江站的特大值不予处理,则用耿贝尔法计算得到的 V_{100} 值为 $38.3m \cdot s^{-1}$,而对特大值加以处理,参数修改后, V_{100} 值为 $31.5m \cdot s^{-1}$,两者相差 $6.8m \cdot s^{-1}$,达后者的 21.6%,差距是较大的。

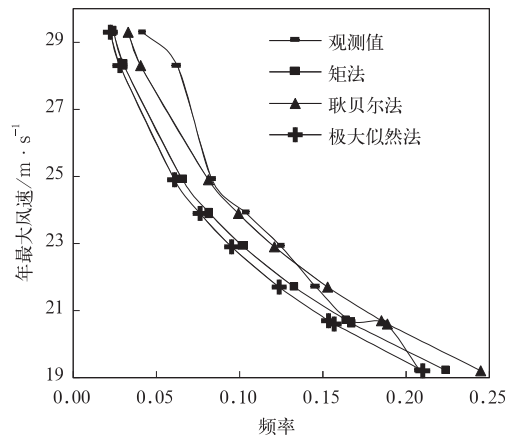


图 1 徐闻站的年最大风速概率分布拟合曲线

在 8 个极大似然法估计为优的计算实例中,极大似然法和耿贝尔法的 3 项优良性指标较为接近,计算得到的 V_{100} 值相差很小,最大为 $0.7m \cdot s^{-1}$,最小为 $0.1m \cdot s^{-1}$,平均为

$0.3\text{m} \cdot \text{s}^{-1}$ 。

表 4 为仅有的 1 例以耿贝尔法估计效果最差的情况,从表中看出,耿贝尔法的 3 项优

良性指标与矩法和极大似然法是较为接近的,其 V_{100} 值较矩法和极大似然法大 $0.8\text{m} \cdot \text{s}^{-1}$ 和 $0.7\text{m} \cdot \text{s}^{-1}$ 。

表 4 耿贝尔法估计效果最差的计算实例

站名	参数估计方法	$S_f^2/\text{m}^2 \cdot \text{s}^{-2}$	$R_f/\%$	K_f	a	$u/\text{m} \cdot \text{s}^{-1}$	$V_{100}/\text{m} \cdot \text{s}^{-1}$
	矩法	0.549	3.441	0.802	0.569	10.529	18.6
德庆	耿贝尔法	0.555	3.780	0.720	0.518	10.487	19.4
	极大似然法	0.545	3.393	0.747	0.566	10.566	18.7

4 结 论

根据广东省 86 个气象站年最大风速序列的计算实例,极值 I 型分布函数的 3 种常用参数估计方法中,在大多数情况下采用耿贝尔法时拟合效果最好,即使在矩法或极大似然法为优的情况下,其拟合优良性指标与耿贝尔法也较为接近,计算得到的 100 年重现期的年最大风速值也相差不大。因此,在广东省应用极值 I 型分布函数进行极值风速的概率计算时,其参数估计方法应首选耿贝尔法。

参考文献

- [1] 孟庆珍,唐谟智. 成都地面气温与风速年极大值的渐近分布及参数估计[J]. 成都气象学院学报,1997,12(4):284-291.
- [2] 段忠东,周道成. 极值概率分布参数估计方法的比较研究[J]. 哈尔滨工业大学学报,2004,36(12):1605-1609.
- [3] 广东省沿海风压研究小组. 风压问题研究及计算方法[R]. 广东,1977:25.
- [4] 屠其璞,王俊德,丁裕国,等. 气象应用概率统计学[M]. 北京:气象出版社,1984:243-244.
- [5] 袁春红,薛桁,杨振斌. 建筑风荷载风压计算中的若干问题[J]. 气象,2002,28(1):39-42.