



黄骅港外航道骤淤量重现值的区间估计*

张靖婷, 陶山山, 董胜, 李晨阳

(中国海洋大学工程学院, 山东青岛266100)

摘要: 黄骅港外航道经常遭受大风骤淤的影响, 对骤淤量重现值的研究有助于泥沙淤积整治标准的确定。重现值的点估计往往因方法的不同而产生较大差异。区间估计方法可以给出一定置信水平下设计重现值的参考区间, 从而为设计和施工人员提供相应的指导。基于黄骅港年最大骤淤量的预测值序列, 分别采用对数正态分布和最大熵分布函数对其进行拟合, 并利用极大似然法对骤淤量重现值进行区间估计。骤淤量重现值的置信区间长度随重现期的增加而增大, 最大熵分布计算结果精度较高, 可为泥沙淤积的整治提供参考。

关键词: 骤淤量; 重现期; 置信区间; 极大似然估计

中图分类号: P 737.14

文献标志码: A

文章编号: 1002-4972(2016)04-0118-05

Interval estimation of return values of sudden siltation volume in outer channel of Huanghua port

ZHANG Jing-ting, TAO Shan-shan, DONG Sheng, LI Chen-yang

(Ocean University of China, Qingdao 266100, China)

Abstract: The outer channel of Huanghua port often confronts with sudden siltation induced by strong winds, and the study on the return values of sudden siltation volumes can be helpful to the treatment of sediment accumulation. The point estimations of return values given by diverse methods may produce large differences. The interval estimation can present some reference intervals of design return values under certain confidence levels, so it will provide corresponding instruction for the designers and constructors. Fitting the predicted sequences of the annual maximum sudden siltation of Huanghua port with the lognormal and maximum entropy distributions separately, we estimate the confidence intervals of the return values of sudden siltation volumes by the maximum likelihood method. With the increase of return periods, the scale of the confidence intervals becomes large. The results given by the maximum entropy distribution have higher precision, so they can be applied as reference for the treatment of sediment accumulation.

Keywords: sudden siltation volume; return period; confidence interval; maximum likelihood estimation

黄骅港所处岸线属于粉沙质海岸, 泥沙易起易沉, 大风或风暴潮(台风、寒潮)过程极易造成外航道的骤淤^[1-2], 从而严重影响航道通航和港口运营^[3]。黄骅港自建港以来已发生多次大风骤

淤灾害, 迫使航道改线, 并于2004—2005年进行了整治与疏浚工程^[4-5]。在整治过程中, 科研技术人员提出了骤淤重现期的概念, 并利用概率统计方法, 根据年最大值序列计算得到了多年一遇的

收稿日期: 2015-10-22

*基金项目: 国家自然科学基金资助(51509227); 山东省自然科学基金资助(ZR2014EEQ030); 中央高校基本科研业务费专项(201413003)

作者简介: 张靖婷(1995—), 女, 研究方向为港口海岸工程及其与海洋环境的相互作用。

骤淤量重现值^[6]。由于该地区在风速达到6级以上且历时达到2 h后,航道才会发生明显淤积,故大风骤淤量可依据有效风能推导的风波淤积公式确定^[7]。

在骤淤量的频率分析中,季泽舟等^[6]选用对数正态和皮尔逊Ⅲ型分布进行年最大骤淤量的拟合,得到了不同重现期下骤淤量的值。但对于不同的参数估计方法,所求骤淤量重现值的点估计有较大差异。因此在工程上需要给出某种置信水平下设计值的取值范围,也即进行骤淤量重现值的区间估计。实际上,设计重现值与对应随机变量长期分布的分位数相对应^[8]。Coles针对广义极值分布,提供了一种基于极大似然方法的分位数区间估计^[9],该方法可推广到其他的分布类型中。Zhang等^[10]提出了一种最大熵分布函数,用以描述波面高程的统计规律。实际上,皮尔逊Ⅲ型分布为这种最大熵分布函数的特例^[11]。

本文根据极大似然区间估计理论,给出对数正态分布和最大熵分布分位数的置信区间;基于黄骅港1979—2003年的年最大骤淤量序列,利用以上两种分布类型进行数据拟合,从而得到骤淤量重现值的点估计和区间估计,进行优选比较。

1 骤淤量重现值

与工程中设计洪水、设计波浪的概念类似,骤淤量的重现期可定义为 $N=1/p$, p 为超越概率。若年最大骤淤量服从某种类型的长期分布(如对数正态分布或最大熵分布),则相应的 N 年一遇骤淤量重现值对应于该分布的上侧 $1/N$ 分位数 $x_{1/N}$ (即 x_p)。

1.1 最大骤淤量的长期分布

这里选用对数正态分布和最大熵分布进行年最大骤淤量序列的频率分析。

二参数对数正态分布(不含位置参数)的概率密度函数为:

$$f(x) = \frac{1}{x\sigma} \frac{\exp\left[-\frac{1}{2\sigma^2}(\ln x - \mu)^2\right]}{\sqrt{2\pi}} \quad (1)$$

式中: μ 和 σ 分别为形状参数和尺度参数。

最大熵分布^[10]的概率密度函数为:

$$f(x) = \xi\beta \frac{\gamma+1}{\xi} \Gamma^{-1}\left(\frac{\gamma+1}{\xi}\right) \cdot (x-a_0)^\gamma \cdot e^{-\beta(x-a_0)^\xi} \quad (2)$$

式中: $a_0 > 0$ 为位置参数; $\xi > 0$ 为矩约束的阶数; β 和 γ 为拉格朗日常数。

令 $\gamma = \alpha - 1$, $\xi = 1$,且 $a_0 = \delta$,即可得到皮尔逊Ⅲ型分布,其概率密度函数如下:

$$f(x) = \beta^\alpha \Gamma^{-1}(\alpha) \cdot (x-\delta)^{\alpha-1} e^{-\beta(x-\delta)} \quad (3)$$

因此,皮尔逊Ⅲ型分布可视为最大熵分布的一种特例。

1.2 骤淤量重现值

若已知骤淤量年最大值序列 x_1, x_2, \dots, x_n ,其独立同分布于随机变量 X ,可选用合适的分布 $F(x)$ 对该序列进行拟合。分布中各参数的估计方法有矩法、极大似然法和适线法等^[9],这里统一选用极大似然法求解。

骤淤量重现期 N 是一个统计平均的概念,它并不表示 N 年内必定会发生这一强度的骤淤现象。若 $N=1/p$,且 $1-F(x_p)=p$,则 N 年一遇骤淤量可由随机变量 X 的上侧分位数 x_p 表出。

若分布 $F(x)$ 的未知参数为 θ ,其参数估计值为 $\hat{\theta}$,则 N 年一遇骤淤量重现值的点估计 \hat{x}_p 可表示为:

$$\hat{x}_p = x_p(\hat{\theta}) = F^{-1}(1-p; \hat{\theta}) = F^{-1}\left(1 - \frac{1}{N}; \hat{\theta}\right) \quad (4)$$

式中: $F^{-1}(\cdot)$ 为分布函数 $F(\cdot)$ 的反函数。

2 基于极大似然法的水重现值区间估计

Coles基于极大似然方法,给出了广义极值分布分位数的区间估计^[9]。本文将该方法应用于二参数对数正态分布和最大熵分布中。

设随机变量 X 的分布函数为 $F(x; \theta), x_1, x_2, \dots, x_n$ 是它的一个样本, $l(\theta)$ 和 $\hat{\theta} = (\hat{\theta}_1, \dots, \hat{\theta}_d)$ 分别表示对数似然函数和分布参数 $\theta = (\theta_1, \dots, \theta_d)$ 的极大似然估计。若渐近协方差矩阵为 $\mathbf{V}_\theta, x_p = x_p(\theta)$ 为

上侧分位数 x_p 与参数 θ 的函数关系式(式(4)), 则 x_p 的极大似然估计 $\hat{x}_p = x_p(\hat{\theta})$ 满足

$$\hat{x}_p \sim N(x_p, V_{x_p}) \tag{5}$$

式中: $V_{x_p} = \nabla x_p^T \mathbf{V}_\theta \nabla x_p$, ∇x_p 取 $\left(\frac{\partial x_p}{\partial \theta_1}, \dots, \frac{\partial x_p}{\partial \theta_d}\right)$ 在 $\hat{\theta}$ 的估

计值; $\mathbf{V}_\theta = [\mathbf{I}_E(\theta)]^{-1}$, 且 $\mathbf{I}_E(\theta) = \left[E\left(\frac{\partial^2 l(\theta)}{\partial \theta_i \partial \theta_j}\right)\right]_{d \times d}$ (Fisher 信息阵), 一般常用观测信息阵 $\mathbf{I}_0(\theta) = \left[\frac{\partial^2 l(\theta)}{\partial \theta_i \partial \theta_j}\right]_{d \times d}$ 代替。于是 $\hat{x}_p \pm u_{\alpha/2} \sqrt{|V_{x_p}|}$ 即为 x_p 置信水平为 $1 - \alpha$ 的近似置信区间, 其中 $u_{\alpha/2}$ 为标准正态分布的上侧 $\alpha/2$ 分位数。

对于不同的分布类型, 置信区间的求解关键是给出关系式 $x_p = x_p(\theta)$ 、参数 θ 的极大似然估计值 $\hat{\theta}$ 以及 V_{x_p} 。其中参数估计 $\hat{\theta}$ 可据分布拟合得到, 下面分别给出二参数对数正态分布和最大熵分布的关系式 $x_p = x_p(\theta)$ 和 V_{x_p} 。

2.1 二参数对数正态分布

相应的对数似然函数 $l(\theta) = l(\mu, \sigma)$ 为:

$$l(\mu, \sigma) = - \sum_{i=1}^n \left[\ln x_i + \ln \sigma + \frac{1}{2} \ln(2\pi) + \frac{(\ln x_i - \mu)^2}{2\sigma^2} \right] \tag{6}$$

故

$$\mathbf{I}_0(\theta) = \begin{bmatrix} \frac{\partial^2 l(\theta)}{\partial \mu^2} & \frac{\partial^2 l(\theta)}{\partial \sigma \partial \mu} \\ \frac{\partial^2 l(\theta)}{\partial \mu \partial \sigma} & \frac{\partial^2 l(\theta)}{\partial \sigma^2} \end{bmatrix}_{(\mu, \sigma) = (\hat{\mu}, \hat{\sigma})} =$$

$$\begin{bmatrix} -\frac{n}{\hat{\sigma}^2} & -\frac{2}{\hat{\sigma}^3} \sum_{i=1}^n \ln(x_i - \hat{\mu}) \\ -\frac{2}{\hat{\sigma}^3} \sum_{i=1}^n \ln(x_i - \hat{\mu}) & \frac{n}{\hat{\sigma}^2} - \frac{3}{\hat{\sigma}^4} \sum_{i=1}^n \ln(x_i - \hat{\mu})^2 \end{bmatrix} \tag{7}$$

另外, 上侧 p 分位数 x_p 与参数间的关系为

$$x_p = \exp[\sigma + \Phi^{-1}(1 - p) + \mu] \tag{8}$$

式中: $\Phi^{-1}(\cdot)$ 为标准正态分布分布函数的逆。因此,

$$\nabla x_p = \left(\frac{\partial x_p}{\partial \mu}, \frac{\partial x_p}{\partial \sigma} \right)^T \Big|_{(\mu, \sigma) = (\hat{\mu}, \hat{\sigma})} =$$

$$e^{\hat{\mu} + \hat{\sigma} \Phi^{-1}(1-p)} \cdot [1, \Phi^{-1}(1 - p)]^T$$

2.2 最大熵分布

相应的对数似然函数 $l(\theta) = l(\beta, \gamma, \xi, a_0)$ 为:

$$l(\beta, \gamma, \xi, a_0) = n \left[\ln \xi + \frac{\gamma + 1}{\xi} \ln \beta - \ln \Gamma\left(\frac{\gamma + 1}{\xi}\right) \right] + \gamma \sum_{i=1}^n \ln(x_i - a_0) - \beta \sum_{i=1}^n (x_i - a_0)^\xi \tag{10}$$

故

$$\mathbf{I}_0(\theta) = \begin{bmatrix} \frac{\partial^2 l}{\partial \beta^2} & \frac{\partial^2 l}{\partial \beta \partial \gamma} & \frac{\partial^2 l}{\partial \beta \partial \xi} & \frac{\partial^2 l}{\partial \beta \partial a_0} \\ \frac{\partial^2 l}{\partial \gamma \partial \beta} & \frac{\partial^2 l}{\partial \gamma^2} & \frac{\partial^2 l}{\partial \gamma \partial \xi} & \frac{\partial^2 l}{\partial \gamma \partial a_0} \\ \frac{\partial^2 l}{\partial \xi \partial \beta} & \frac{\partial^2 l}{\partial \xi \partial \gamma} & \frac{\partial^2 l}{\partial \xi^2} & \frac{\partial^2 l}{\partial \xi \partial a_0} \\ \frac{\partial^2 l}{\partial a_0 \partial \beta} & \frac{\partial^2 l}{\partial a_0 \partial \gamma} & \frac{\partial^2 l}{\partial a_0 \partial \xi} & \frac{\partial^2 l}{\partial a_0^2} \end{bmatrix}_{(\beta, \gamma, \xi, a_0) = (\hat{\beta}, \hat{\gamma}, \hat{\xi}, \hat{a}_0)} \tag{11}$$

另外, 上侧 p 分位数 x_p 与参数间的关系为

$$x_p = x_p(\beta, \gamma, \xi, a_0) = \left[\frac{1}{\beta} F_G^{-1}(1 - p; \frac{\gamma + 1}{\xi}, 1) \right]^{\frac{1}{\xi}} + a_0 \tag{12}$$

式中: $F_G(x; a, b)$ 表示参数为 a 和 b 的 Gamma 分布的分布函数。

对固定的概率值 p , $y_p = F_G^{-1}\left(1 - p; \frac{\gamma + 1}{\xi}, 1\right)$ 和参数 γ 、 ξ 的关系式(此处 γ 、 ξ 可在 $\hat{\gamma}$ 和 $\hat{\xi}$ 附近取值)可通过线性拟合得到。假设此关系式表为

$$F_G^{-1}\left(1 - p; \frac{\gamma + 1}{\xi}, 1\right) \cong c_0 + c_1 \gamma + c_2 \xi \tag{13}$$

式中: $c_i (i = 1, 2, 3)$ 为线性系数。因此,

$$\nabla x_p = \left(\frac{\partial x_p}{\partial \beta}, \frac{\partial x_p}{\partial \gamma}, \frac{\partial x_p}{\partial \xi}, \frac{\partial x_p}{\partial a_0} \right)^T \Big|_{(\beta, \gamma, \xi, a_0) = (\hat{\beta}, \hat{\gamma}, \hat{\xi}, \hat{a}_0)} = [-d_2 y_p t^k, d_1 c_1 t^k, (d_1 c_2 - d_3 \ln t) t^k, 1]^T \tag{14}$$

式中: $t = \hat{\beta}^{-1} y_p$; $k = \hat{\xi}^{-1} - 1$; $d_1 = (\hat{\xi} \hat{\beta})^{-1}$; $d_2 = (\hat{\xi} \hat{\beta}^2)^{-1}$; 且 $d_3 = (\hat{\xi}^2 \hat{\beta})^{-1}$ 。

3 黄骅港外航道骤淤量重现值的统计分析

根据风况可以推导出不同的骤淤量统计模型, 包括以变化过程、含沙量和有效风能为基础建立的计算公式。侯志强^[4]利用依据有效风能概念建立的风波淤积公式, 计算出了黄骅港整治前年最大骤淤量的序列(1979—2003年), 见图1。本文

利用该序列进行黄骅港外航道骤淤量重现值的统计分析。

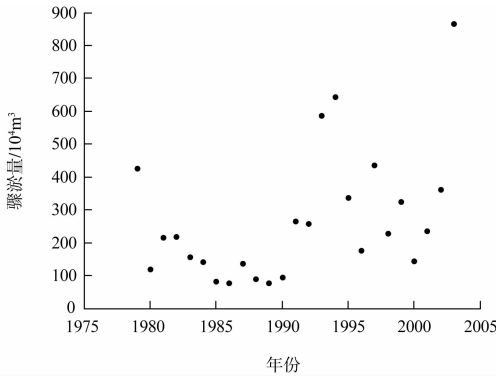
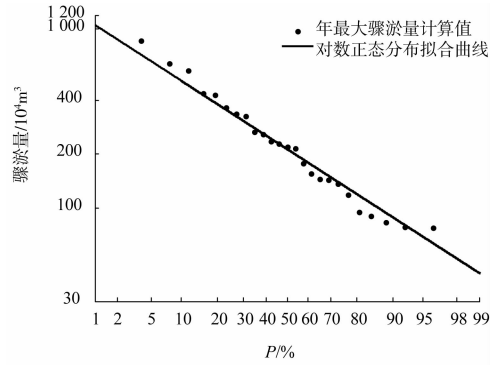


图1 1979—2003年黄骅港外航道年最大骤淤量

3.1 最大骤淤量的概率分布拟合

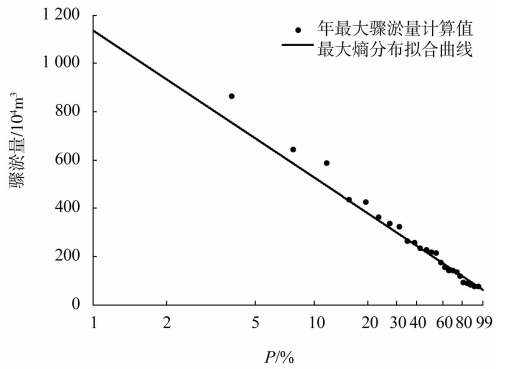
选用对数正态分布和最大熵分布拟合黄骅港年最大骤淤量序列, 结果见表1及图2。

由表1中K-S检验结果可知, 两种分布的检验统计量 D_n 均小于 $D_n(\alpha)$, 因此这两种拟合均通过检验。而拟合曲线与数据序列的离差平方和结果表明, 最大熵分布的拟合误差(0.0093)稍小于对数正态分布误差(0.0131)。



注: 纵坐标为指数坐标。

a) 对数正态分布拟合



b) 最大熵分布拟合

图2 黄骅港外航道年最大骤淤量拟合曲线

表1 黄骅港外航道年最大骤淤量分布拟合及检验结果

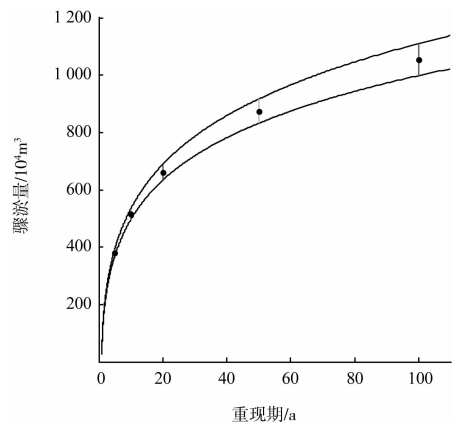
分布类型	参数估计值					K-S 检验 ($n=25, \alpha=0.05$)		离差平方和检验 Q	
	$\hat{\mu}$	$\hat{\beta}$	$\hat{\gamma}$	$\hat{\sigma}$	$\hat{\xi}$	\hat{a}_0	$D_n(\alpha)$		D_n
对数正态分布	5.363 6			0.686 5			0.264 1	0.080 9	0.013 1
最大熵分布		2.838 6	2.969 2		0.300 0	48.118 7	0.264 1	0.093 3	0.009 3

两种拟合得到的多年一遇骤淤量重现值见表2(区间外的值), 结果说明当重现期大于(等于)10 a时, 最大熵分布推算的骤淤量重现值高于对数正态分布结果。

表2 黄骅港外航道多年一遇骤淤量重现值点估计和区间估计

重现期/a	骤淤量/ 10^4 m^3	
	对数正态分布	最大熵分布
5	381 [368, 392]	377 [370, 384]
10	515 [496, 534]	525 [515, 534]
20	660 [633, 688]	689 [676, 701]
50	874 [832, 917]	932 [915, 948]
100	1 054 [998, 1 110]	1 136 [1 117, 1 155]

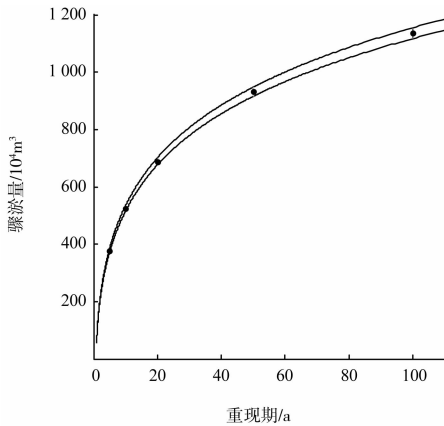
用对数正态分布和最大熵分布的求解公式, 给出黄骅港外航道骤淤量重现值的置信区间($\alpha=0.05$), 结果见表2(区间中的数据)及图3。



a) 对数正态分布结果

3.2 骤淤量重现值的区间估计

根据重现值的极大似然区间估计方法, 分别利



b) 最大熵分布结果

注：图中沿纵轴上部曲线表示置信区间的上限，下部曲线为置信区间下限，圆点表示重现值的点估计结果。

图3 黄骅港外航道骤淤量重现值的区间估计

结果表明，对于相同的骤淤量重现期，对数正态分布求得置信区间的长度明显大于最大熵分布结果。这说明在同一置信水平下，前者区间估计的精度要弱于后者。另外，二者所推导的置信区间在重现期较小时区间长度小，与点估计结果相近；重现期大时区间长度显著增大。这与实际理论相符，因为外推结果必然会产生较大误差。通过比较可知，对于本文中黄骅港外航道重现值的区间估计，最大熵分布所给出的结果较优。

4 结语

1) 最大熵分布对年最大骤淤量序列的拟合误差小于对数正态分布。这说明在已有信息下，制定黄骅港外航道骤淤量的防范标准时，根据最大熵分布进行计算结果较优。

2) 利用极大似然法给出的基于对数正态分布和最大熵分布的区间估计，均能获得黄骅港外航道骤淤量重现值的合理置信区间。

3) 对于同一重现期，基于最大熵分布的置信区间长度小于对数正态分布结果，故其区间估计的精度较高，该结果可对泥沙淤积的整治标准提供参考。

4) 随着重现期的增大，置信区间的长度也随之增大。这说明在现有信息量的条件下，外推的时间跨度越大，误差也会增加，与预期相符。

参考文献：

- [1] 吴今权, 章始红, 孙林云, 等. 粉沙质海岸港区航道减淤工程措施研究[J]. 海洋工程, 2012, 30(1): 98-105.
- [2] 章卫胜, 张金善, 赵红军, 等. 黄骅港外航道寒潮风暴潮及大浪作用下泥沙骤淤数值模拟[J]. 中国港湾建设, 2010(S1): 32-37.
- [3] 季泽舟. 考虑泥沙骤淤影响的航道通航标准研究[J]. 港工技术, 2011, 48(6): 1-4.
- [4] 侯志强. 粉沙质海岸航道淤积重现期研究[M]. 北京: 人民交通出版社, 2013.
- [5] 侯志强, 杨华. 黄骅港外航道骤淤分析[J]. 水道港口, 2010, 25(4): 213-215.
- [6] 季泽舟, 侯志强. 粉沙质海岸港口航道骤淤防治标准研究[J]. 港工技术, 2010, 47(1): 11-13.
- [7] 孔令双, 曹祖德, 李炎保. 利用有效风能预报航道淤积[J]. 水道港口, 2004, 25(4): 209-212.
- [8] Borgman L E. Risk criteria [J]. Journal of the Waterways and Harbors Division, 1963, 89(3): 1-35.
- [9] Coles S G. An introduction to statistical modeling of extreme values [M]. London: Springer-Verlag, 2007.
- [10] Zhang L Z, Xu D L. A new maximum entropy probability function for the surface elevation of nonlinear sea waves [J]. China Ocean Engineering, 2005, 19(4): 637-646.
- [11] Dong S, Tao S S, Lei S H, et al. Parameter estimation of the maximum entropy distribution of significant wave height [J]. Journal of Coastal Research, 2013, 29(3): 597-604.

(本文编辑 武亚庆)