

【水文·泥沙】

遗传算法在 P - III 型分布曲线参数估值中的应用

王占海, 陈元芳, 倪夏梅, 魏琳, 黄琴

(河海大学 水文水资源学院, 江苏 南京 210098)

摘 要:通过加入历史洪水等多种方案进行统计试验, 较为全面地分析了遗传算法在 P - III 型分布曲线参数估值中的应用问题。结果表明: 基于绝对值准则的遗传算法具有良好的不偏性和有效性, 统计性能优于矩法、适线法, 且可以充分利用样本信息来优化参数, 在设计值的不偏性上略优于线性矩法; 基于遗传算法的适线求解中, 绝对值准则优于平方和准则; 遗传算法可以提高 P - III 型分布曲线参数估值的精度, 但尚需在实践中提高该方法的应用价值。

关 键 词: 适线准则; 不偏性; 有效性; 遗传算法; 参数估值; P - III 型曲线

中图分类号: P333

文献标识码: A

doi: 10.3969/j.issn.1000-1379.2009.09.010

P - III 型分布曲线是我国水文计算规范中指定采用的分布线型, 其常用的参数估值方法有矩法、权函数法、概率权重法、适线法、线性矩法、最大熵值法等。优化适线法因其简单实用、对资料拟合程度较好而被广泛采用。丛树铮等^[1]将统计试验应用于水文频率分析计算, 证明了绝对值准则对适线估计参数较好, 但该方法基于模型搜索, 有时会得到局部最小解。近些年来, 随着计算机技术和智能算法的发展, 遗传算法(GA)被引入到水文频率优化适线中, 金菊良^[2]采用基于加速遗传算法的绝对值适线准则来估计 P - III 分布曲线参数, 姜铁兵等^[3]采用平方和准则估计, 均说明了遗传算法优于常规的适线法; 康玲^[4]采用基于遗传模拟退火算法的加权适线法, 证明该方法提高了遗传算法全局寻优能力。

上述研究成果表明, GA 使 P - III 型曲线的参数估值更为稳定和可靠, 但是仅通过一次实测资料来检验, 且未考虑加入历史洪水后的不连续系列, 表现不了该方法的统计性能。此外, GA 从计算原理上区别于常规的优化方法, 其优化目标函数(即绝对值准则)是否适用仍需研究。基于此, 笔者根据 Monte - Carlo 统计试验探讨了 GA 适应性和相应的优化适线准则。

1 基于 GA 的优化适线参数估值方法

1.1 P - III 型分布适线法

P - III 型分布密度函数为

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} (x - a_0)^{\alpha-1} e^{-\beta(x-a_0)} \quad (x > a_0) \quad (1)$$

式中: a_0 、 α 、 β 为 P - III 型分布密度函数的参数, $\alpha > 0$, $\beta > 0$ 。

a_0 、 α 、 β 与统计参数 E_x 、 C_V 、 C_S 的关系为

$$E_x = (\alpha/\beta) + a_0, C_V = \sqrt{\alpha}/(\alpha + \beta a_0), C_S = 2/\sqrt{\alpha} \quad (2)$$

优化适线法通过优选 E_x 、 C_V 、 C_S , 使得理论频率曲线与经验点距的拟合程度最好, 即拟合曲线与经验点距距离差最小, 度量该离差有多种指标, 一般形式为

$$f = \sum_{i=1}^n |x_{pi} - x_{pi}^*|^m \quad (3)$$

通常取 $m = 1$ 或 2 , $m = 1$ 时称为优化适线绝对值准则, $m = 2$ 时为平方和准则。设计值 $x_p = E_x(\varphi_p C_V + 1)$, 其中 φ_p 为离均系数, p 为频率, 可查用已编制好的 $C_S - p - \varphi_p$ 关系数字表。目前, 有许多数值方法^[5]来求解 φ_p , 但是各类数值计算仅适用于特定范围内的 C_S 值, 不具有通用性, 故笔者仍采用插值 $C_S - p - \varphi_p$ 表来编程计算, 采用二元三次样条曲线插值^[6]计算离均系数值以提高精度。

1.2 GA 优化适线

遗传算法的基本原理和算法详见文献[7-8], 这里仅介绍遗传算法优化适线的基本步骤:

(1) 参数编码。 E_x 用矩法估计的精度较好, 一般对 C_V 和 C_S 优化, 对两者进行二进制编码, 把矩法估计的 C_V 和 C_S 作为遗传算法优化初值。

(2) 适应度函数的选择。根据具体需要选择目标函数——绝对值优化准则或平方和优化准则。

(3) 确定遗传算法的初始参数。初始种群规模 $N = 100$, 交叉概率 $P_c = 0.8$, 变异概率 $P_m = 0.02$ 。

(4) 根据群体规模随机生成初始种群, 并计算每个个体的适应度 f , 按轮盘赌方法进行选择, 被选中的个体放入配对库中。

(5) 交叉和变异操作。通过该操作产生优秀的新一代个体。

(6) 确定终止条件。设定算法最大重复执行的次数为 500, 适应度限 (Fitness Limit) 为 10^{-5} , 若已经找到达到终止条

收稿日期: 2008-10-25

基金项目: 水利部公益性行业科研专项(200801076)。

作者简介: 王占海(1983—), 男, 河北秦皇岛人, 硕士研究生, 研究方向为随机水文与工程水文。

通讯作者: 陈元芳(1963—), 男, 教授。

E-mail: yfchen@gmail.com

件的最优参数 C_V 和 C_S , 则结束, 否则返回步骤(5), 直到找到最优解为止。

2 GA 优化适线与常规优化适线的比较

GA 与传统优化算法不同, 不依赖于梯度信息, 而是通过模拟自然进化过程来搜索最优解, 具有群体搜索特性。该方法采用同时处理群体中多个个体的方法, 即同时对搜索空间中多个解进行评估, 这一特点使 GA 具有较好的全局搜索性能, 因而在参数估值方面表现出更好的稳定性和可靠性。

2.1 寻找最优解的能力

随机生成若干组 P-III 型随机数来比较 GA 与常规优化算法在优化适线中的区别。令 P-III 型随机数总体分布 $E_x = 100$ 、 $C_V = 1$ 、 $C_S = 3$, 由于实际工程中的资料一般不会太长, 因此生成样本长度为 $n = 50$ 。按照绝对值优化准则(不同优化准则的选取问题在统计试验中加以论证)对生成的随机数按上面两种方法分别估计参数, 发现两者结果有时很接近, 有时相差较大, 原因是基于模型搜索法的优化适线在有些情况下会陷入局部极小点, 而群体搜索的 GA 可以找到全局最小点。

2.2 运算效率

GA 具有优秀的寻找全局最优解的能力, 但是作为一种随机群体搜索优化方法, 其运算效率也是一个值得考虑的问题, 如果耗费的时间很长, 效率很低, 那么具有再高的优化能力也是没有实际应用价值。为验证其效率, 在 CPU 为 AMD Turion 64 X2(主频 1.8 GHz)、内存为 1 G 的电脑上, 随机模拟出 100 组样本容量为 50 的 P-III 型分布随机数, 总体参数与上面相同, 计算优化适线的平均时间为 2.797 s, 而 GA 计算的平均时间为 56.078 s(仅是优化方法的计算时间, 不包括随机数生成、数据读入等时间)。虽然 GA 计算时间明显慢于常规优化算法, 但是计算容量为 50 的样本数据平均时间不到 1 min, 为获得更高精度, 在实际应用中是可以接受的, 加之现今计算机高速发展, 高配置的硬件设备将会提高该种算法的运算效率。

3 统计试验研究

对简单样本和加入历史洪水资料的非简单样本进行统计试验分析, 并与常用的参数估计方法——矩法(MOM)、适线法(FIT)、线性矩法(L-M)相比较。此外, 由于 GA 优化机理与常规优化算法有一定区别, 因此对优化准则——绝对值准则(ABS)和平方和准则(OLS)进行研究, 以较为全面地反映 GA 对 P-III 型分布曲线参数估计的适应性。

(1) 方法优劣评价标准。以统计参数、设计值的不偏性及有效性为依据评价估计方法的有效性^[9]。

(2) 试验方案设计。①简单样本。实测系列长度 n 取 30、50, 总体参数 $E_x = 100$, $C_{V_0} = 0.3, 0.5, 1.0$, $C_{S_0}/C_{V_0} = 2, 3, 4$, 共计 18 个方案。②非简单样本。最大重现期 $N = 100, 200$, 实测期 $n = 30, 50$, 历史洪水个数取 1、3 个, 总体参数 $E_x = 100$, $C_{V_0} = 0.3, 0.5, 1.0$, $C_{S_0}/C_{V_0} = 2, 3, 4$, 共 72 个方案。各方案中设计频率 $p = 1\%, 0.1\%, 0.01\%$, 统计试验次数 N_s 为 1 000。

4 试验结果分析

4.1 GA 优化适线准则比较

(1) 参数估计结果比较。GA(ABS)和 GA(OLS)估计参数 C_V 基本不偏, 两种准则下估值结果较接近, GA(OLS)估值稍偏大, 两者估值的有效性基本相当。从 C_S 估计不偏性来看, GA(OLS)估值略微偏大些, 但当 $C_{V_0}, C_{S_0}/C_{V_0}$ 较大时, GA(OLS)偏离程度会变大, 从整体估值 C_S 的结果来看, GA(OLS)有些偏大, GA(ABS)优于 GA(OLS)。

(2) 设计值估计结果比较。从估计值的不偏性(B_p)来看, 绝大多数方案 B_p 值均能保持在 0 ~ 5% 之内, 只有极少数方案, 如 $C_{V_0} = 1, C_{S_0} = 4, N = 200, n = 30, a = 3$ 时, GA(OLS)的 $B_{p, 0.01\%}$ 达到了 6.98%。对全部 90 个方案进行统计后, 只有 8 个方案偏离程度超过 5%, 最大偏离不超过 8%, 可见这两种估值方法表现出很好的不偏性; 从估计值的有效性来看, GA(ABS)好于 GA(OLS)。

从整体来看, 无论是参数估值还是设计值估计, GA(ABS)和 GA(OLS)的不偏性和有效性都较好, 对参数估计表现出较好的稳定性, 且 GA(ABS)具有更好的参数估计效果, 故绝对值准则对于遗传算法优化适线是较为适用的。

4.2 GA 与传统参数估计方法比较

(1) GA 适线法与传统适线法比较。无论是绝对准则还是平方和准则, GA 估计参数值和设计值的效果均优于基于模型搜索的适线法。从统计试验结果可以看出, GA(OLS)好于 FIT(OLS)的程度大于 GA(ABS)优于 FIT(ABS)的程度, 原因是模型搜索法容易找到局部极小点, 非线性形式的平方和准则相对于线性表达式的绝对值准则更容易陷入局部极小点, 然而 GA 对目标函数和约束条件没有限制, 均可以找到全局最优解, 因此 GA 对非线性形式的目标函数表现出更好的优化能力。

(2) GA 适线法与传统参数估计法比较。①参数估值结果比较。对于 C_V 的估计, MOM 效果较差, L-M、GA(ABS)估值结果较为接近, GA(ABS)略微偏大, 但均优于 FIT(ABS); 有效性方面, L-M、GA(ABS)较好, FIT(ABS)次之, MOM 较差。对于参数 C_S 的估计, GA(ABS)和 L-M 估计效果最好; 有效性方面, GA(ABS)与 L-M 相比时好时坏, 总体来说两者参数估值的不偏性和有效性基本相当。②设计值估计结果比较。GA(ABS)估计的设计值不偏性略优于 L-M 法, FIT 法大多数情况下与 L-M、GA(ABS)较为接近, 但有时偏离较大, 在稳定性上不如 GA(ABS)、L-M 好。GA(ABS)、L-M 和 FIT(ABS)的有效性基本相当, MOM 法较差。

总之, 从参数估值和设计值估计效果来看, GA(ABS)的参数估值较线性矩估值大些。由于对资料拟合较好, 因此不偏性略优于 L-M, 而 FIT(ABS)效果稍差。

5 实例分析

为了验证各种参数估计方法对实测资料的适用性, 选取长江上游干滩站进行频率分析计算。该站集水面积为 866 559 km², 是长江干流较为重要的控制站。选取了 1893 ~ 2004 共计

12年最大洪峰流量资料,分别采用以上方法进行参数估值,结果见表1。

表1 不同参数估计方法对寸滩站最大洪峰流量的估计结果

项目	MOM	L-M	GA(ABS)	FIT(ABS)
C_v	0.233 0	0.241 3	0.241 1	0.241 0
C_s	0.557 1	0.619 9	0.626 6	0.616 8
$f(\text{ABS})/(\text{m} \cdot \text{s}^{-1})$	160 410	119 740	115 250	126 560

通过表1可以看出,GA(ABS)优于FIT(ABS),略好于线性矩法,矩法估值效果较差,因此GA(ABS)估计的参数值更稳定、可靠。

6 结 语

(1)通过加入历史洪水等多种方案进行统计试验,较为全面地分析了GA在P-Ⅲ型分布曲线参数估值方面的应用问题。试验表明,GA(ABS)具有良好的不偏性和有效性,统计性能优于其他传统估值方法(矩法、适线法),且可以充分利用样本信息来优化参数,在设计值的不偏性方面略优于线性矩法。

(2)GA作为一种随机群体搜索算法,可以找到全局最优解,基于GA的适线求解中,绝对值准则优于平方和准则。

(3)经分析验证,GA计算速度慢于模型搜索优化适线,但计算50个样本数据平均机时不到1 min,加之现今计算机硬件设备的高速发展,使得该方法具有实际的应用价值。

GA可以提高P-Ⅲ型分布曲线参数估值的精度,但尚需将该方法应用到更多的人文实测序列之中,并对其他线型,如极值分布、对数正态分布进行研究,在实践中提高该方法的应用价值。

参考文献:

- [1] 丛树铮,谭维炎,黄守信,等.水文频率计算中的参数估计方法的统计试验[J].水利学报,1980(3):1-15.
- [2] 金菊良.遗传算法及其在水问题中的应用[D].南京:河海大学,1998.
- [3] 姜铁兵,康玲,沈同林,等.GA适线法软件设计在水文频率计算中应用[J].华中理工大学学报,1999(5):52-53.
- [4] 康玲,陈辉,沈同林.基于遗传模拟退火算法的洪水频率计算研究[J].广西水利水电,2003(2):37-40.
- [5] 王建刚,刘亚萍.P-Ⅲ型分布 ϕ 值数值计算方法比较[J].水文,1997(5):10-14.
- [6] 冯果忱,黄明游.数值分析[M].北京:高等教育出版社,2007.
- [7] 苑希民,李鸿雁,刘树坤,等.神经网络和遗传算法在水科学领域的应用[M].北京:中国水利水电出版社,2002.
- [8] 付玉娟,蔡焕杰,张旭东,等.基于遗传算法的树状灌溉管网优化设计[J].人民黄河,2006,28(7):42-44.
- [9] 陈元芳.统计试验方法及应用[M].哈尔滨:黑龙江人民出版社,2000.
- [10] 李娟,陈元芳,王文鹏,等.最大嫡值原理在P-Ⅲ型分布参数估计中的应用研究[J].水电能源科学,2007(5):31-35.

【责任编辑 翟成亮】

(上接第20页)不同粒径组沙含量越均匀,形成的临界粗化层剪切流速 u_{*cr} 越大,即对水流的抗冲能力越强。

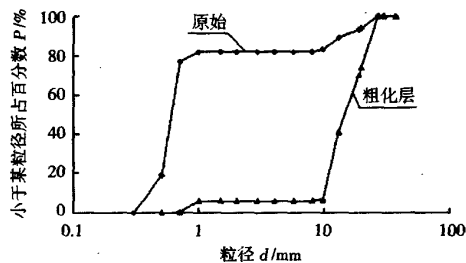


图3 混合沙3粗化层粒径级配随剪切流速变化情况

(2)缺少中等粒径的床沙也能形成粗化层,但较大粒径间的孔隙粗沙很难有效地充填进去,这种粗化层厚度较大,且对下层原床沙的保护效果较差^[6]。

(3)缺少中等粒径的床沙需要冲走更多的细沙才能形成粗化层。由此可见,临界粗化层的形成需要特殊的床沙组成条件。临界粗化层犹如一把标尺,当床沙粗化程度还未达到临界粗化层状态时,床面还有承受和抵挡更大水流冲刷的能力;如果来流强度超过临界粗化层稳定条件,伴随粗化层彻底破坏河床将发生强烈冲刷。

河流水沙条件的改变,尤其是水库枢纽下游河床发生冲刷粗化,对河流开发利用、堤防保护、取水及桥梁等涉河工程的安全,必将产生重大的影响。应深入研究粗化层形成、发展和稳

定机理,为生产实际服务。

参考文献:

- [1] 韩其为,向熙琰,王玉成.床沙粗化[C]//第二次河流泥沙国际学术讨论论文集.北京:水利电力出版社,1983:356-367.
- [2] 秦荣昱,王崇浩.河流推移质运动理论及应用[M].北京:中国铁道出版社,1996.
- [3] 陆永军.宽级配河床粗化稳定结构[G]//水科学青年学术论文集(一).北京:水利电力出版社,1990:296-303.
- [4] 杨具瑞,徐一民,谭毅源.粗化层破坏的试验研究和机理分析[J].水力发电学报,2008,27(3):115-119.
- [5] 孙志林,孙志峰.粗化层试验与预报[J].水力发电学报,2000,21(4):40-48.
- [6] Chin C O, Melville B W, Raudkivi A J. Streambed Armoring[J]. Journal of Hydraulic Engineering, 1994, 120(8): 899-918.
- [7] 尹学良.黄河下游的河性[M].北京:中国水利水电出版社,1995.
- [8] 张瑞瑾,谢鉴衡,陈文彪.河流动力学[M].武汉:武汉大学出版社,2007.
- [9] Lane E W, Carlson E J. Some factors affecting the stability of canals constructed in coarse granular material[C]//Proc. 5th Congress, IAHR. University of Minnesota, Minneapolis, Minn., 1953:37-38.
- [10] Lane E W, Carlson E J. Some observations on the effect of shape and movement of coarse sediments[J]. Trans. Amer. Geophysical Union, 1954(35): 453-462.

【责任编辑 栗志】