文章编号: 1000 - 582X(2005)04 - 0082 - 04

P - Ⅲ型分布参数估计的改进混合遗传优化适线法*

任伯帜[1,2],龙腾锐[1]
(1.重庆大学 城市建设与环境工程学院,重庆 400030; 2. 长沙理工大学 河海工程学院,湖南 长沙 410076)

摘 要:选用多元插值原理对 $\Phi-(p, c_s)$表进行细化和扩展,对传统遗传算法加以改进。在此基础上提出 P - Ⅲ型分布参数估计的混合遗传优化适线法。通过研究表明:该法简易通用、适应性强,可精确地对 $\bar{x}$、c_v、c_s 3个参数进行估计;克服了适线法未能解决 $\bar{x}$,特别是 c_v的估计精度,以及计算过程与结果受人为因数的影响等不足,为准确估计 P - Ⅲ型分布参数提供了一种十分有效的方法。

关键词: P - Ⅲ型分布;参数估计;适线法;混合遗传算法
中图分类号: TU832.6 **文献标识码**: A

城市暴雨强度公式是地区暴雨管理的重要基础模型,是计算暴雨地面径流和确定城市排水工程设计流量的重要依据。根据文献[1]规定:编制城市暴雨强度公式时,一般要对实测暴雨统计资料进行调整,而 P - Ⅲ型分布作为我国水文分析计算规定的线型,长期以来在其调整中得到广泛应用。对 P - Ⅲ型分布通过变换可得如下理论计算模型[2]:

$$x_p(p;\bar{x}, c_v, c_s) = (1+\Phi c_v)\bar{x} \qquad (1)$$

式(1)中:x_p为理论计算降雨强度 (mm/min);$\bar{x}$、c_v和Φ分别为选取样本的均值、离差系数和离均系数;Φ是关于频率 p和选取样本的偏差系数 c_s的函数,在实际应用中,若已知 p和 c_s,可由 $\Phi-(p, c_s)$表查得对应的Φ值。显然,该方法应用的一个关键问题就是 P - Ⅲ型分布参数 $\bar{x}, c_v, c_s$的合理估计。许多年来,我国水文工作者对其参数估计问题作了大量研究工作,先后提出矩法、适线法、极大自然法、权函数法、密度函数法等[3-4]。适线法是一种优选法,即优选一种组参数 $(\bar{x}, c_v, c_s)$使之相应的理论曲线与经验点拒最好吻合。适线法尤其是试错适线法在实际工程中应用较多[2],但由于适线法较多针对 c_s的调整,对 $\bar{x}$,特别是 c_v的估计精度考虑较少,在全面精确地估计 P - Ⅲ型分布参数方面还存在一些欠缺;其次,适线过程带有一定的盲目性和不确定性,受计算者的实际经验和判断水准的制约,结果因人而异,往往达不到优选参数目的;再者,适线法中离均系数 Φ人为直线内插,计算烦琐,精度不高,从而影响 3个参数的精确估计。随后,许多学者将计算机及优化技术应用于适线法中,提出模型搜索适线法、模糊加权优化适线法等[5]。这些方法虽避免了适线法中存在的一些缺点,适当提高了参数估计的精度,但在增强参数估计方法的精确性和减少人为因素影响等方面还存在局限性。目前,迅速发展起来的遗传算法 (Genetic Algorithm,简称 GA),因不受模型结构、优化准则、参数数目、约束条件等的限制,具有直观、简便、通用、适应性强、全局寻优等特点,引起人们广泛的重视[6],但如何将遗传算法技术应用于适线法中,寻求一种简易通用、适应性强、不受人为因素影响且精度较高的自适应优化适线法,全面优选参数 $(\bar{x}, c_v, c_s)$,是值得探讨的问题。鉴于此,作者,通过计算机辅助,应用多元插值理论和改进混合遗传算法联合估计参数,提出经验点与所配曲线拟合的混合遗传优化适线法,以期弥补现行优化适线法的不足。

1 改进混合遗传优化适线法的构建

1.1 改进混合遗传算法

许多实际工程优化问题常常是维数高,具有连续的解空间。对这类问题若采用二进制编码的基本遗传算法 (BGA)[6],必然使得编码长度很长,导致搜索空间的增大,且编码和解码还要耗费大量的时间,使得遗

* **收稿日期**: 2004 - 12 - 12
基金项目:国家自然科学基金重点资助项目 (59838300);湖南省自然科学基金 (03jjy6020)资助
作者简介:任伯帜 (1968 -),男,四川南充人,重庆大学博士研究生,主要从事废水处理及雨水资源化利用研究。

© 1994-2008 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

传算法的效率大大降低;若采用浮点编码的基本遗传算法,因浮点编码相当于使用十进制位串进行编码,且它采用概率规则加适应度函数的方法进行群体搜索,而不是沿梯度方向搜索,所以不能保证每一步搜索都能取得向前进化的良好效果。实际上,不需待求问题的梯度信息是遗传算法的重要特点,也是与使用梯度信息的常规优化方法最本质的差别之一,也正因如此,导致其收敛速度较慢,相反,一些采用梯度信息的常规优化方法尽管容易陷入局部极值区域,但能以很快的速度收敛[7-9]。由于实际工程中往往是多变量多峰值问题,其目标函数常常不可导,这使得梯度信息根本无法使用,但由于梯度的实质就是目标函数值在某点产生最大的变化。因此,结合三元浮点编码法,并在 BGA算法中引进类似梯度信息的高斯变异下降方向因子,在保证算法的通用性的同时,加快算法的收敛率,提高算法的效率,从而构造出改进混合遗传算法。

为阐述改进混合遗传算法,首先给出下面 2个定义[10]:

定义 1　给定一维空间中任意两点 x^{t+1} 和 x^{t+2},如果

$$\vee(x^{t+1}) = \mathrm{sing}(x^{t+1} - x^{t}) = \begin{cases} 1 & \text{当 } x^{t+1} > x^{t} \\ 0 & \text{当 } x^{t+1} = x^{t}, \\ -1 & \text{当 } x^{t+1} < x^{t} \end{cases}$$

称 $\vee(x^{t+1})$ 为 x^{t+1} 对 x^{t} 的变化方向。

定义 2　对求最小值的优化问题 $\min f(X)$ ° $X = [x_1, x_2, x_3, \cdots, x_n]^{\mathrm{T}}$,其中在 n维空间中某一点 $x^{t+1} = [x_1^{t+1}, x_2^{t+1}, x_3^{t+1}, \cdots, x_n^{t+1}]^{\mathrm{T}}$,如果它在其它任意点 $x^{t} = [x_1^{t}, x_2^{t}, x_3^{t}, \cdots, x_n^{t}]^{\mathrm{T}}$ 有:

(1) 当 $f(x^{t+1}) < f(x^{t})$ 时,记为: $\vartheta(x^{t+1}) = [\vee(x_1^{t+1}), \vee(x_2^{t+1}), \vee(x_3^{t+1}), \cdots, \vee(x_n^{t+1})]$;

(2) 当 $f(x^{t+1}) \geq f(x^{t})$ 时,记为: $\vartheta(x^{t+1}) = 0$;

称 $\vartheta(x^{t+1})$ 为函数 $f(x)$ 在点 x^{t+1} 的下降方向因子。

1.1.1　编码方法

个体编码采用 $s = [X, \vartheta(x), \alpha]$,其中 s表示群体的个体,它由 3部分组成, X为待优化变量参数的实际值组成的向量, $\vartheta(x)$ 为方向下降因子, α为特征因子,表示下降方向因子的特征,如果 $\vartheta(x) = 0$, $\alpha = 0$;否则 $\alpha = 1$。这种个体编码方法,既减少了因编码和解码所耗费的计算时间,又避免了编码长度与求解精度之间的矛盾。虽然在编码中引入了下降方向因子和特征因子,但 $\vee(x_i) \in \{1, 0, -1\}$, $\alpha \in \{0, 1\}$,它们的变化范围小,所以对编码的长度不会产生很大的影响;同时 $\vartheta(x)$ 的计算简单,对运算时间影响也非常有限。之所以在编码中引入 $\vartheta(x)$,正是要将当前群体或个体的某种有利于进化的趋势继续保留和加强,下降方向因子记录了进化趋势:当 $\alpha \neq 0$ 时, $\vartheta(x)$ 记录的是当前个体在由父代个体进化而来时的变化方向,而这一变化正是使问题进一步优化的方向。由于下降方向因子的计算对适应度函数本身没有任何影响,所以对遗传算法的通用性没有任何影响。

1.1.2　遗传变异

MGA与 BGA的变异区别较大,MGA的变异不再同 BGA那样在个体的每一位上进行,而是以参数为单位进行,变异不再是简单地改变编码位上的值,而是通过某种运算得到被变异参数的新值。由浮点编码的特点知,杂交点只能选在各参数之间,单靠杂交是不能给群体引进新的参数值,只有靠变异在解空间进行全局搜索,所以应加强变异的作用,此时变异成为 MGA的主要成分。采用基于下降方向因子的高斯变异方法,其变异规则如下:

若 $\alpha_j^t = 1$, 则: $\delta^i = \xi_1 \cdot f(s_i^t) / \sum_{j=1}^{N_{\mathrm{pop}}} f(s_j^t)$, $x_{i,j}^{t+1} = x_{i,j}^{t} + \vee(x_{i,j}^{t}) \cdot |N(0, \delta^i)|$;

否则: $\delta^i = \xi_2 \cdot f(s_i^t) / \sum_{j=1}^{N_{\mathrm{pop}}} f(s_j')$, $x_{i,j}^{t+1} = x_{i,j}^{t} + N(0, \delta^i)$;

式中: s_i^t为第 t群体中的第 t个个体, $i = 1, 2, 3, \cdots, N_{\mathrm{pop}}$; N_{pop}　N_{pop}为群体规模; $x_{i,j}^{t}$为 s_i^t中的第 j个待优化参数, $j = 1, 2, 3, \cdots k$, k为参数总数; $\vee(x_{i,j}^{t})$为该参数在 s_i^t点的变化方向; $\vartheta(s_i^t)$即为适应度函数在点 s_i^t的下降方向因子; $|\cdot|$表示绝对值; $\vee(x_{i,j}^{t}) \cdot |N(0, \delta^i)|$表示沿 $\vee(x_{i,j}^{t})$ 方向的一个高斯分布随机变量; ξ_1 和 ξ_2 为正的权系数。

对于每个个体 s_i^t,按上述方法进行变异后,便可得到向前进化的一个新个体 s_i^{t+1}。

1.1.3　遗传杂交与精英选择

在进化过程中,当变异已不能使群体进一步优化时,可采用杂交使整个群体进化。由于此时群体已不在进化,即个体的下降方向因子及特征因子都已为 0,所以杂交只需在个体编码段上进行既可。用以进化杂交的父代个体采用随机概率 p_c从匹配集中成对选取,假设从第 t代群体的匹配集中已选择好待杂交的 2个父代个体为: $s_v^t = (v_1, v_2, v_3, \cdots, v_k)$ 和 $s_w^t = (w_1, w_2, w_3, \cdots, w_k)$,则子代个体的参数由父代个体的参数加权求和产生,即: $s_v^{t+1} = \zeta s_v^t + (1 - \zeta) s_w^t$, $s_w^{t+1} = \zeta s_w^t + (1 - \zeta) s_v^t$。式中 ζ为 0～1间的常数。为保证优良个体不会丢

© 1994-2008 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

失,加快算法收敛,本文采用精英选择策略,即当父代群体中的个个体生成个子代个体后,从这 2个个体中选择最好的个个体作为新一代群体,进行进化操作,这样可较好地避免群体的早熟。

利用以上改进措施,结合 BGA算法[7],改进混合遗传算法便可实现。

1.2 优化适线的构成及实现

1.2.1 (p, c_s) - Φ值表细化及扩展

用 P-Ⅲ型分布拟合时,首先要知不同设计频率 p 和不同 c_s 对应的 Φ值。为此本文选用多元插值原理将文献[2]中 Φ - (p, c_s) 附表细化和扩展,得出混合遗传优化适线所需引用的 Φ - (p, c_s) 表,具体方法如下[10]:

设 p为 x方向,c_s为 y方向,Φ为插值点对应的函数值 $z(x, y)$。已知函数 $z(x, y)$ 的第一个变量 x的节点为 x_i(不一定等距,$i = 0, 1, 2, \cdots, n$),第二变量 y的节点为 y_j(不一定等距,$j = 0, 1, 2, \cdots, m$),其对应节点上的函数值为 z_{ij}($i = 0, 1, 2, \cdots, n$; $j = 0, 1, 2, \cdots, m$)。对于给定的不是节点的值 (x, y),分别选取靠近 x的 3个点 (x_q, x_{q+1}, x_{q+2}) 和靠近 y的 3个点 (y_u, y_{u+1}, y_{u+2}),按以下式(2)计算出相应的 $z(x, y)$,即可得所需的 Φ - (p, c_s) 表。显然,插值点分布越均匀越密,其对应的函数值相差越小,插值结果越精确。

$$z(x, y) = \sum_{i=q}^{q+2}\sum_{j=u}^{u+2}\left(\prod_{\substack{\theta=q\\ \mu\neq i}}^{q+2}\frac{x-x_\theta}{x_i-x_\theta}\right)\left(\prod_{\substack{r=u\\ r\neq j}}^{u+2}\frac{y-y_r}{y_j-y_r}\right) \qquad (2)$$

1.2.2 目标函数及适应度函数

由适线目的可知,为使由 P-Ⅲ型分布曲线与实测值拟合最好,则只需使由 P-Ⅲ型分布曲线的理论计算降雨强度值 x_p (mm/min) 与经验频率点对应的实测降雨强度值 x_l (mm/min) 的离差平方和或数学期望为最小。为此,结合式(1)可得混合遗传优化适线所需目标函数,下式中 n为实测样本容量。

$$\min f(x) = \sum_{l=1}^{n}(x_p - x_l)^2 = \sum_{l=1}^{n}\left[(1+\Phi c_v)\bar{x} - x_l\right]^2 \qquad (3)$$

由式(3)可构造本文混合遗传优化适线所需的适应度函数:

$$\mathrm{eval}(s_i) = 1 - (f_i - f_{\min})/(f_{\max} - f_{\min}) \qquad (4)$$

式(4)中:$\mathrm{eval}(s_i)$ 和 f_i 分别为当前种群第 i个父体的适应度函数值及目标函数值,$f_{\max}$ 和 $f_{\min}$ 分别为当前种群成员的最大和最小计算值。

1.2.3 优化适线的实现

根据上述理论,结合 BGA算法[6],便可实现 P-Ⅲ型分布参数估计的改进混合遗传优化适线法。运行中首先给出参数:群体规模 N_{pop}、遗传变异所需参数 ξ_1 和 ξ_2、杂交概率 p_c 及参数 ζ、最大进化代数 T_{max},并采用最大进化代数作为停止准则,按以下计算流程图 1进行运行,从中选用拟和最好的一组 $(\bar{x}, c_v, c_s)$ 作为 P-Ⅲ型分布参数的计算结果。

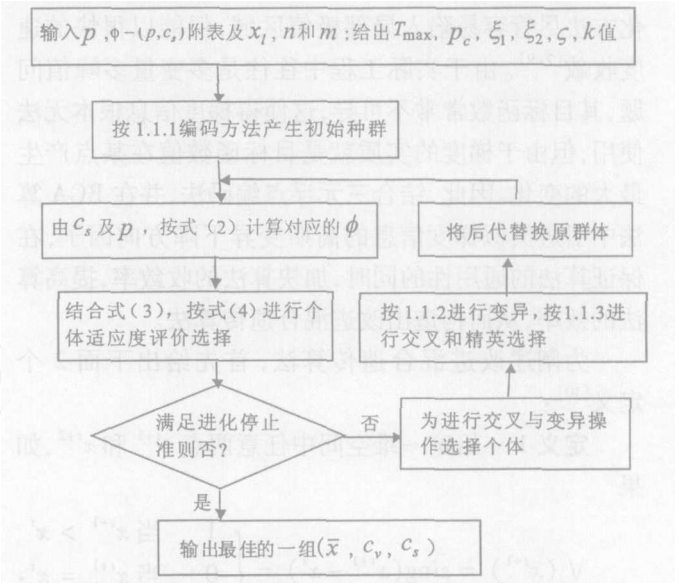

图 1 改进混合遗传播算法适线优化法计算流程

2 实践应用

湖南省长沙地区有连续 35 a(1966 - 2001)的自记录暴雨降雨资料,根据《室外排水设计规范》规定,每场暴雨取 5、10、15、20、30、45、60、90、120 min等 9个历时段的最大降雨强度,采用年多个样法每年取 6个最大值,在每个历时的暴雨降雨强度不分年次从大到小排序,选取其中前 n(n一般取 4倍年份数)个样本作为暴雨频率分析统计的基础资料。经验频率按 $p = M/(1+n)$ 计算(M为序号,n为实测样本容量)。

分别采用矩法、试错适线法及混合遗传优化适线法求得统计参数,其中改进混合遗传优化适线法所取参数为:$N_{pop} = 50$, $\xi_1 = 4.5$和 $\xi_2 = 0.05$, $p_c = 0.9$, $\zeta = 0.03$, $T_{max} = 100$,在上述给定参数下运行 30次,从中选用拟和最好的一组 $(\bar{x}, c_v, c_s)$ 作为计算结果。各种方法计算结果见表 1。根据表 1的参数,由绘图模块在海森机率格纸上绘制出长沙市暴雨公式统计资料 P-Ⅲ型分布 3种计算方法求得的理论累积频率曲线与经验点据的拟合情况,见图 2 ～图 4。

© 1994-2008 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

表 1　不同方法估计 P - Ⅲ型分布参数电算结果

历时 t/min	矩法				试错适线法				混合遗传优化适线法			
	$\bar{x}$	c_v	c_s	$\sigma_{总}$	$\bar{x}$	c_v	c_s	$\sigma_{总}$	$\bar{x}$	c_v	c_s	$\sigma_{总}$
5	1.61	0.20	1.46	0.080	1.61	0.21	1.15	0.079	1.58	0.23	1.13	0.071
10	1.36	0.20	1.07	0.044	1.36	0.21	1.25	0.039	1.33	0.22	1.22	0.035
15	1.21	0.19	0.89	0.039	1.21	0.20	1.51	0.035	1.20	0.21	1.03	0.032
20	1.10	0.18	0.86	0.036	1.10	0.19	1.10	0.033	1.08	0.21	1.04	0.033
30	0.95	0.16	0.48	0.034	0.95	0.17	0.63	0.033	0.92	0.19	0.41	0.028
45	0.79	0.14	0.30	0.033	0.79	0.14	0.33	0.032	0.77	0.16	0.27	0.030
60	0.67	0.14	0.35	0.034	0.67	0.15	0.43	0.034	0.61	0.16	0.10	0.031
90	0.52	0.18	0.94	0.031	0.52	0.19	1.13	0.031	0.52	0.20	1.08	0.029
120	0.44	0.20	1.45	0.036	0.44	0.22	1.75	0.037	0.44	0.23	1.71	0.035
总的绝对均方差	$\sigma_{总}$ = 0.045 0				$\sigma_{总}$ = 0.041 6				$\sigma_{总}$ = 0.038 7			

说明: $\sigma_{绝}=\sqrt{\frac{1}{n}\sum_{l=1}^{n}(x_p-x_l)^2}$, $\sigma_{总}=\sqrt{\frac{1}{n}\frac{1}{l}\sum_{l=1}^{n}\sum_{y=1}^{k}(x_p-x_{ly})^2}$, 式中 k为历时总数。

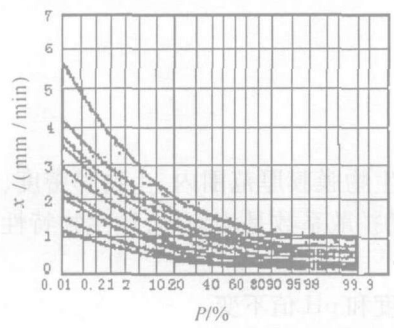

图 2　矩法

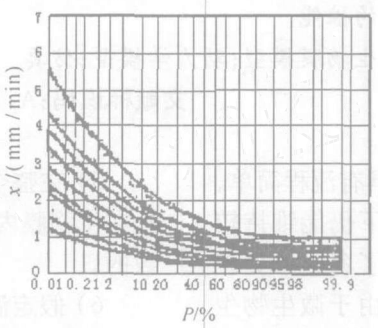

图 3　试错适线法

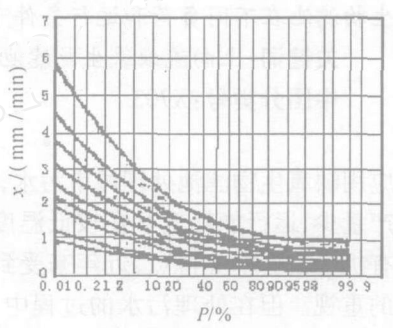

图 4　混合遗传优化适线法

从表 1可知:在拟合准则和比较指标相同的条件下,无论是对比各个历时的绝对均方差还是比较所有历时的总绝对均方差,混合遗传优化适线法优于矩法和试错适线法。由图 1 - 3可知,混合遗传优化适线法与经验点据的拟合情况最佳。

3　结　语

通过上述方法,结合具体的实践验证,得出如下结论:

1) 结合三元浮点编码法,并在进化中引入高斯变异下降方向因子,从而构造出混合遗传算法。该法不但保留了进化算法的通用性,而且求解的精度和算法的收敛速度都有很大的改进,同时算法简单,易于编程实现。

2) 选用多元插值原理对 $\Phi-(p, c_s)$表进行了细化和扩展,在此基础上结合混合遗传算法进行优化适线,可实现同时对 3个参数 $\bar{x}$、c_v、c_s进行精确估计,克服了适线法未能解决 $\bar{x}$、特别是 c_v的估计精度,以及计算过程与结果受人为因数的影响等不足。

3) 通过具体工程计算体现了本文所述方法的准确性,为 P - Ⅲ型分布参数估计提供了一种新的方法,可广泛推广使用。

参考文献:

[1] GBJ14 - 1987. 室外排水设计规范 [S].
[2] 王俊德. 水文统计 [M]. 水利电力出版社, 1993.
[3] 刘光文. 皮尔逊 Ⅲ型分布参数估计 [J]. 水文, 1990, (4、5): 1 - 15.
[4] HOSKING J R M L. Moments Analysis and Estimation of Distribution Using Linear Combination of Order Statistics [J]. Journal of the Royal Statistical Society, Senes B, 1990, 52: 105 - 124.
[5] 邱林,陈守煜,潘东,等. P - Ⅲ型分布参数估计的模糊加权优化适线法 [J]. 水利学报, 1998 (1): 33 - 38.
[6] 周明,孙树栋. 遗传算法原理及应用 [M]. 北京:国防工业出版社, 2001.
[7] 袁晓辉,张双全,等. 非线性马斯京根模型参数率定的新方法 [J]. 水利学报, 2001 (5): 77 - 81.
[8] LIN W, DELGADO - FRIAS JG. Hybird Newton - Raphson Genetic Algorithm for Traveling Salesman Problem [J]. Cybernetics and Systems, 1995, 26 (5): 387 - 412.
[9] MICHALEWICZ Z. A Modified Genetic Algorithm for Optimal Control Problems [J]. Computers Math Application, 1992, 23 (12): 83 - 94.
[10] 唐焕文, 秦学志. 最优化方法 [M]. 大连:大连理工大学出版社, 1994.
(下转第 94页)

© 1994-2008 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

Modal Analysis and Vibration Control of Benfield Frame

WANG Zhi-song[1], LI Zheng-liang[1], ZHANG Yu[2], TANG Yi-ke[3], XIE Zhi-jiang[3]
(1. Civil Engineering College, Chongqing University, Chongqing 400030, China;
2. Civil Engineering College, Southwest University of Scince and Technology, Mianyang 621002, China;
3. Mechanical Engineering College, Chongqing University, Chongqing 400030, China)

Abstract: There is a low frequency and great amplitude on Benfield frame of Jianfeng chemical plant. With finite element modal, the reasons and mechanism of the vibration on Benfield frame are analyzed and confirmed. And a method of vibration control about the main vibration resource (100V10) is proposed to reduce the amplitude of Benfield frame vibration.

Key words: low frequency and great amplitude vibration; modal analysis; resonance; vibration control

(编辑 姚 飞)

(上接第 85页)

Hybrid Genetic Algorithm Optimum Curve - fitting Method for Estimating the Parameters of Pearson type - Ⅲ distribution

REN Bo-zhi[1,2], LONG Teng-rui[2]
(1. College of Urban Construction and Environmental Engineering, Chongqing University, Chongqing 400030, China;
2. College of Sea and River Engineering of Changsha Science and Technology, Changsha 410076, China)

Abstract: Based on subdivided and extended the table of Φ - (p, c_s) by using the binary interpolation method, and analyzing traditional genetic algorithm shortcoming, the hybrid genetic algorithm optimum curve - fitting method is proposed. By the conclusion of application in Changsha, it is proved that this method can estimating the parameters of Pearson type-Ⅲ distribution at the same time. This method overcomes the evaluating accuracy problem of $\bar{x}$、C_v the tradition curve-fitting method, and the calculating process and results can not be effected by the personal objective attitude, and this method supplies one kind of fully valid means to estimate the parameters of Pearson type-Ⅲ distribution.

Key words: pearson type-Ⅲ distribution; the parameters estimation; the curve-fitting method; hybrid genetic algorithm; optimization

(编辑 姚 飞)

© 1994-2008 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net