

文章编号: 1000-0585(2001)01-0097-06

水质模型参数的非数值随机优化^{*}

郑红星, 李丽娟

(中国科学院地理科学与资源研究所, 北京 100101)

摘要: 以模拟退火算法为核心着重讨论了水质模型参数的非数值随机优化方法。实例分析表明, 利用非数值随机优化方法(包括模拟退火算法和遗传算法)对水质模型参数进行估计, 可以获得较为理想的结果。不同参数估计方法的比较进一步阐述了非数值随机优化方法在参数估计中的优点。

关键词: 水质模型参数; 非数值算法; 随机优化; 模拟退火算法

中图分类号: P342; O224 文献标识码: A

水质模型参数估计是水质系统辨识的一个重要环节, 所研究的是当模型结构已经确定, 如何确定模型参数的问题。在水质模型参数物理意义的范围内, 使模拟结果与实际观测值之间误差最小的参数估计问题本质上则属于参数优化的研究范畴。传统的水质模型参数的确定方法有实验室法、经验公式法和试错法等。实验室法和经验公式法获取的参数值往往很难反映水体的实际情况, 因而在实际使用中有一定的局限性。试错法虽然可以获得相对满意的结果, 但是需要较多的计算机时, 而且参数估计结果并非最优。水质模型参数的同时估计法是近年来逐渐发展成熟起来的一类方法, 其基本思想是应用多维参数的最优化估值方法, 同时确定模型中的各个参数。这种方法从模型的整体出发, 充分利用计算机计算速度快的特点, 因而具有搜索时间短, 参数可靠性较高的优点, 已逐渐成为人们研究的热点, 并在实际应用中发挥了较大的作用。在水质模型参数的同时估计中, 应用较多的有最速下降法、复合形法(complex)和计算机扫描计算-图解-梯度搜索法。但是这些方法也同样存在着诸如受初始条件影响明显、算法过于复杂等不足。当前, 遗传算法、模拟退火算法等非数值随机优化算法的研究应用, 为水质模型参数的最优估计提供了新的途径。

1 非数值随机优化

一般而言, 数值算法是指基于代数关系运算的一类诸如矩阵运算、多项式运算、解三角形线性系等的计算方法, 基本上是属于数值分析(以数字形式表达的问题求数值解)的范畴。对于参数优化问题, 数值算法的一般方法是利用普通极值的充分必要条件来寻优, 即

^{*} 本文承蒙吴国元教授指正, 谨致谢忱!

收稿日期: 2000-03-17; 修订日期: 2000-06-15

基金项目: 国家重点基础研究发展规划项目(G1999043602)和中国科学院地理科学与资源研究所创新项目资助(CX10G-A00-07)

作者简介: 郑红星(1973-), 男, 福建永春人, 博士研究生, 研究方向为水资源和水环境。 <http://www.cnki.net/journals/1000-0585/200101097.htm>

指标函数 $Z(S^T)$ 为最小值的充分必要条件为: ① 一阶偏导数为零(必要条件); ② 二阶偏导数阵(记为 $H(S)$) 为正定阵(充分条件)。如果 $Z(S^T)$ 可以写成解析式, 而且能求出其一次及二次偏导数, 那么可以从一阶偏导数为零的条件中给出的 n 个非线性方程求得 $S^T = [s^0, s^1, \dots, s^n]$, 然后再代入 $H(S)$ 中校验 $H(S)$ 是否正定, 如果正定, 则这组 S^T 值即为所求使 $Z(S)$ 为极小值的 S^{T*} 。然而, 在控制系统参数最优化问题中, 由于指标函数 $Z(S^T)$ 一般很难写成解析式, 或解析式形式较为复杂等原因, 上述数值方法应用起来有较大的困难。

与数值算法不同, 非数值算法是基于关系运算的一类诸如搜索、选择、排序、匹配以及图论等方面的计算方法, 基本上是属于符号(如字符、数字、图形或其它记号)处理的范畴。其特点是算法直接、简单。具体地讲, 非数值算法的参数优化过程是先给出任意的初始点 $S^T(0) = S^T_0$, 然后解系统动态方程得到 $Z(S^T_0)$, 判断 $Z(S^T_0)$ 是否为最小, 若最小则停止计算, 否则按某种规律改变 S^T , 然后再解动态方程, 如此循环反复, 直到 $Z(S^T)$ 为最小。根据 S^T 改变规律差异, 非数值优化方法可分为定向寻优和随机寻优两类。前者如最速下降法、单纯形法^[3]等, 其特点是寻优过程中每一步的方向都十分明确; 后者如遗传算法^[2, 8]和模拟退火算法等, 其最大的特点表现在搜索方向的不确定性上。本文以模拟退火算法为重点, 说明水质模型参数的非数值随机优化。

2 模拟退火算法

模拟退火(Simulated Annealing, 简称 SA) 算法, 是基于蒙特卡罗(Monte Carlo) 迭代求解法的一种启发式随机搜索算法。模拟退火算法用于求解优化问题的出发点, 是基于物理中固体物质的退火过程与一般组合优化问题间的相似性。在对固体物质进行退火处理时, 通常先将其加温熔化, 使其中的粒子可以自由运动, 然后随着温度的逐渐下降, 粒子也逐渐形成了低能态的晶格。若在凝结点附近的温度下降速率足够慢, 则固体物质一定会形成最低能量的基态, 这一过程和金属的初始状态无关。根据这一过程, N. Metropolis 等人在 1953 年提出了 SA 算法的基本思想, 后来由 S. Kirkpatrick 等人先后研究发展了 SA 算法的理论, 并应用到寻找函数最优解问题上(S. Kirkpatrick et al., 1983; Cenry, 1985, Geman, 1984)。理论上已经证明, 通过模拟退火过程, 可找到全局(或近似)最优解。

2.1 基本原理

模拟退火算法的基本思想可以表述为: 把某种状态的 S 看成某一物质系统的微观状态, 而 $Z(S)$ 为该状态下物质系统的内能。利用控制参数 T 类比温度, 让 T 从一个足够高的温度慢慢下降, 对每个 T , 用 Metropolis 抽样法在计算机上模拟该体系在此 T 下的热平衡态, 即对当前状态 S 作随机扰动产生一个新状态 S' , 计算增量 $\Delta Z = Z(S') - Z(S)$, 并以概率 $\exp(-\Delta C/kT)$ 接受 S' 作为新的当前状态。当如此重复地随机扰动足够次数后, 状态 S 表现为当前状态的概率将服从玻耳兹曼分布, 即

$$f = z(T) \exp[C(S)/kT]$$

其中

$$z(T) = \frac{1}{\int \exp[-C(S)/kT]}$$

式中 k 为玻耳兹曼常数。若 T 下降足够慢, 且 $T \rightarrow 0$, 则将求得具有最小能量 $Z(S)$ 的状态 S^* 。

2. 2 算法的实现

模拟退火算法的实现关键在于以下相关函数的选取设置:

(1) 以一定的概率密度从当前状态 S 跃迁到新的状态 S' , 这个概率密度函数称为生成函数 (generating function)。本文取

$$S' = \text{Generate}(S) + \beta \sin \alpha$$

式中 β 为给定的步长, α 为 -1 和 1 之间根据均匀分布产生的随机数, S 相当于水质模型中的待估参数 $[k_1, k_2, \dots, k_n]$ 。

(2) 以一定的概率密度接受目标函数值的偶然上升 (即当 $\Delta > 0$ 时), 这一概率密度函数称为接受函数 (acceptance function)。取概率密度函数为

$$f = \exp(-\Delta/T)$$

这里, $\Delta = Z(S') - Z(S)$, T 表示温度, 又设 γ 为用均匀分布概率产生的 0 和 1 之间的随机数, 则这一过程 (记作 $\text{Accept}(\gamma, S')$) 可以表述为:

如果 $\gamma < f$, 则 $S = S'$

(3) 以一定的冷却方式降低等效温度 T , 等效温度是生成函数与接受函数中的控制参量, 确定了所引入的随机扰动 (噪声) 的强度。设温降函数 (cooling function) 形如:

$$T_{i+1} = \eta T_i \quad (0 < \eta < 1)$$

应用上述选取的函数形式, 模拟退火算法水质模型参数估计的计算流程见图 1。图中初始化过程包括: 初始温度 T_0 、初始状态 S 、温降系数 η 、搜索步长 β 、内循环次数 n 以及外循环结束条件 (收敛条件) 的确定。

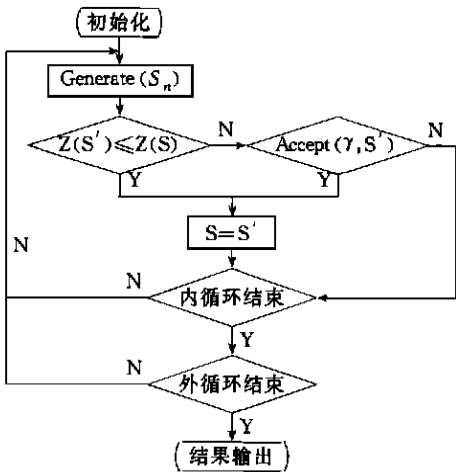


图1 模拟退火算法水质模型参数估计计算图
Fig. 1 Flow chart of parameter estimation with simulated annealing algorithm

3 实例分析

作为一个例子^①, 设某均匀河段起始断面 BOD 浓度为 20 mg/l , 溶解氧浓度为 10 mg/l , 饱和溶解氧的浓度为 10 mg/l , 河流流速 $u = 4 \text{ km/d}$ 。该河段 $8、28、36、56 \text{ km}$ 处的 BOD 与 DO 的实测浓度值如下表所示, 要求估计 BOD 降解系数 k_1 和复氧系数 k_2 。

3. 1 模拟退火算法水质模型参数估计

选取模型结构较为简单又具有较强代表性的 $S-P$ 模型, 并取实测值与模拟值的误差平方和最小为目标函数, 则水质模型参数估计问题可用数学式子表示如下:

$$\min Z = \sum_{i=1}^n \sum_{j=1}^m [\lambda (B_{ij} - B_i^*)^2 + (1 - \lambda) (C_{ij} - C_i^*)^2] \tag{1}$$

式中 B_{ij} 、 C_{ij} 表示 i 断面第 j 次的实测值, $i = 1, \dots, n, j = 1, \dots, m$; B_i^* 、 C_i^* 为最优参数下的水质模拟值; λ 为加权因子 ($0 < \lambda < 1$), 表示 BOD 和 DO 实测值的相对可靠程度。且:

① 根据参考文献 [6] 算例二。

$$C = C_s - (C_s - C_0)e^{-k_1x/u} + k_1C_0(e^{-k_1x/u} - e^{-k_2x/u})/(k_1 - k_2)$$
$$B = B_0e^{-k_1x/u}$$

(2)

(3)

式中 B 、 B_0 、 C 、 C_0 分别为 $x=x$ 和 $x=0$ 处河水的 BOD 、 DO 浓度 (mg/l); x 为离排污口的距离 (m); u 为河水平均流速 (m/s); k_1 为 BOD 衰减系数 ($1/\text{d}$); k_2 为河水复氧系数 ($1/\text{d}$); C_s 为河水在某温度下的饱和溶解氧浓度 (mg/l)。

表 1 某河流各断面不同时刻的水质实测值

Tab. 1 Actual concentration of BOD and DO of a river concerned

断面号	断面距离/km	BOD 实测浓度值/ $\text{mg} \cdot \text{l}^{-1}$			DO 实测浓度值/ $\text{mg} \cdot \text{l}^{-1}$		
1#	$x=8$	17.928	16.891	18.304	8.335	8.657	8.032
2#	$x=28$	13.783	13.156	13.478	6.750	6.565	6.830
3#	$x=36$	12.392	11.934	12.283	6.654	6.631	6.823
4#	$x=56$	6.932	6.879	6.951	9.501	9.343	9.677

取 $\lambda=0.5$, 利用模拟退火算法, 给定不同的初始值及搜索条件, 获得参数的估计结果如下表 2 所示。由表 2 可以看出, 不论初始值 k_1 、 k_2 以及搜索条件如何设置, 得到的参数估计值都较为接近, 且指标函数值也稳定在一定的精度范围内。水质模型参数的最优估计为 $k_1=0.055$, $k_2=0.203$, 此时指标函数值 $Z=1.04$ 。这说明, 模拟退火算法在水质模型参数估计中是有效的。另一方面, 就参数优化速度看, 搜索时间表现出较大的随机性。优化速度一般因收敛条件的宽严而有相应的快慢变化。除此之外, 搜索时间与初始值 k_1 、 k_2 及其它搜索条件的关系不明显。

表 2 模拟退火算法水质模型参数估计结果

Tab. 2 Parameter estimation results of simulated-annealing algorithm

初始值		搜索条件				参数估计值		指标函数值	搜索时间
k_1	k_2	T_0	η	β	ϵ	k_1	k_2	Z	t/s
0.4	0.6	2 000	0.85	0.01	1.5	0.052	0.191	1.469	18.13
0.8	0.2	2 000	0.85	0.01	1.5	0.052	0.192	1.385	0.88
0.4	0.6	200	0.85	0.01	1.5	0.055	0.210	1.091	0.38
0.4	0.6	20	0.85	0.01	1.5	0.055	0.214	1.152	3.24
0.4	0.6	200	0.5	0.01	1.5	0.056	0.214	1.152	13.45
0.4	0.6	200	0.85	0.001	1.5	0.056	0.219	1.275	9.68
0.4	0.6	200	0.85	0.01	1.05	0.055	0.203	1.040	8.35

注: 表中 T_0 为初始温度, η 为温降系数, β 为搜索步长, ϵ 为收敛条件, 取内循环次数 $n=10$ 。

3.2 方法的比较

为进一步分析模拟退火算法在水质模型参数估计中的作用, 阐明非数值随机优化方法的特性, 本文利用另一非数值随机优化算法——遗传算法 (GA) 和具有代表性的非数值定向搜索方法——最速下降法 (GD) 分别对上述算例进行参数优化, 结果如下表 3、表 4。

由表 2、表 3 和表 4 可以看出, 三种方法都可以求得较为准确的参数估计值。就参数优化的速度而言, 最速下降明显要快于两种非数值随机寻优方法 (SA 和 GA), 但是其解的稳定性则明显要逊于 SA 和 GA。表 4 表明, 最速下降法水质模型参数估计结果受初始搜索条件的影响较大, 不同的搜索条件下获得的水质模型参数有较大的差别。这主要是因为该方

法是一类有向的搜索策略, 受“出发点”位置、搜索步长等条件的影响, 寻优过程容易陷入局部最优。就两种非数值随机寻优方法而言, 它们在理论上都被证明能找到或接近全局最优, 从实际的例子也可以看出, 其参数优化效果良好。但是, 在一定程度上, 这是以牺牲求解速度为代价的。

表 3 遗传算法水质模型参数估计结果
Tab. 3 Parameter estimation results of genetic-algorithm

搜索条件				参数估计值		指标函数值 Z	搜索时间 / s
p_{op}	p_c	p_m	ϵ	k_1	k_2		
40	0.85	0.01	1.5	0.056	0.200	1.163	29.11
10	0.85	0.01	1.5	0.054	0.185	1.168	5.94
20	0.85	0.01	1.5	0.056	0.190	1.299	7.19
20	0.95	0.01	1.5	0.054	0.214	1.213	13.18
20	0.85	0.001	1.5	0.052	0.190	1.469	4.67
20	0.85	0.01	1.2	0.056	0.220	1.168	75.9

注: 表中 pop 表示种群数, pc 为交叉概率, pm 为变异概率, ϵ 为收敛要求。

表 4 最速下降法水质模型参数估计结果
Tab. 4 Parameter estimation results of steepest-decent algorithm

初始值		搜索条件				参数估计值		指标函数值 Z	搜索时间 / s
k_1	k_2	P	q	q_1	q_2	k_1	k_2		
0.4	0.6	0.5	0.01	0.25	0.001	0.056	0.600	20.224	0
0.4	0.6	0.5	0.01	0.005	0.001	0.052	0.598	21.287	0.05
0.4	0.6	0.01	0.001	0.25	0.000 1	0.055	0.237	1.720	0.44
0.4	0.6	0.1	0.000 1	0.05	0.001	0.055	0.209	1.078	0.93
0.8	0.2	0.1	0.000 1	0.05	0.000 4	0.055	0.206	1.057	0.51

注: 表中 p 为搜索步长, q 为参数增量, q_1 为步长修正系数, q_2 为指标函数的收敛要求。

就遗传算法和模拟退火算法的差别而言, 模拟退火算法是一种基于热力学退火原理建立起来的随机搜索算法, 该算法已在理论上被证明是以概略 1 收敛至全局最优的一类方法, 其解的质量也较高, 但在控制其主要参数如初始温度、温降系数等问题上却存在一定难度。温度的管理常需要实验地加以调整。与模拟退火算法不同, 遗传算法从生物进化的规律得到启发, 其最大的特点在于它是一种群体的搜索策略, 即同时从多个状态出发, 通过交叉、变异等手段, 不断向最优解逼近。尽管理论上已经证明 GA 算法最终能收敛至全局最优, 但收敛过程需要较多的时间, 这一时间复杂度问题仍亟待解决。如前所述, 在两种非数值随机优化方法 (SA 和 GA) 间, 虽然参数优化的效果都较为理想, 但模拟退火算法在解的精度及优化速度上都优于遗传算法。两种随机优化方法 (SA 和 GA) 虽然在参数优化的具体过程上有所不同, 但是都是基于随机搜索思想的寻优方法, 其最大的特点表现在搜索方向的不确定性。正是由于这一特性, 才使寻优过程有可能跳出局部最优而搜索到全局最优解。

4 结语

的应用。实例计算及分析表明:

(1) 非数值参数优化避开了数值方法求导和求解方程组的过程, 算法简洁有效。

(2) 非数值随机优化 (SA、GA) 在求解精度上明显高于定向寻优方法 (如最速下降法 GD), 但在一定程度上却是以牺牲求解速度为代价的。

(3) 就两种非数值随机优化方法 (SA 和 GA) 而言, 两者在理论上都已被证明能以概率 1 收敛至全局最优, 实例研究也表明这两种算法都能获得较高精度的解, 而且解的稳定性较好。但 SA 的求解精度和速度都优于 GA。

(4) 当前, 非数值随机优化方法虽然在求解速度上存在着一定的不足, 但是, 有理由相信随着计算机技术的发展 (特别是并行处理技术的发展与推广) 以及算法理论的进一步完善, 模拟退火算法和遗传算法将因其自身固有的优点而受到越来越广泛的关注。

参考文献:

- [1] 康立山, 谢云 等. 非数值并行算法——模拟退火算法[M]. 北京: 科学出版社, 1998. 22~55.
- [2] 陈国良, 王熙法 等. 遗传算法及其应用[M]. 北京: 人民邮电出版社, 1996. 101~127.
- [3] 傅国伟. 河流水质数学模型及其模拟计算[M]. 北京: 中国环境科学出版社, 1987. 181~265.
- [4] 赵振宇, 徐用懋. 模糊理论和神经网络的基础与应用[M]. 北京: 清华大学出版社, 广西: 广西科学技术出版社, 1996. 103~115.
- [5] 翟国静. 马斯京根模型参数估计方法探讨[J]. 水文, 1997(3): 40~43.
- [6] 王黎, 马光文. Streeter-Phelps 模型参数估计的遗传算法[J]. 水科学进展, 1997(3): 32~37.
- [7] 谭炳卿. 水文模型参数自动优选方法的比较分析[J]. 水文, 1996(5): 8~13.

Stochastic optimization on parameters of water quality model

ZHENG Hong-xing, LI Li-juan

(Institute of Geographic Sciences and Natural Resources Research, CAS, Beijing 100101, China)

Abstract: This paper focuses on stochastic parameter optimization for water quality model with simulated-annealing algorithm (SA) which is discussed in detail. For comparison, genetic-algorithm (GA) and steepest-decent algorithm (SD) are also discussed. Simultaneously, the typical S-P water quality model is adopted in a case study. Result of the case study shows that the stochastic optimization methods (SA and GA) are more effective than the other methods such as the steepest-decent method. What are testified include not only in the aspect of theory but also in the case study, both SA and GA are able to reach the global optimal results. However, concerning SA and GA, GA is weaker in local optimization and spends more time in parameter optimization.

Key words: stochastic optimization; simulated annealing algorithm; parameters; water quality model