

基于多初值 GRG 算法与遗传算法的 ReNuMa 模型校准模块优化

李川, 沙健, 赵罡, 王玉秋

(南开大学 环境科学与工程学院, 天津 300071)

摘要: ReNuMa 模型设计用于在大流域尺度上评估营养盐通量。其校准模块采用的是广义简约梯度算法 (GRG)。在实际应用中, 该校准模块有许多不足之处。为提高 ReNuMa 模型校准模块的校准效率和全局寻优能力, 提出了具有多初始点的 GRG 算法、遗传算法和遗传算法与 GRG 算法连用的方法。以练江流域月径流量模拟为案例开展比较研究。结果表明: 对 ReNuMa 模型校准模块的优化有效的提高了校准结果的有效性和参数的全局寻优能力。

关键词: 区域营养盐管理; 参数校准; 广义简约梯度算法; 遗传算法

中图分类号: X522; O224

文献标识码: A

文章编号: 1672-643X(2014)01-0095-05

Optimization of calibration module of ReNuMa model based on multi-start GRG and GA

LI Chuan, SHA Jian, ZHAO Gang, WANG Yuqiu

(College of Environmental Science and Engineering, Nankai University, Tianjin 300071, China)

Abstract: ReNuMa model is designed to estimate nutrient fluxes at the scale of large watershed. The calibration module of ReNuMa uses Generalized Reduced Gradient (GRG) optimal algorithm. In practical application, the optimal module of ReNuMa has many shortcomings. For the purpose of improving the calibration efficiency of ReNuMa and the capacity of finding global optimal result, the paper put forward multi-start GRG, Genetic Algorithm (GA) and combination of both to improve the optimal results. The improved calibration module of ReNuMa shows a better performance on the validity of calibration result and the capacity of finding the global optimal result on the basis of the case in Lianjiang river basin.

Key words: regional nutrient management (ReNuMa) model; parameter calibration; generalized reduced gradient method (GRG); genetic algorithm (GA)

在现代流域管理中, 模型以及其他分析方法是重要的工具^[1]。模型对环境状态的模拟均通过复杂的数值计算^[2]。当模型被应用在一个具体的地点去描述相关过程时, 该模型应当根据当地多年观测数据进行校准, 以获取更好的模拟效果^[2-3]。

GWLF 模型是 1987 年美国 Haith 和 Shoemaker 共同设计开发的。该模型能够对流域水系中的溶解态氮磷以及总氮总磷月负荷量进行评估, 同时还可提供土壤侵蚀、河川径流量以及沉积物量。GWLF 模型已在美国多地得以应用^[4-6], 是被美国 EPA 认证的中尺度流域模型, 同时该模型也在天津市于桥

水库流域得到了应用^[7-8]。区域营养盐污染负荷模型 ReNuMa 是美国康奈尔大学 Bongghi Hong 和 Dennis P. Swaney 基于 GWLF 模型开发设计的, 该模型是一个靠水文驱动的半经验模型, 设计用于在大流域尺度上评估营养盐通量。ReNuMa 模型已在我国长春石头口门水库流域^[9]以及天津市饮用水水源于桥水库上游沙河子流域^[1,10]得以应用, 为流域管理决策的制定提供了科学支持。

ReNuMa 是基于 Microsoft Excel 实现的, 他所采用的编程语言是 Visual Basic for Application (VBA), Excel 作为其图形用户界面, 用于管理模型的输入以

收稿日期: 2013-08-19; 修回日期: 2013-09-18

基金项目: 环境保护部环境规划院水污染综合防治项目 (2013A009)

作者简介: 李川 (1990-), 男, 河南安阳人, 硕士研究生, 主要从事流域地表水环境模型研究。

通讯作者: 王玉秋 (1965-), 男, 黑龙江齐齐哈尔人, 教授, 博士生导师, 主要从事水环境管理决策与技术支持, 流域及水源保护战略等研究。

及结果输出。ReNuMa 参数校准模块采用了广义简约梯度算法。在实际应用中,该校准模块存在诸多不足之处。本文总结了 ReNuMa 模型校准模块的缺陷,提出了若干改进措施,并以新安江支流练江流域为例,基于 2000 - 2010 年 11 年月径流量数据为观测值进行校准测试,对比验证了改进措施的效果。

1 ReNuMa 校准模块的缺陷

ReNuMa 模型的校准模块基于 GRG 算法^[11-12]以拟合一个或多个参数,使得月总量模拟结果与实测结果尽可能接近,即残差平方和最小。目前,ReNuMa 的校准模块存在多种缺陷,主要包括:

1.1 校准过程无法保证参数的物理意义

ReNuMa 的校准是在满足参数物理意义的基础上进行的,不满足参数物理意义的校准结果是不被接受的。ReNuMa 原始校准模块仅仅能够保证参数值大于等于 0.000001,而对每个参数的上限却没有规定。在多参数校准中,这种缺陷一方面会增大求解规模;另一方面会使部分参数的校准结果失去物理意义。

1.2 GRG 算法全局寻优能力较差

广义简约梯度法是将简约梯度法推广到求解具有非线性约束优化问题的一种新方法。这种方法是目前求解一般非线性优化问题的比较有效的算法。

GRG 算法求解的非线性规划问题可归纳为如下形式:

$$\begin{aligned} \min: & f(x) \\ \text{s. t.} & C(x) = 0 \\ & \alpha \leq x \leq \beta \end{aligned} \quad (1)$$

式中: $x \in R^n$, α, β 为 R^n 上常向量, $C(x) = (c_1(x), c_2(x), c_3(x), \dots, c_m(x))^T$ ($m \leq n$), $f(x), c_i(x)$ 都是连续可微函数。

假定对任一可行解 x 可分解成基向量 $x^B \in R^m$ 和非基向量 $x^N \in R^{n-m}$, 相应地 $\alpha = \begin{pmatrix} \alpha^B \\ \alpha^N \end{pmatrix}$, $\beta = \begin{pmatrix} \beta^B \\ \beta^N \end{pmatrix}$, 且

$$\alpha^B < x^B < \beta^B \quad (2)$$

又假定向量组

$$\begin{aligned} \nabla_B c_i(x) &= \left(\frac{\partial c_i(x)}{\partial x_{B_1}}, \dots, \frac{\partial c_i(x)}{\partial x_{B_m}} \right)^T, \\ i &= 1, 2, 3, \dots, m \end{aligned} \quad (3)$$

是线性无关的,则 $m \times m$ 矩阵是非奇异的。

$$\nabla_B c(x) = (\nabla_B c_1(x), \dots, \nabla_B c_m(x)) \quad (4)$$

令:

$$\nabla_N c(x) = (\nabla_N c_1(x), \dots, \nabla_N c_m(x)) \quad (5)$$

$$\nabla_N c_i(x) = \left(\frac{\partial c_i(x)}{\partial x_{N_1}}, \dots, \frac{\partial c_i(x)}{\partial x_{N_{n-m}}} \right)^T$$

$$i = 1, 2, 3, \dots, m$$

由多元微积分学中隐函数存在定理知,在 x 的某领域内可由非线性方程组

$$C(x^B, x^N) = 0 \quad (6)$$

确定 x^B 为 x^N 的函数:

$$x^B = x^B(x^N) \quad (7)$$

可计算出

$$r(x^N) = \nabla_N f(x) - \nabla_N c(x) (\nabla_B c(x))^{-1} \nabla_B f(x) \quad (8)$$

称 $r(x^N)$ 为 $f(x)$ 的简约梯度。

令下降可行方向 P^N 的分量为

$$P_j^N = \begin{cases} 0, & \text{当 } x_j^N = \alpha_j^N, \text{ 且 } r_j(x^N) > 0 \text{ 或} \\ & x_j^N = \beta_j^N \text{ 且 } r_j(x^N) < 0 \text{ 时} \\ -r_j(x^N), & \text{其它情形} \end{cases} \quad (9)$$

令 $x = x_k$ 为第 k 次可行迭代点,设搜索步长为 θ ,则可通过求解一维极小问题确定搜索步长 θ :

$$\begin{aligned} \min & f(x_k + \theta P^N) \\ \text{s. t.} & C(x_k + \theta P^N) = 0 \\ & \alpha \leq x_k + \theta P^N \leq \beta \end{aligned} \quad (10)$$

获取搜索步长,即得到后继可行迭代点 x_{k+1} 。

ReNuMa 中存在很多不连续的函数关系和多极值函数关系。根据 GRG 算法可知,GRG 算法需要根据函数的偏导式估值和梯度信息去计算参数的变化方向和变化步长,而函数的偏导式估值和梯度信息是基于函数的连续性构建的。针对不连续问题,GRG 算法无法在不连续点处计算函数的偏导式估值和梯度信息,所以 GRG 算法无法保证最后的寻优结果是全局最优值。当针对多极值问题时,其最终的校准结果是一个临近初始给定值的局部最优值。

2 ReNuMa 校准模块的改进措施

2.1 设置参数取值范围

在 ReNuMa 模型中,每个参数都有具体物理意义以及科学含义的。原始的 ReNuMa 校准模块在算法上允许参数出现不符合其物理意义的校准结果。通过修改 ReNuMa 的源代码,为每个被校准的参数添加约束条件。以 ReNuMa 径流模拟为例,典型的退水系数为 0.01 ~ 0.2,渗滤系数的推荐初始值为 0,由其物理意义可知其参数范围为 0 ~ 1,每种土地利用类型的 CN 值依据《U. S. Soil Conservation Serv-

ice's Technical Release No. 55》设定,每种土地类型的蒸发覆盖因子依据《GWLf Users Manual》相关说明均设定为 0.3~1。

经过修改后的 ReNuMa 校准模块可以灵活的设定参数取值范围,上述参数并非被设定不变,而是可以通过 ReNuMa 文件“calibrate”工作表中新增的“Auto-Calibration”两列单元格人工输入。使用者可以依据相关文献、模型使用经验和具体应用地点的特殊场景灵活设置。

2.2 添加具有多初始点的 GRG 算法

针对 GRG 算法全局寻优能力差的缺点,一个较好的解决办法就是在进行校准的时候,从多个初始点运行 GRG 算法,进而可以得到多个局部最优值,通过比较这些局部最优值,可以得到一个更好的校准结果。虽然具有多初始点的 GRG 算法能在很大程度上提高最终校准结果,但是该算法并不保证最后寻优结果为全局最优值。不过针对 ReNuMa 参数校准而言,该算法已经极大地提高了找到全局最优参数集的概率。

2.3 添加遗传算法

遗传算法是模仿自然界生物进化机制发展起来的随机全局搜索和优化方法,它借鉴了达尔文的进化论和孟德尔的遗传学说。遗传算法本质上是一种高效、并行、全局搜索的方法,它能在搜索过程中自动获取和积累有关搜索空间的知识,并自适应地控制搜索过程以求得最优解^[13]。

基本遗传算法可表示为:

$$SGA = (C, E, P_0, M, \phi, \Gamma, \Psi, T) \quad (11)$$

式中: C 为个体的编码方法; E 为个体适应度评价函数; P_0 为初始种群; M 为种群大小; ϕ 为选择算子; Γ 为交叉算子; Ψ 为变异算子; T 为遗传运算终止条件。

遗传算法并不关注问题本身的数学特性^[14],其本质上是一个具有定向制导的随机搜索技术^[15],而它的这一特点则非常有利于解决 GRG 算法所难以解决的不连续问题和多局部最优值问题。

和基本遗传算法相比,ReNuMa 中新添加的遗传算法并不需要明确的指明遗传算法的终止进化代数,也不需要明确指明交叉概率。ReNuMa 中遗传算法通过“无改进的最大时间”来控制繁衍的代数。这里指定“无改进的最大时间”为 300 s,即当后代个体在繁衍和进化的 300 s 内,如果没有出现更好的个体,则校准结束。

2.4 遗传算法与 GRG 算法连用

遗传算法对求解问题本身的数学特性一无所知,正因为如此,它并不能确切地检测最后寻优结果是否为全局最优值或局部最优值。遗传算法有较强的全局搜索能力,但是局部寻优能力不足^[16]。它能比较当前寻优结果是否比之前的寻优结果有所提高。在大多数情况下,遗传算法只是找到它所遍历的参数集中的最优结果。但是由于遗传算法自身的算法机制,只要能够提供合适的突变概率和运算时间,就能够确保它很好的遍历整个参数空间,最后的寻优结果也往往能够落在一个较优的局部参数区域内。

GRG 算法具备较好的局部寻优能力,当用遗传算法得出一个较好的局部参数集时,在此基础上利用 GRG 算法进行寻优,就能够得出该局部参数区域中的最优参数集。从理论上讲,上述两种算法的连用可以极大的提高寻优效果。

通过对 ReNuMa 校准模块的修改,在经过遗传算法寻优之后,继续使用 GRG 算法寻优。其中,设置连用方法中遗传算法的“无改进的最大时间”仍为 300 s。

3 算法测试

为了验证改进算法的特性,以练江流域月径流量模拟为案例,对其进行河川径流参数集的校准。以校准耗时,校准后参数合理性以及校准后模拟值和实测值之间的残差平方和为评价指标,分别利用 ReNuMa 原始校准方法,多初始值的 GRG 算法,遗传算法和遗传算法与 GRG 算法连用的方法进行多参数校准。在多初始值的 GRG 算法,遗传算法和遗传算法与 GRG 算法连用的方法中,按照参数物理意义为每个参数设定取值范围。同时为了更好地证明 GRG 算法与改进后的算法在算法特性上的差别,增设一组添加范围的 GRG 算法。由于遗传算法的寻优结果与其繁衍代数紧密相关,为了验证该算法寻优能力,增设一组“无改进的最大时间”为 1 000 s 的遗传算法测试组。

由于 GRG 算法受初始给定值影响较大,所以按照之前设定的参数取值范围,取由各个参数上限、下限和上下限平均值组成的 3 组参数集,作为初始参数集。同时,由于遗传算法内部的选择算子、交叉算子和变异算子受随机数影响,所以每组校准各运行 10 次。校准所需时间以及残差平方和用 10 次校准所得平均值加减标准差表征。测试所用计算机采用 Intel i5 3210m CPU 以及双通道 8GB 1600 Mhz 内存。

4 结果分析

测试结果见表1。

表1 校准测试结果

校准算法	初始值	时间/s	残差平方和
GRG	低	315.90 ± 11.28	303.4 ± 0.00
	中	663.11 ± 12.65	170.7 ± 0.00
	高	521.63 ± 12.76	171.1 ± 0.00
多初始值 GRG	低	94597.11 ± 683.97	227.6 ± 0.00
	中	99858.14 ± 1471.37	227.6 ± 0.00
	高	109448.16 ± 1557.62	227.4 ± 0.00
遗传算法	低	1102.47 ± 271.72	271.82 ± 20.01
	中	1427.79 ± 200.63	268.50 ± 30.83
	高	1060.83 ± 251.26	266.61 ± 15.83
遗传算法与 GRG 连用	低	1804.00 ± 125.07	227.47 ± 0.06
	中	1864.25 ± 680.59	227.6 ± 0.42
	高	1985.16 ± 412.71	240.7 ± 11.26
GRG*	低	354.87 ± 4.23	380.7 ± 0.00
	中	482.82 ± 2.27	269.2 ± 0.00
	高	738.62 ± 17.06	231.2 ± 0.00
遗传算法*	低	3575.28 ± 420.61	236.02 ± 7.95
	中	3287.56 ± 669.17	233.56 ± 7.14
	高	3047.44 ± 472.60	240.84 ± 15.83

注：“GRG*”表示添加参数取值范围的 GRG 算法测试组，“遗传算法*”表示“无改进的最大时间”为 1 000 s 的测试组。

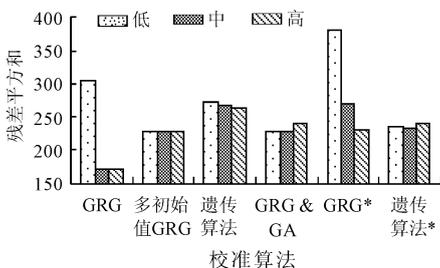


图1 各校准算法在不同初始条件下所得残差平方和

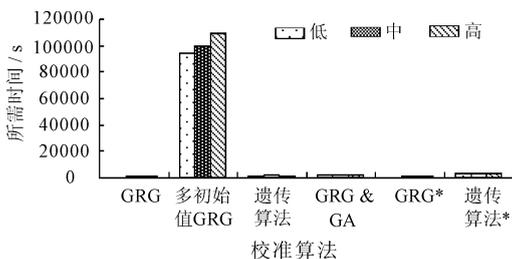


图2 各校准算法在不同初始条件下校准所需时间

(1) ReNuMa 原始校准方法虽然在中值参数组和高值参数组下模拟值和实测值之间的残差平方和最小,但是在它校准之后,存在较多的不合理参数。以中值参数组下的 GRG 校准算法为例,不合理参数如下:

表2 中值参数组下不合理参数

参数名称	校准值	合理值
et cover factor; JAN	1.390995640	0 ~ 1
et cover factor; FEB	1.606781366	0 ~ 1
et cover factor; APR	1.133465896	0 ~ 1
et cover factor; SEPT	1.363598659	0 ~ 1
et cover factor; DEC	2.492486256	0 ~ 1

(2)从图1中可以看出,GRG 算法受初始值影响较大,在不同初始值下,校准结果差异较大,且添加参数范围的 GRG 算法校准结果并不理想。

(3)具有多初始值的 GRG 校准算法,虽然有较好的校准结果,但是,校准耗时较长,见图2。不过其校准后的残差平方和在所有方法中属于较高水平,且无论初始值大小,校准结果十分稳定。

(4)遗传算法相比于设定参数范围的 GRG 算法而言,具备较好的校准效果,但和多初始值 GRG 算法以及遗传算法与 GRG 算法连用的方法相比,还有明显不足。而且其校准结果并不稳定。同时,与“无改进的最大时间”为 1 000 s 的遗传算法组相比,可以得出,延长“无改进的最大时间”,遗传算法的校准结果会进一步提高,见图3。

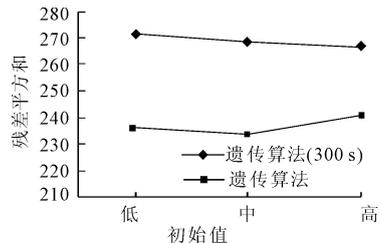


图3 不同“无改进的最大时间”的遗传算法的校准结果对比

(5)遗传算法与 GRG 连用的方法具备相对较好的校准结果,同时所用时间也相对较短,在参测算法中,是比较理想的算法,但是其校准结果并不稳定,这与其前期使用遗传算法有较大关系。

5 结语

对 ReNuMa 模型校准模块的改进能够很好的提高模型校准效果,能够确保校准后参数的合理性。为 ReNuMa 模型在实际应用中取得更好的拟合效果提供支持。

本文主要测试了 ReNuMa 径流模拟的校准,针对不同的模拟指标,ReNuMa 模型在实际应用中,还需要针对溶解态氮,沉积物和总氮进行校准。针对校准指标的不同,校准参数数量的不同,不同参数和目标函数之间关系的不同,往往一种校准算法很难

保证校准结果。根据不同问题的问题特性使用不同的校准算法以及针对同一问题切换不同的校准算法都有可能提高校准结果。本文针对 ReNuMa 校准模块所做的改进,保障了 ReNuMa 参数校准结果的有效性并提高了其对多参数问题的全局寻优能力。

参考文献:

- [1] Sha Jian, Liu Min, Wang Dong, et al. Application of the ReNuMa model in the Sha He river watershed: Tools for watershed environmental management[J]. *Journal of Environmental Management*, 2013,124(30):40-50.
- [2] Gallagher M, Doherty J. Parameter estimation and uncertainty analysis for a watershed model[J]. *Environmental Modeling and Software*, 2007,22(7):1000-1020.
- [3] Eary L E, Jenne E A, Vail L W, et al. Numerical models for predicting watershed acidification[J]. *Archives of Environmental Contamination and Toxicology*, 1989,18(1-2):29-53.
- [4] Swaney D P, Sherman D, Howarth R W. Modeling water, sediment and organic carbon discharges in the Hudson - Mohawk Basin: coupling to terrestrial sources[J]. *Estuaries and Coasts*, 1996,19(4):833-847.
- [5] Lee K Y, Fisher T R, Jordan T E, et al. Modeling the hydrochemistry of the Choptank River Basin using GWLF and Arc/Info: 1. Model calibration and validation[J]. *Biogeochemistry*, 2000,49(2):143-173.
- [6] Lee K Y, Fisher T R, Emma Rochelle - Newall E. Modeling the hydrochemistry of the Choptank River Basin using GWLF and Arc/Info: 2. Model Validation and Application [J]. *Biogeochemistry*, 2001,56(3):311-348.
- [7] 杨光,刘伟江,孙韧,等. 天津市于桥水库饮用水源地保护对策研究[J]. *水资源与水工程学报*, 2011,22(6):43-45+49.
- [8] 何因,秦保平,李云生,等. GWLF 模型的原理、结构及应用[J]. *城市环境与城市生态*, 2009,22(6):24-27.
- [9] 刘敏,谢阳村,王东,等. 基于 ReNuMa 模型的长春石头口门水库流域非点源污染负荷模拟[J]. *水资源与水工程学报*, 2012,23(6):70-72+78.
- [10] 王东,赵越,徐敏,等. 区域营养盐管理模型在沙河流域的应用[J]. *水资源与水工程学报*, 2012,23(5):98-101.
- [11] Stuart Smith, Leon Lasdon. Solving large sparse nonlinear programs using GRG[J]. *ORSA Journal on Computing*. 1992,4(1):2-15.
- [12] Lasdon L S, Waren A D, Jain A, et al. Design and testing of a generalized reduced gradient code for nonlinear programming [J]. *ACM Transactions on Mathematical Software*, 1978,4(1):34-50.
- [13] 葛继科,邱玉辉,吴春明,等. 遗传算法研究综述[J]. *计算机应用研究*, 2008,25(10):2911-2916.
- [14] 李宏,焦永昌,张莉,等. 一种求解全局优化问题的新混合遗传算法[J]. *控制理论与应用*, 2007,24(3):343-348.
- [15] 雷英杰,张善文,李续武,等. MATLAB 遗传算法工具箱及应用[M]. 西安:西安电子科技大学出版社, 2005,4:1-61.
- [16] 郎茂祥,胡思继. 用混合遗传算法求解物流配送路径优化问题的研究[J]. *中国管理科学*, 2002,10(5):51-56.