

文章编号:1000-6788(2006)02-0118-09

# 软计算技术在环境复杂模型参数识别中的应用研究

王建平,程声通

(清华大学环境科学与工程系,北京 100084)

**摘要:** 以 WASP 模型在密云水库水质模拟中的应用为例,研究考察分析了软计算技术用于环境复杂模型参数识别的性能和效率.参数分析表明,WASP 模型存在一些低灵敏参数且受相关参数的影响明显.为对比参数灵敏度和相关性的影响,共设计了 4 组数值试验.数值试验表明,全局搜索法能较好解决参数全局寻优问题,是获取参数辅助信息的重要手段.MCMC 法可有效对参数后验分布进行采样,采样序列稳定收敛到参数后验分布上,同时 MCMC 法较好地处理了相关参数采样的问题.最后给出案例表明基于软计算技术的复杂模型参数识别技术路线是高效的、实用的.

**关键词:** 环境模型;参数识别;软计算技术;案例研究

**中图分类号:** X192

**文献标识码:** A

## Parameter Identification of Complicated Environmental Model Using the Soft-computing Approach

WANG Jian-ping, CHENG Sheng-tong

(Department of Environmental Science and Engineering, Tsinghua University, Beijing 100084, China)

**Abstract:** In order to study the performance and efficiency of soft-computing technology in application to parameter identification of complicated environmental model, a case study was presented, which was an application of water quality simulation in the Miyun Reservoir using WASP model. Parameter analysis indicated that there were many low sensitivity parameters in WASP model and several groups of parameters were correlative apparently. For analyzing effects of sensitivity and correlativity, four numerical experiments were designed. Resulted indicated that global searching method could solve global optimization well and also was an efficient step to acquire auxiliary information of parameters. And MCMC method could sample posterior distributions of parameters effectively; moreover, sampling series could converge to the posterior distributions of parameters eventually. Case study indicated that it was efficient and practicable for the soft-computing approach presented by the paper to identify parameters of complicated model.

**Key words:** environmental model; parameter identification; soft-computing technology; case study

### 1 软计算参数识别技术路线

研究表明<sup>[1,2]</sup>,基于贝叶斯理论的不确定性参数识别思路得到了研究人员的广泛认同,但目前常见的贝叶斯离散化方法尚存在如下不足:1)计算限制和相关性采样问题.参数越多,问题越严重;2)参数先验信息缺乏.马尔科夫链蒙特卡罗法(Markov Chain Monte Carlo, MCMC)是一种基于贝叶斯理论的后验分布采样方法,能高效可靠地采集并收敛到参数的后验分布上,相比 Monte Carlo 法,MCMC 法可大大降低计算量,并可有效解决相关参数采样的问题.同时,要想获取更多的先验知识,文献研究是一方面,更重要的是从实际问题中挖掘信息,参数灵敏度分析、相关性分析和全局最优解搜索等均是非常有用的获取先验信息的手段.基于以上两点,本文提出了如图 1 所示的实用参数识别技术路线,该技术路线以贝叶斯理论为基础,以 MCMC 法为核心,并融合了参数灵敏度分析、相关性分析和参数全局搜索法,充分利用了软计算技术简单实用和全局搜索的优点.

收稿日期:2004-12-21

资助项目:北京市重大科技项目(H012110010119)

作者简介:王建平(1977-),男,内蒙古清水河县人,博士,主要从事环境模拟和水利政策的研究,现在水利部发展研究中心工作, E-mail: wjp@waterinfo.com.cn;通讯联系人:程声通, E-mail: cst-den@tsinghua.edu.cn.

### 1.1 先验知识的获取

先验知识来源包括模型技术文档、国内外相关研究和典型实验数据等,重点掌握参数物理意义、灵敏度特征、取值范围和分布等信息.先验知识的获取是模型研究的基础工作.

### 1.2 参数分析

参数分析一般包括参数灵敏度分析和相关性分析.参数灵敏度和相关性构成了参数识别研究内因分析的两大要素,因此参数分析是参数识别研究中必不可少的环节,也是MCMC法确定参数先验分布和推荐分布的基础工作.

### 1.3 参数搜索

参数搜索是复杂模型参数识别的关键所在,针对复杂模型参数识别的难点,如参数灵敏度低、参数相关以及数据稀缺等问题,本文提出在全局搜索算法的支持下利用MCMC法进行参数后验分布采样,继而进行参数不确定性分析.首先利用全局搜索算法在参数空间内进行粗搜索,确定参数空间的希望区域或重点搜索区域,进一步掌握参数灵敏度和相关性信息.在这些辅助信息的支持下,合理确定参数的先验分布和推荐分布,以使MCMC法高效可靠地获取参数的后验分布.当然,如图1所示,对于结构清楚,参数物理意义明确的模型,可直接利用全局优化算法搜索最优参数.

### 1.4 参数统计

在参数搜索结果的基础上,进行参数识别结果统计与检验.

本文将WASP模型在密云水库水质模拟中的应用为案例,进行参数识别的实际应用研究.

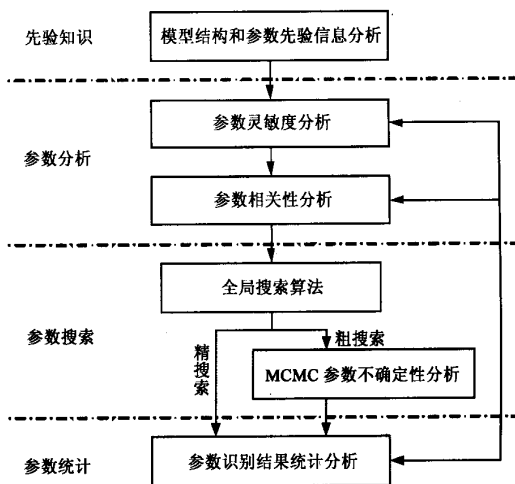


图1 复杂模型参数识别技术路线

## 2 模型与数据

### 2.1 WASP模型

研究采用的水质模型为WASP模型系统.WASP<sup>[3]</sup>是由美国国家环保局开发的用于地表水水质模拟的模型,它提供了一个灵活的动态模拟系统.如图2所示,WASP可以模拟八个指标,分别为:氨氮( $\text{NH}_3$ )、硝酸盐氮( $\text{NO}_3$ )、溶解性磷酸盐( $\text{OPO}_4$ )、叶绿素a( $\text{Chl-a}$ )、碳生化需氧量(CBOD)、溶解氧(DO)、有机氮(ON)和有机磷(OP).在WASP模型系统中水质模块EUTRO5的水质参数有42个之多,通过灵敏度和相关性分析确定了模型待识别参数表,结果如表1所示.

### 2.2 数据序列产生

本文采用合成的“观测”数据进行研究,即在已知参数值的情况下(参数值见表1)应用WASP模型,产生“真实值”时间序列,得到“观测值”时间序列,然后利用这些数据进行模型参数识别.这样做的目的是为了使系统的真实情况在掌握之中,同时模型没有结构上的误差,参数估计的所有误差仅来源于参数初值和参数识别本身.这使得有可能在排除结构误差的情况下,单独讨论研究提出的实用复杂模型参数识别技术路线的识别性能和效率.合成数据方法在许多模型分析中有广泛的使用.为简洁起见,在不引起混淆的情况下,不再对“观测值”和“真实值”加上引号.

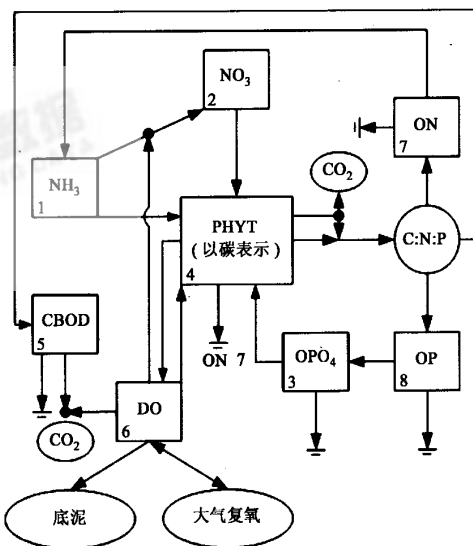


图2 水质模拟反应动力学关系图

用于参数识别的合成数据序列为：

1)没有扰动的数据序列.模型最优解已知,目标函数值为零.后文中简称“没有扰动情况”.

2)两组带有扰动的数据序列.为了模拟观测误差的影响,给观测值(没有扰动情况的数据序列)加上了一定的随机扰动.随机扰动为：

$\xi = N(0, \sigma)$ ,即均值为 0,标准差为  $\sigma$  的正态分布, $\sigma$  分别为各水质状态变量浓度均值的 5% 和 10%.观测值为： $C'(t) = C(t) + \xi$ ,显然,观测误差满足独立、正态和同方差的性质.叙述中分别简称为“5% 扰动情况”和“10% 扰动情况”.

表 1 待识别水质模型参数表

| 参数名称 | 物理意义                                                                    | 参数取值  | 参数范围        |
|------|-------------------------------------------------------------------------|-------|-------------|
| SOD  | 底泥耗氧量, $\text{g}/(\text{m}^2 \cdot \text{day})$                         | 2     | 0 ~ 3       |
| K12C | 20℃ 条件下的硝化速度系数/ $\text{d}^{-1}$                                         | 0.2   | 0.05 ~ 0.35 |
| K20C | 20℃ 条件下的反硝化速度系数/ $\text{d}^{-1}$                                        | 0.05  | 0.01 ~ 0.2  |
| KNO3 | 反硝化氮限制半饱和系数, $\text{mgO}_2/\text{L}$                                    | 0.5   | 0.1 ~ 1.0   |
| K1C  | 浮游植物的饱和生长率/ $\text{d}^{-1}$                                             | 2.5   | 1.5 ~ 4     |
| CCHL | 浮游植物碳与叶绿素的比,仅当 LGHTS = 1 使用<br>$\text{mg C}/\text{mg Chl-a}$ , 缺省值为 30. | 30    | 20 ~ 40     |
| K1RC | 20℃ 条件下浮游植物的内源呼吸速率/ $\text{d}^{-1}$                                     | 0.1   | 0.05 ~ 0.2  |
| NCRB | 浮游植物内的氮碳比, $\text{mg}/\text{mg}$ , 缺省值为 0.25                            | 0.25  | 0.2 ~ 0.3   |
| PCRB | 浮游植物内的磷碳比, $\text{mg}/\text{mg}$ , 缺省值为 0.025                           | 0.025 | 0.02 ~ 0.03 |
| KDC  | 20℃ 条件下的 CBOD 降解速率/ $\text{d}^{-1}$                                     | 0.03  | 0.01 ~ 0.1  |
| K2   | 20℃ 条件下,水体的复氧速度常数/ $\text{d}^{-1}$                                      | 0.15  | 0.1 ~ 0.2   |
| K71C | 溶解有机氮的矿化速度/ $\text{d}^{-1}$                                             | 0.03  | 0.01 ~ 0.1  |
| K83C | 溶解有机磷的矿化速度/ $\text{d}^{-1}$                                             | 0.03  | 0.01 ~ 0.1  |

注:参数范围确定参考文献包括 DiToro and Mtystik 1980 (Lake Huron & Saginaw Bay)<sup>[4]</sup>, Thomann and Fitzpatrick 1982 (Potomac Estuary Model)<sup>[5]</sup>, Bowie et al. 1985 (Various locations)<sup>[6]</sup>.

3 参数分析

3.1 参数灵敏度分析

参数灵敏度分析的方法很多<sup>[7]</sup>, Hamby 对不同方法进行了综述<sup>[7]</sup>, 并进行了数值实验对比<sup>[8]</sup>, 研究将灵敏度分析方法概括为三类:1)基于单变量操作的方法,具体包括差分分析法(Differential Sensitivity Analysis)、单参数分析法(One-at-a-time Sensitivity Methods)、灵敏因子法(Sensitivity Index)等.2)基于输入矩阵和输出向量的分析方法,如散点图法(Scatter Plots)、重要系数法(Importance Index)、偏相关系数法(Partial Correlation Coefficient)等.3)基于输出向量对相应的输入变量进行划分的方法,如 Smirnov 检验法、Mann-Whitney 检验法等.不同的方法

有不同的应用范围,同时实施难度也不同,单参数分析法是最简单的方法,同时也是应用最广泛的一种方法,本研究将采用单参数灵敏度方法进行分析,基于以下三点考虑:1)整套参数识别技术路线简单实用;2)并不要求严格准确的参数灵敏度排序;3)参数识别过程有一个反馈过程,局部灵敏度分析结果可得到检验.

参数灵敏度分析结果按表 2 进行分类.结果表明,模型有 2 个灵敏参数(K1C, K1RC),9 个较灵敏参数(K12C, K20C, KNO3, CCHL, NCRB, PCRB, KDC, K71C, K83C),其余参数均不太灵敏.

3.2 参数相关性分析

环境模型中常用的参数相关性分析方法有两种<sup>[9]</sup>:参数相关系数法和奇异值分解法(Singular Value

表 2 参数灵敏度分类

| 灵敏度   | 灵敏性  | 分类值 |
|-------|------|-----|
| < 0.1 | 不灵敏  | 1   |
| < 0.5 | 一般灵敏 | 2   |
| < 1.0 | 较灵敏  | 3   |
| > 1.0 | 灵敏   | 4   |

Decomposition, SVD 法).Hill et al.2003 对参数相关系数法和 SVD 法的研究指出<sup>[9]</sup>:灵敏矩阵存在误差时,相关系数法很难有效识别参数之间的相关性,同时相关系数法仅能识别两个参数之间的相关性,而不能确定两个以上参数的相关性.SVD 法可有效克服这些问题.故本研究将采用 SVD 法进行参数相关性分析.首先对灵敏矩阵  $X$  进行比例变化,然后进行 SVD 分解.SVD 法的分析结果的奇异值和右奇异向量如表 3 所示.

表 3 SVD 法分析结果表

| 参数   | $v_1$  | $v_2$  | $v_3$  | $v_4$  | $v_5$  | $v_6$  | $v_7$  | $v_8$  | $v_9$  | $v_{10}$ | $v_{11}$ | $v_{12}$ | $v_{13}$ |
|------|--------|--------|--------|--------|--------|--------|--------|--------|--------|----------|----------|----------|----------|
| SOD  | 0.003  | 0.130  | 0.629  | -0.167 | -0.188 | 0.167  | 0.007  | 0.290  | 0.527  | 0.132    | 0.098    | -0.328   | 0.008    |
| K12C | 0.018  | 0.349  | 0.253  | -0.076 | 0.621  | -0.643 | -0.023 | -0.074 | 0.054  | 0.000    | 0.001    | -0.001   | 0.000    |
| K20C | -0.033 | -0.459 | 0.180  | 0.144  | -0.232 | -0.424 | -0.016 | 0.027  | -0.019 | -0.001   | -0.001   | 0.017    | 0.707    |
| KNO3 | -0.033 | -0.459 | 0.180  | 0.146  | -0.231 | -0.423 | -0.016 | 0.028  | -0.019 | 0.000    | 0.001    | -0.017   | -0.707   |
| K1C  | 0.446  | -0.053 | 0.004  | -0.019 | 0.021  | 0.024  | -0.462 | 0.007  | 0.002  | -0.382   | 0.660    | 0.044    | 0.000    |
| CCHL | -0.446 | -0.050 | -0.009 | -0.051 | 0.066  | 0.045  | -0.426 | -0.244 | -0.310 | 0.226    | 0.168    | -0.611   | 0.015    |
| K1RC | -0.438 | 0.047  | -0.004 | 0.017  | -0.019 | -0.027 | 0.606  | -0.040 | -0.038 | -0.143   | 0.641    | 0.047    | 0.000    |
| NCRB | -0.031 | -0.465 | 0.098  | -0.232 | 0.326  | 0.230  | 0.045  | -0.618 | 0.399  | -0.039   | -0.036   | 0.120    | -0.003   |
| PCRB | -0.447 | 0.047  | -0.002 | 0.009  | -0.013 | 0.011  | -0.436 | 0.162  | 0.186  | 0.344    | 0.155    | 0.636    | -0.015   |
| KDC  | -0.001 | -0.023 | 0.250  | 0.880  | 0.296  | 0.271  | -0.004 | -0.017 | 0.005  | 0.005    | 0.004    | -0.012   | 0.001    |
| K2   | -0.003 | -0.130 | -0.631 | 0.170  | 0.159  | -0.166 | 0.004  | 0.250  | 0.568  | 0.124    | 0.092    | -0.307   | 0.008    |
| K71C | 0.031  | 0.441  | -0.078 | 0.251  | -0.498 | -0.205 | -0.069 | -0.601 | 0.282  | 0.003    | 0.004    | -0.001   | 0.000    |
| K83C | 0.455  | -0.049 | 0.002  | -0.012 | 0.016  | -0.003 | 0.197  | -0.120 | -0.156 | 0.794    | 0.285    | 0.036    | 0.000    |
| 奇异值  | 2.179  | 1.950  | 1.486  | 1.002  | 0.778  | 0.611  | 0.464  | 0.189  | 0.070  | 0.022    | 0.010    | 0.005    | 0.0001   |

SVD 分解结果的分析可参考矩阵分析相关的书籍,结果分析如下:

1)表 3 最后一行为奇异值,可以看出  $\sigma_{12}$ 和  $\sigma_{13}$ 较小,尤其是  $\sigma_{13}$ 非常小,基本接近于 0.表明灵敏矩阵中的两列或更多列线性相关,等价于模型的两个或多个参数高度相关.

2)公式(1)~(3)为奇异值最小的后三个分量.由公式(1)可知,参数 K1C 的变化  $\Delta\theta_{K1C}$ 和参数 K1RC 的变化  $\Delta\theta_{K1RC}$ 可通过相互调整,达到对模型输出结果影响很小,说明参数 K1C 和 K1RC 相关.同理,由公式(2)和(3)可知,参数 CCHL、PCRB、NCRB、SOD 和 K2 彼此相关,K20C 和 KNO3 相关.从相关性的影响和识别难易程度来看,K20C 和 KNO3 高度相关,影响最大且最难识别,CCHL、PCRB、SOD 和 K2 其次,最后为 K1C 和 K1RC.

$$\Delta v_{11} = 0.660\Delta\theta_{K1C} + 0.641\Delta\theta_{K1RC} + \cdots, \quad \sigma_{11} = 0.010, \tag{1}$$

$$\Delta v_{12} = -0.611\Delta\theta_{CCHL} + 0.636\Delta\theta_{PCRB} - 0.328\Delta\theta_{SOD} - 0.307\Delta\theta_{K2} + 0.120\Delta\theta_{NCRB} + \cdots, \quad \sigma_{12} = 0.005, \tag{2}$$

$$\Delta v_{13} = 0.707\Delta\theta_{K20C} - 0.707\Delta\theta_{KNO3} + \cdots, \quad \sigma_{13} = 0.0001. \tag{3}$$

综上,本研究中共提取出 13 个待识别参数,如表 1 所示,包括 2 个灵敏参数、9 个较灵敏参数和 2 个低灵敏度参数,其中有数组参数间存在显著的相关性.待识别参数为 10 个的情况是指除 SOD、KNO3 和 CCHL 外的参数集合,目的是考察受灵敏度和相关性影响较小的参数识别效率和性能.

为对比分析不同扰动条件下的参数识别性能和效率,研究设计了以下四组数值试验.

试验 I:在 5% 扰动条件下识别 10 个参数:K12C、K20C、K1C、K1RC、NCRB、PCRB、KDC、K2、K71C 和 K83C.试验重复运行 5 次.

试验 II:在 10% 扰动条件下识别 10 个参数,参数同上.试验重复运行 5 次.

试验 III:在 5% 扰动条件下识别 13 个参数:K12C、K20C、KNO3、K1C、CCHL、K1RC、NCRB、PCRB、KDC、

K2、K71C、K83C 和 SOD. 试验重复运行 5 次.

试验Ⅳ: 在 10% 扰动条件下识别 13 个参数, 参数同上. 试验重复运行 5 次.

4 参数全局搜索

混合算法被认为是求解非线性复杂模型采用随机搜索方法最成功的途径之一<sup>[10]</sup>, 本文将采用应用较广且效率较高的遗传单纯形混合算法进行参数全局优化搜索.

试验Ⅰ~Ⅳ的参数全局搜索结果汇总为表 4 和表 5. 研究为避免随机数产生器种子(seed)的影响, 每一组试验随机模拟了 5 次. 考察 5 次平行试验目标函数值的“标准差/平均值”, “标准差/平均值”反映目标函数值的离散程度, 可指示算法是否收敛到全局最优解. 结果分析如下:

1) 如表 4 和表 5 所示, 目标函数值的“标准差/平均值”范围在 0.0% ~ 0.05%, 5 次平行试验均收敛到或近似收敛到一个目标函数值, 说明算法全局收敛.

表 4 参数全局搜索结果(一)

| 参数名称 | 参数最优值 | 试验Ⅰ    |       |        |        | 试验Ⅱ    |       |        |        |
|------|-------|--------|-------|--------|--------|--------|-------|--------|--------|
|      |       | 平均值    | 方差    | 相对误差   | 方差/平均值 | 平均值    | 方差    | 相对误差   | 方差/平均值 |
| K12C | 0.2   | 0.219  | 0.001 | 9.52%  | 0.26%  | 0.196  | 0.000 | 1.99%  | 0.02%  |
| K20C | 0.05  | 0.053  | 0.000 | 5.56%  | 0.08%  | 0.063  | 0.000 | 25.84% | 0.00%  |
| K1C  | 2.5   | 2.419  | 0.000 | 3.25%  | 0.02%  | 2.764  | 0.000 | 10.57% | 0.00%  |
| K1RC | 0.1   | 0.097  | 0.000 | 3.27%  | 0.02%  | 0.112  | 0.000 | 11.98% | 0.00%  |
| NCRB | 0.25  | 0.262  | 0.000 | 4.99%  | 0.11%  | 0.220  | 0.000 | 12.03% | 0.01%  |
| PCRB | 0.025 | 0.027  | 0.000 | 9.90%  | 0.06%  | 0.025  | 0.000 | 0.06%  | 0.02%  |
| KDC  | 0.03  | 0.029  | 0.000 | 1.77%  | 0.10%  | 0.032  | 0.000 | 7.71%  | 0.06%  |
| K2   | 0.15  | 0.185  | 0.000 | 23.30% | 0.10%  | 0.158  | 0.000 | 5.00%  | 0.01%  |
| K71C | 0.03  | 0.030  | 0.000 | 0.45%  | 0.06%  | 0.029  | 0.000 | 3.73%  | 0.03%  |
| K83C | 0.03  | 0.032  | 0.000 | 6.73%  | 0.03%  | 0.032  | 0.000 | 6.98%  | 0.02%  |
| 目标值  |       | 79.681 | 0.002 | -      | 0.00%  | 350.57 | 0     | -      | 0.00%  |

注: “参数最优值”为数据序列合成时的参数取值; “平均值”为 5 次试验参数识别结果的平均值; “标准差”为 5 次试验参数识别结果的标准偏差; 相对误差为|优化参数平均值 - 参数最优值|/参数最优值. “目标值”为算法优化结果的目标函数值<sup>[11]</sup>. 相对误差反映了参数优化结果偏离“最优目标值”的距离, 而标准差/平均值反映了参数优化结果的离散程度, 离散程度太大, 说明参数识别不唯一.

2) 结果表明, 存在观测误差的情况下, 遗传单纯形混合算法能够较好进行参数全局搜索, 除个别参数外, 大多数参数的相对误差能控制在 15% 以下, 基本趋势是扰动越大, 参数识别误差越大. 个别参数误差较大的主要原因是参数灵敏度和相关性的影响.

3) 参数灵敏度的影响: 理论上分析表明, 在足够的观测数据下, 只要参数的变化产生了模拟结果的变化, 不管多小, 该参数都是可以识别的. 本研究的数值试验在一定程度上也验证了这一点结论. 然而在实际应用中, 一方面数据非常有限甚至会稀缺, 另一方面优化算法也不可能无休止地搜索, 因此参数灵敏度需要达到一定量级才能有效识别. 然而, 究竟灵敏度多大才能识别, 很难定量回答, 在全局搜索算法的支持下, 本研究给出一个定性描述: 只要参数变化引起的目标函数变化大于全局算法在一定迭代次数下的优化误差, 参数就可有效识别. 需要说明的是, 以上结论是在目标函数最优或接近最优的情况下得出的, 若目标函数允许一定范围内的误差, 可以理解, 低灵敏度参数的不确定性会很大.

4) 参数相关性的影响: 数值试验结果表明, 参数相关会导致参数估值不唯一. 具体表现在几方面: ①参数相关导致各参数优化结果的“标准差/平均值”整体明显增大, 对比表 4 和表 5 即可得出; ②3.2 节相关性



分析指出的几组相关参数,它们的“标准差/平均值”要大于其它参数,同时 5 次平行试验的参数优化结果也证明了相关性的存在,如图 3~图 6,相关系数均在 0.99 以上;③几个独立参数或受相关性影响不太严重的参数,如 K12C、KDC、K71C 和 K83C 等,均可有效识别.同时数值试验表明参数相关性显著降低了算法搜索效率.

表 5 参数全局搜索结果表(二)

| 参数名称 | 参数最优值 | 试验Ⅲ    |       |        |        | 试验Ⅳ    |       |        |        |
|------|-------|--------|-------|--------|--------|--------|-------|--------|--------|
|      |       | 平均值    | 方差    | 相对误差   | 方差/平均值 | 平均值    | 方差    | 相对误差   | 方差/平均值 |
| K12C | 0.2   | 0.217  | 0.002 | 8.55%  | 0.95%  | 0.200  | 0.003 | 0.06%  | 1.55%  |
| K20C | 0.05  | 0.071  | 0.018 | 42.77% | 25.83% | 0.085  | 0.026 | 69.62% | 30.74% |
| KNO3 | 0.5   | 0.386  | 0.118 | 22.70% | 30.46% | 0.405  | 0.141 | 18.98% | 34.75% |
| K1C  | 2.5   | 2.419  | 0.002 | 3.25%  | 0.09%  | 2.759  | 0.006 | 10.38% | 0.21%  |
| CCHL | 30    | 29.309 | 1.914 | 2.30%  | 6.53%  | 31.206 | 2.081 | 4.02%  | 6.67%  |
| K1RC | 0.1   | 0.097  | 0.000 | 3.27%  | 0.11%  | 0.112  | 0.000 | 11.76% | 0.25%  |
| NCRB | 0.25  | 0.269  | 0.018 | 7.44%  | 6.60%  | 0.216  | 0.016 | 13.45% | 7.56%  |
| PCRB | 0.025 | 0.028  | 0.002 | 12.72% | 6.43%  | 0.024  | 0.002 | 2.59%  | 7.79%  |
| KDC  | 0.03  | 0.029  | 0.000 | 2.53%  | 1.00%  | 0.034  | 0.001 | 12.11% | 1.63%  |
| K2   | 0.15  | 0.200  | 0.000 | 33.31% | 0.04%  | 0.127  | 0.009 | 15.53% | 7.35%  |
| K71C | 0.03  | 0.030  | 0.000 | 1.02%  | 0.50%  | 0.030  | 0.000 | 1.61%  | 0.86%  |
| K83C | 0.03  | 0.032  | 0.000 | 6.54%  | 0.14%  | 0.032  | 0.000 | 7.71%  | 1.30%  |
| SOD  | 2     | 2.342  | 0.006 | 17.12% | 0.24%  | 1.397  | 0.157 | 30.14% | 11.21% |
| 目标值  |       | 79.23  | 0.037 | -      | 0.05%  | 349.70 | 0.159 | -      | 0.05%  |

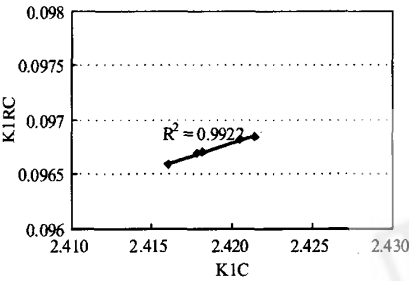


图 3 K1C-K1RC 关系图(试验 I)

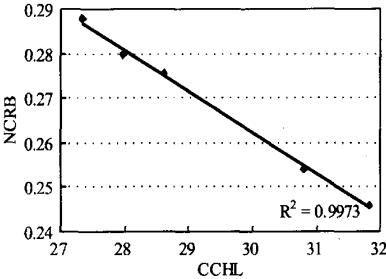


图 5 CCHL-NCRB 关系图(试验 I)

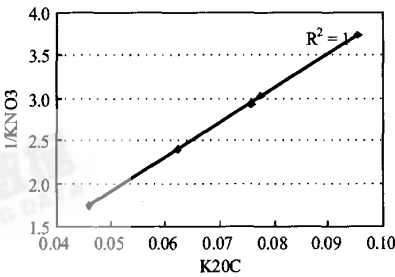


图 4 K20C-1/KNO3 关系图(试验 I)

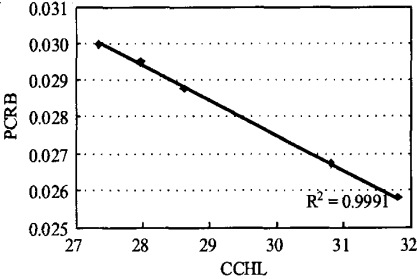


图 6 CCHL-PCRB 关系图(试验 I)

数值试验表明,启发式搜索法为通用全局搜索技术,能较好解决全局寻优问题,是获取参数辅助信息

的重要手段。

5 参数不确定性分析

文献中常用的采样算法有<sup>[12]</sup>:Metropolis-Hastings 算法和自适应 Metropolis 算法。Metropolis-Hastings 算法(简称 M-H 法)是基于贝叶斯推理框架下描述参数不确定性的最早、最通用的一类 MCMC 采样器。自适应 Metropolis 算法(Adaptive Metropolis, 简称 AM 法)是 Haario 在 2001 年提出的一种改进型 MCMC 采样器<sup>[12]</sup>。相比传统 M-H 算法,AM 算法不再需要事先确定参数的推荐分布,而是由后验参数的协方差矩阵来估算,而且参数同步更新,采样效率高。本研究采用 AM 法进行参数后验分布采样。参数搜索范围如表 1 所示,每次采样 20000 次,平行运行 5 次,初始化阶段为 10000 次,这样 5 次平行试验共采集了 50000 个样本用于参数后验分布的统计分析。MCMC 法参数后验分布采样结果分析如下:

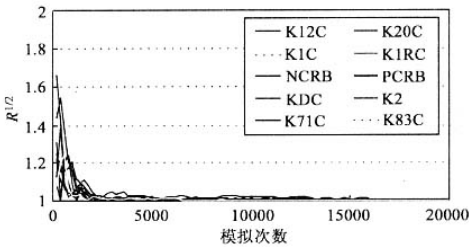


图 7 比例缩小得分  $R^{1/2}$  演化过程

1) Gelman 采样序列收敛判断准则<sup>[13]</sup> 计算表明,如图 7 所示,在搜索初期,即迭代次数  $i < 5000$ ,不同参数的  $\sqrt{R}$  (比例缩小得分, Scale Reduction Score) 变化剧烈,从 1.6 ~ 2.0 快速下降到 1.05 左右,  $i > 5000$  后,  $\sqrt{R}$  继续缓慢下降并最终稳定到一个略大于 1.0 的数值上。各参数均呈现出这一规律,说明不同参数的 MCMC 采样序列均能稳定收敛到参数的后验分布上。

2) 参数后验分布统计结果如表 6 所示,以试验 I 为例给出。表中分别统计了平均值、众数、标准差、斜度、最大值、最小值以及 7 个百分点数据,其中众数不是通常意义上序列中出现概率最高的数,而是将参数空间等分为 25 组,取概率最高的一组的中值为众数。百分点数据有助于确定置信区间,进行参数的不确定性分析,如 90% 置信度的参数区间为 [5% 百分点, 95% 百分点], 80% 置信度的参数区间为 [10% 百分点, 90% 百分点], 其它同理。表 6 中的最小值和最大值与表 1 中的参数搜索范围基本一致,说明参数空间得到有效搜索。另外,表 6 中的平均值和众数存在显著差别,说明参数后验分布为非对称分布,大多数参数呈偏正态分布,斜度值也说明了这一点,因此进行参数不确定性分析时,采用众数分析相对合理。

表 6 参数后验分布统计表

| 参数   | 样本数   | 平均值    | 众数     | 标准差    | 斜度      | 最小值    | 最大值    | 百分点    |        |        |        |        |        |        |
|------|-------|--------|--------|--------|---------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
|      |       |        |        |        |         |        |        | 5      | 10     | 25     | 50     | 75     | 90     | 95     |
| K12C | 50000 | 0.2187 | 0.2000 | 0.0762 | -0.1575 | 0.0501 | 0.3500 | 0.0890 | 0.1145 | 0.1603 | 0.2204 | 0.2814 | 0.3213 | 0.3357 |
| K20C | 50000 | 0.0660 | 0.0518 | 0.0364 | 1.4486  | 0.0101 | 0.1998 | 0.0228 | 0.0305 | 0.0428 | 0.0566 | 0.0783 | 0.1170 | 0.1478 |
| K1C  | 50000 | 2.8747 | 2.5500 | 0.6006 | -0.0471 | 1.5010 | 3.9999 | 1.8732 | 2.0771 | 2.4303 | 2.8615 | 3.3637 | 3.7051 | 3.8456 |
| K1RC | 50000 | 0.1208 | 0.1010 | 0.0338 | 0.2504  | 0.0500 | 0.2000 | 0.0681 | 0.0780 | 0.0959 | 0.1176 | 0.1444 | 0.1686 | 0.1819 |
| NCRB | 50000 | 0.2491 | 0.2340 | 0.0288 | 0.0356  | 0.2000 | 0.3000 | 0.2048 | 0.2095 | 0.2242 | 0.2483 | 0.2740 | 0.2893 | 0.2944 |
| PCRB | 50000 | 0.0253 | 0.0274 | 0.0028 | -0.1216 | 0.0200 | 0.0300 | 0.0206 | 0.0212 | 0.0228 | 0.0254 | 0.0277 | 0.0290 | 0.0295 |
| KDC  | 50000 | 0.0422 | 0.0370 | 0.0225 | 0.8603  | 0.0100 | 0.1000 | 0.0141 | 0.0179 | 0.0256 | 0.0355 | 0.0549 | 0.0794 | 0.0891 |
| K2   | 50000 | 0.1542 | 0.1580 | 0.0275 | -0.1762 | 0.1000 | 0.2000 | 0.1078 | 0.1144 | 0.1318 | 0.1564 | 0.1778 | 0.1906 | 0.1948 |
| K71C | 50000 | 0.0457 | 0.0262 | 0.0225 | 0.6897  | 0.0100 | 0.1000 | 0.0173 | 0.0212 | 0.0281 | 0.0401 | 0.0610 | 0.0817 | 0.0907 |
| K83C | 50000 | 0.0387 | 0.0334 | 0.0155 | 1.5413  | 0.0100 | 0.1000 | 0.0210 | 0.0241 | 0.0290 | 0.0347 | 0.0438 | 0.0597 | 0.0732 |

(3) 图 8 ~ 图 11 为 AM 算法全局收敛情况下的部分参数后验分布直方图,以参数 K20C 和 K2 为例给出。K2 为低灵敏度参数。参数 K20C 较灵敏,在 10 个待识别参数情况下受相关性影响较小,而在 13 个待识别参数情况下与参数 KNO3 显著相关。从图 8(a)和图 9(a)可以看出,在试验 I 和 II 中,由于受参数相关性

的影响较小,参数 K20C 的不确定性较小,峰值明显,但随着扰动的增大,参数不确定性也在增大.图 10(a)和图 11(a),由于受相关性参数 KNO3 的影响,参数不确定性明显增大.参数灵敏度的影响在参数后验分布中也可反应出来,如图 8(b)~图 11(b)中参数 K2 的后验分布,低灵敏度导致参数的不确定性增大,基本呈均匀分布,而且扰动越大,不确定性越大.

可见,MCMC 法能有效搜索并收敛到参数后验分布,适合于复杂模型的参数识别和不确定性分析研究.

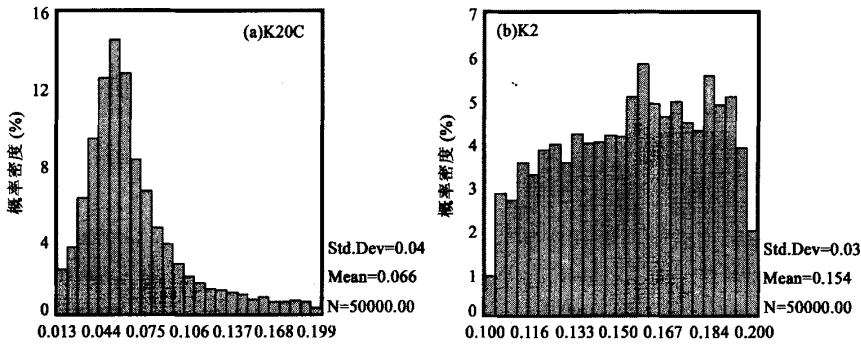


图 8 试验 I 参数 K20C 和 K2 的后验分布

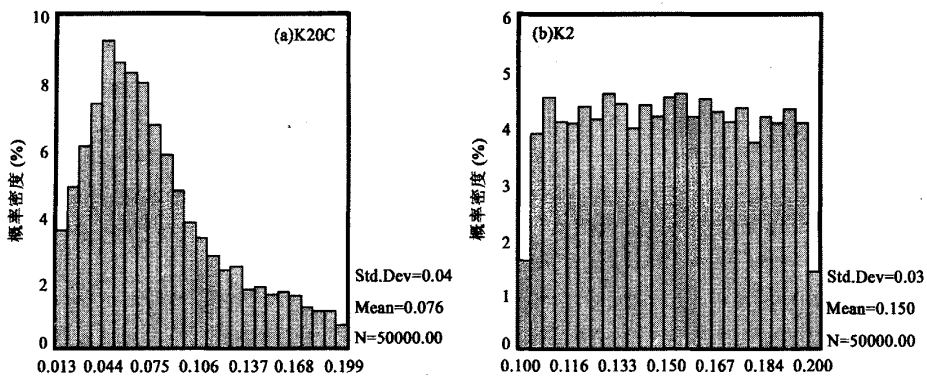


图 9 试验 II 参数 K20C 和 K2 的后验分布

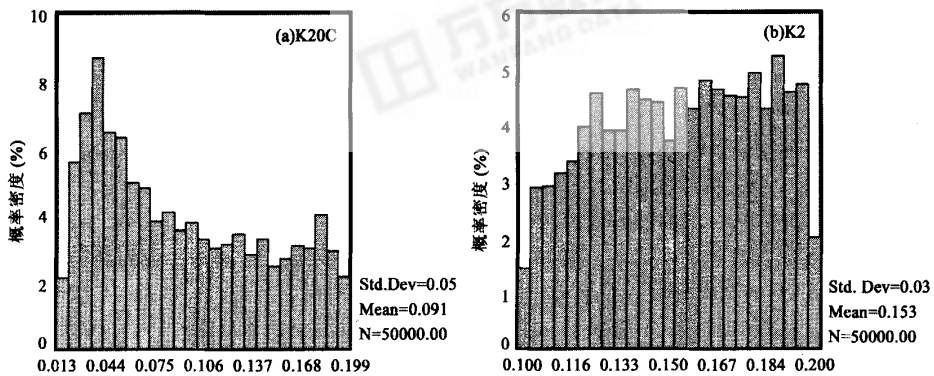


图 10 试验 III 参数 K20C 和 K2 的后验分布

## 6 小结

1)研究以 WASP 模型在密云水库水质模拟中的应用为例进行了数值试验.结果表明,启发式搜索法为通用全局搜索技术,能较好解决全局寻优问题,是获取参数辅助信息的重要手段.



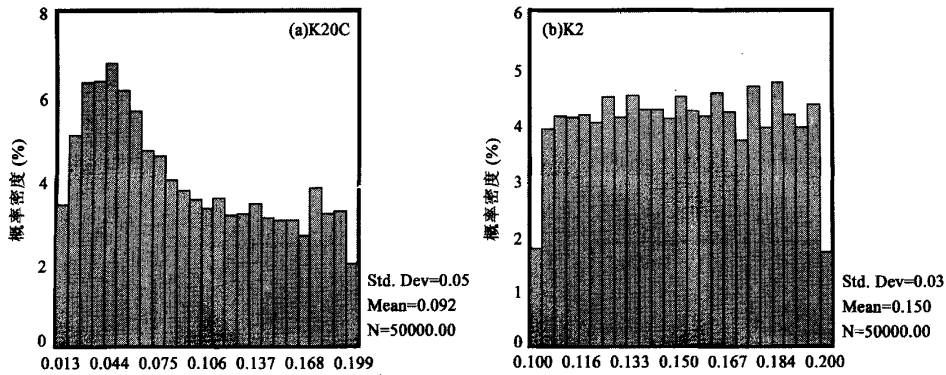


图 11 试验Ⅳ参数 K20C 和 K2 的后验分布

2) 案例研究表明, MCMC 法可有效对参数后验分布进行采样, 采样序列稳定收敛到参数后验分布上, 同时 MCMC 法较好地处理了参数灵敏度和相关参数采样的问题. 此外, 参数后验分布的最小值和最大值与参数搜索范围基本一致, 说明参数空间得到有效搜索.

3) 综上所述, 本文提出的基于软计算技术的复杂模型参数识别技术路线是高效的、实用的. 技术路线的理论基础是贝叶斯推理, 核心方法是 MCMC 法, 并辅以灵敏度分析、相关性分析和全局搜索算法等.

#### 参考文献:

- [1] Reichert P, Omlin M. On the usefulness of overparameterized ecological models[J]. *Ecological Modelling*, 1997, 95: 289 - 299.
- [2] Tierney L. Markov-chains for exploring posterior distributions[J]. *Annals of Statistics*, 1994, 22: 1701 - 1762.
- [3] Ambrose R B, Wool T A, Martin J L, et al. WASP5.x, A Hydrodynamic and Water Quality Model Model Theory, User's Manual, and Programmer's Guide[M]. Draft: Environmental Research Laboratory, US Environmental Protection Agency, 1993.
- [4] DiToro D M, Matystik J W F. Mathematical Models of Water Quality in Large Lakes \_ Part 1: Lake Huron and Saginaw Bay. U.S [R]. EPA/600/3-80-056. Duluth, Minnesota: Environmental Protection Agency, 1980.
- [5] Thomann R V, Fitzpatrick J J. Calibration and Verification of a Mathematical Model of the Eutrophication of the Potomac Estuary [R]. Prepared for Department of Environmental Services. Washington, D.C.: Government of the District of Columbia, 1982.
- [6] Bowie G L, Mills W B, Porcella D B, et al. Rates, Constants, and Kinetic Formulations in Surface Water Quality Modeling, Second Edition[R]. EPA/600/3-85/040. Athens: U. S. Environmental Protection Agency, 1985.
- [7] Hamby D M. A review of techniques for parameter sensitivity analysis of environmental models[J]. *Environmental Monitoring and Assessment*, 1994, 32: 35 - 154.
- [8] Hamby D M. A comparison of sensitivity analysis techniques[J]. *Health Physics*, 1995, 68(2): 195 - 204.
- [9] Hill M C, Osterby O. Determining extreme parameter correlation in ground water models[J]. *Ground Water*, 2003, 41(4): 420 - 430.
- [10] Duan Q, Sorooshian S, Gupta V K. Effective and efficient global optimization for conceptual rainfall-runoff models[J]. *Water Resources Research*, 1992, 28: 1015 - 1031.
- [11] 王建平, 程声通, 贾海峰. 水质模型参数优化的遗传算法实现及控制参数分析[J]. *环境科学*, 2005, 26(3): 65 - 69. Wang J P, Cheng S T, Jia H F. Parameter optimization of water quality model: Implementation of genetic algorithm and its control parameters Analysis[J]. *Environmental Science*, 2005, 26(3): 65 - 69.
- [12] Haario H, Saksman E, Tamminem J. An adaptive metropolis algorithm[J]. *Bernoulli*, 2001, 7(2): 223 - 242.
- [13] Gelman A, Rubin D B. Inference from iterative simulation using multiple sequences[J]. *Statistics Science*, 1992, 7(4): 457 - 511.