

# 优化、RSA 和 GLUE 方法在非线性和环境模型参数识别中的比较

邓义祥<sup>1</sup>, 王琦<sup>2</sup>, 赖斯芸<sup>1</sup>, 陈吉宁<sup>1\*</sup> (1. 清华大学环境科学与工程系, 北京 100084, E-mail: jchen1@tsinghua.edu.cn; 2. 沈阳航空工业学院, 沈阳 110034)

**摘要:** 参数识别是数学模型应用的前提, 本文对非线性环境模型常用的 3 种参数识别方法进行了比较分析. 优化方法是出现最早、应用最广泛的参数识别方法之一. 但在观测误差存在的情况下, 采用优化方法识别的最优参数进行模型预测, 存在很大的决策风险. 考虑到这种不足, RSA 方法和 GLUE 方法摒弃了识别单一最优参数的传统思维, 而把识别参数扩大到多点组成的参数集. RSA 方法与 GLUE 方法不同的是, RSA 方法把可行的参数点看成是同等接受的, 而 GLUE 方法则根据模拟值与实测值的差别, 确定其似然度, 代表参数的可信度水平. 除参数识别以外, RSA 方法和 GLUE 方法也是全局灵敏度分析的重要工具.

**关键词:** 参数识别; 不确定性分析; 优化方法; RSA 方法; GLUE 方法

中图分类号: X143 文献标识码: A 文章编号: 0250-3301(2003)06-07-0009

## Comparison of Optimum, RSA and GLUE Methods in Parameter Identification of a Nonlinear Environmental Model

Deng Yixiang<sup>1</sup>, Wang Qi<sup>2</sup>, Lai Siyun<sup>1</sup>, Chen Jining<sup>1</sup> (1. Department of Environmental Sciences and Engineering, Tsinghua University, Beijing 100084, China E-mail: jchen1@tsinghua.edu.cn; 2. Shenyang Institute of Aeronautical Engineering)

**Abstract:** Parameter identification plays a key role in environmental model application. The optimization method is one of the earliest and most widely used methods. However, as the parameters by optimization may not fully fit the observations, there is a risk that the errors may be enhanced in the decision make stage. With this deficiency in consideration, the RSA and GLUE algorithms search for the feasible parameters not only to the optimum but also around the neighbors. The difference between RSA and GLUE is that the RSA accepts the estimated parameters equally as the candidates for application; while the GLUE keeps the difference among the parameters as measured by likelihood. In addition for parameter identification, both RSA and GLUE are efficient tools for global sensitivity analysis.

**Keywords:** parameter identification; uncertainty analysis; optimization; RSA; GLUE

参数识别是数学模型应用的前提. 优化方法是参数识别最常用的方法之一. 但由于数据的稀疏性和观测误差的存在, 仅仅依赖优化方法所得到的最优参数进行预测, 具有无法把握的风险<sup>[1]</sup>. 考虑到优化方法在稀疏数据条件下进行模型识别的困难, 在环境数学模型的发展中, 先后出现了 RSA 和 GLUE 2 种方法, 两者既有相似的特点, 也有显著的差别. 优化方法、RSA 方法和 GLUE 方法, 是目前环境数学模型应用较多的 3 种参数识别方法, 它们体现了不同的参数识别思想, 具有一定的代表性. 本文对这些方法的异同进行了比较和探讨.

### 1 方法介绍

优化方法是最直观、最传统的参数识别方法之一. 经过不断的发展, 涌现了大量的优化技术, 例如梯度方法、单纯性方法等<sup>[2,3]</sup>. 近年来, 由于计算速度的提高, 基于随机搜索的现代优化算法在非线性和环境模型的参数识别中日益受到欢迎, 例如禁忌搜索算法、模拟退火算法、遗传算

基金项目: 国家环境保护总局资助项目: 985 项目

作者简介: 邓义祥(1974~), 男, 博士研究生, 主要从事环境数学模型的研究.

收稿日期: 2003-01-14; 修订日期: 2003-05-26

\* 通讯联系人

法和人工神经网络算法等<sup>[4]</sup>.各种优化方法的有效性和收敛速度各有不同,但本质是一样的.本文不对这些方法进行比较,而是采用 SCE-UA 方法(Shuffled Complex Evolution methods Developed at the University of Arizona).SCE 方法是 Duan 等人于 1992 年在吸收了控制随机搜索方法和遗传算法的基础上提出的<sup>[5]</sup>,1993 年对其进行了修正<sup>[6,7]</sup>,称为 SCE-UA 方法.SCE-UA 方法在水文学模型中得到了广泛应用,是到目前为止对非线性复杂模型采用随机搜索方法寻找最优值最成功的方法之一<sup>[8,9]</sup>.

20 世纪 70 年代末、80 年代初,由于认识模型识别的困难,Hornberger 和 Spear 将过于强硬的优化条件弱化,即将其转化为一些可以用定量或定性的语言描述的条件来决定参数的取舍,从而在一定程度上克服了采用优化方法进行参数识别带来的不确定性问题,这就是 RSA 方法(Regionalized Sensitivity Analysis)<sup>[10]</sup>.RSA 方法是基于行为和非行为的二元划分进行参数识别的,换句话说,给定一组参数,如果系统的模拟行为满足事先设定的条件,那么这组参数就是可接受的,否则是不可接受的<sup>[11,12]</sup>.

GLUE(Generalized Likelihood Uncertainty Estimation)方法 1992 年由 Beven 提出,它吸收了 RSA 方法和模糊数学方法的优点<sup>[13]</sup>.GLUE 方法认为与实测值最接近的模拟值所对应的参数具有最高的可信度,离实测值越远,可信度越低,似然度越小.当模拟值与实测值的距离大于规定的指标时,就认为这些参数的似然度为 0.可见,GLUE 方法不同于 RSA 方法对参数集“是”和“否”的二元划分,而是采用似然度对不同的参数进行区分.GLUE 方法既考虑到最优即最好这一直观事实,也避免了采用单一的最优值进行预测而带来的风险<sup>[14,15]</sup>.

以上 3 种方法都依赖于随机采样技术.为提高采样效率,本文采用 Latin Hypercube<sup>[16]</sup>采样方法替代了传统的 Monte Carlo 采样.

## 2 模型描述

考虑一个简单的子过程方程,即污染物衰减模型:

$$\frac{dc}{dt} = -k'c \quad (1)$$

式中, $c$  是污染物的浓度(mg/L), $k'$  是衰减系数(1/d), $t$  是时间(d).

这里假设  $k'$  不是常数,而与污染物浓度有关:

$$k' = \frac{kc}{c + M} \quad (2)$$

式中, $k$  是系数(1/d), $M$  是限制因子(mg/L).

根据式(2),当  $c \gg M$  时, $k'$  近似为常数;当  $c \ll M$  时,则  $k'$  近似与  $c$  成正比.

联立(1)、(2)式:

$$\frac{dc}{dt} = -\frac{kc^2}{c + M} \quad (3)$$

积分可得:

$$-\ln c + \frac{M}{c} = kt - \ln c_0 + \frac{M}{c_0} \quad (4)$$

其中, $c_0$  为初始时刻污染物的浓度(mg/L).

注意到式(4)中  $c$  是  $t$  的单调函数,因此在  $k$ 、 $M$  和  $c_0$  已知的情况下, $c$  是可唯一确定的.因此记  $c$  为  $t$  的函数:

$$c = c(c_0, k, M, t) \quad (5)$$

选择该模型作为研究对象,主要考虑到以下几个方面因素:首先,它是许多环境模型的组成部分,在实际中有广泛的应用;其次模型是动态和非线性的,参数之间具有较强的相互作用;最后模型的结构比较简单,参数之间的关系容易把握.

## 3 数据序列的产生

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采用(4)式,产生“真实值”时间序列,然后加上随机扰动以模拟观测误差,得到“观测值”时间序列,然后利用这些数据进行模型的识别和验证.这样做的目的是为了系统的真实情况在掌握之中,同时模型没有结构上的误差,参数估计的所有的误差仅来源于初值和观测系列.这使得有可能在排除结构误差的情况下,单独研究观测误差对于参数识别的影响.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.

令  $c_0 = 20 \text{ mg/L}$ ,  $k = 0.2 / \text{d}$ ,  $M = 10 \text{ mg/L}$ , 由于方程(4)是一个非线性方程,因此对于给定的时刻  $t$ ,采用牛顿方法解非线性方程求出  $c$  值<sup>[17]</sup>,再加上均值为 0、标准方差为  $1 \text{ mg/g}$  的随机扰动,即扰动项为:

$$\xi \sim N(0, 1) \tag{6}$$

观测值  $c'$  为:

$$c'(t) = c(t) + \xi \tag{7}$$

显然,观测误差满足独立、正态和同方差的性质.

#### 4 目标函数

由于具有多个时刻的采样值,因此必须将多目标转化为单目标.考虑到上述观测误差是正态的独立同分布,此时最大似然方法与最小二乘法是等价的,即目标函数为模拟值与实测值差的平方和.大量的经验表明,如果观测误差满足上述关系,则最小二乘法目标函数是最优的目标函数.由于模型没有结构误差,本文不再对方法进行相关性分析.

由于初值对于模拟结果具有重要的影响,因此如果直接采用观测初值进行模拟,可能会带来很大的误差,因此,将  $c_0$  也作为参数进行识别.参数取值空间为:

$$\Omega = \left\{ (k, M, c_0) \mid 0 \leq k \leq 2, \right. \\ \left. 0 \leq M \leq 100, 15 \leq c_0 \leq 30 \right\} \tag{8}$$

### 5 识别结果

#### 5.1 优化方法

采用真实值进行识别,得  $k = 0.20$ 、 $M = 10.00$ 、 $c_0 = 10.02$ ,说明在没有观测误差的情况下,系统能够很好地被识别,优化方法是有效的.采用具有观测误差的数据进行参数识别,结果为: $k = 0.26$ 、 $M = 15.08$ 、 $c_0 = 20.05$ ,误差分别为 30.5%、50.7%和 0.25%,模型对于  $c_0$  的识别结果与真实值非常接近,而其它 2 个参数则具有一定的误差.优化方法拟合时观测值与最佳拟合曲线的对比见图 1,从图中可以看出,观测值基本上在拟合曲线上分布比较均匀,且误差均比较小.

令  $c_0 = 100 \text{ mg/L}$ ,将  $k$  和  $M$  值的识别结果

代入模型进行检验,模拟值与真实值的最大差别为 19.92%,见图 2.

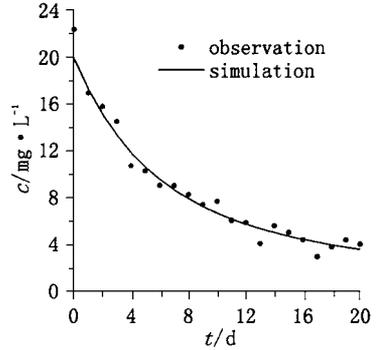


图 1 优化方法的参数识别

Fig. 1 Parameter identification by optimum method

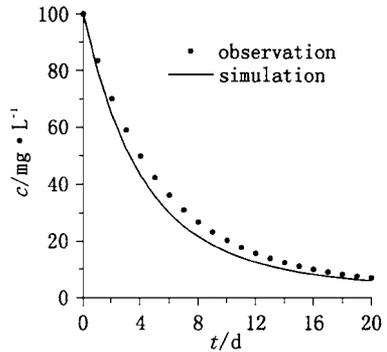


图 2 优化方法的模型验证

Fig. 2 Validation of optimum method

#### 5.2 RSA 方法

采用 Latin Hypercube 方法,在参数空间中进行随机采样,计算每组参数对应的目标函数.采样 1 万次,取目标函数最小 5% 所对应的参数为可接受的参数.经过识别后,各参数的频率直方图见图 3.

从图 3 可以看出,参数分布的区间分别为: $k = [0.1, 1.5]$ ,  $M = [5, 100]$ ,  $c_0 = [16.5, 26.5]$ ,而概率最大的区间是  $k = [0.8, 1.0]$ ,  $M = [80, 90]$ ,  $c_0 = [21, 22]$ ,可见,在目标函数取前 5% 的情况下,参数的取值范围分别占了取值空间的 70%、95% 和 33.3%,因此可以认为在这样的观测数据下,模型的可识别性不高.另外,对于出现概率(频率)最大的值,除了  $M$  值与真实值比较接近以外,其它 2 个值都远离于

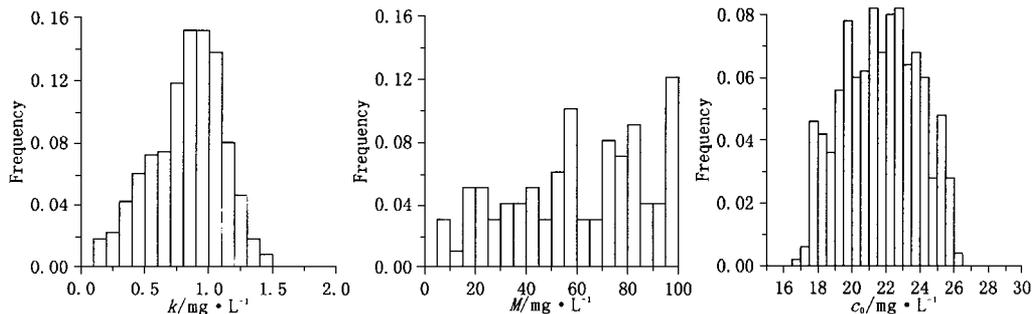


图 3 RSA 参数识别的分布

Fig. 3 Marginal distribution of identified parameters by RSA

真实值,因此,RSA 方法识别的概率最大的参数与真实的参数存在很大的偏差.图 4 中,真实值所在的点(0.2,10)位于参数分布区域的一角.由于这个区间的点数相对较少,因此其概率反而较小.如果只是采用识别概率最大的参数值进行预测,显然与真实值具有相当的偏差.

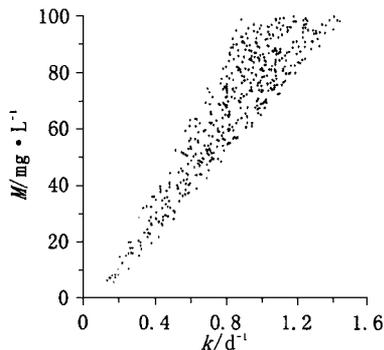
图 4 RSA 可行参数  $k$ - $M$  的对应关系Fig. 4  $k$ - $M$  points of RSA behavioral parameters

图 5 是当初始浓度  $c_0 = 100 \text{ mg/L}$  时的模

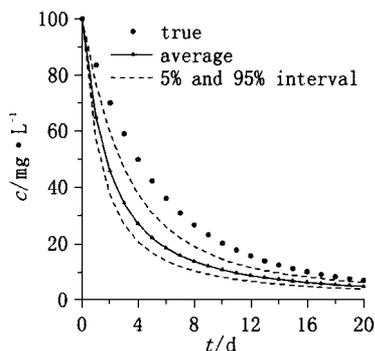


图 5 RSA 方法的模型验证

Fig. 5 Validation of RSA

型验证.从图 5 中可以看出,预测结果的范围能够完全覆盖真实值,而预测的平均值与真实值具有相当的偏差.这说明,RSA 方法能够有效地估计预测结果的风险,但由此失去了一定的预测精度.另外,从图 5 可以看出,如果采用平均值进行预测,那么与优化方法一样具有严重的风险.

除了参数识别,RSA 还具有全局灵敏度分析的功能.图 6 是可接受参数的边缘分布与原始均匀分布的对比.如果参数对目标函数具有较大的影响,则目标函数对参数应具有较强的筛选能力.因此,可接受参数的分布离原始分布越远,说明该参数对目标函数的影响越显著,其灵敏度越高,重要性越突出.

这一结论受到参数之间相关性的影响.如果某个参数的变化可以通过其它参数的变化抵消,那么有可能导致可接受参数值的分布与原始分布的差别并不是很大.但是,由于参数的取值是在封闭的区间,参数在整个区间的取值不太可能被其它参数同等地抵消,因此 RSA 方法仍不失为一种有效的全局灵敏度分析方法.

图 6 中,可接受参数集的分布与真实分布之间的最大差异,即  $K$ - $S$  统计量分别为: $k$  0.34,  $M$  0.27,  $c_0$  0.27.因此,根据 RSA 方法,3 个参数之间的重要性没有显著的差别.

### 5.3 GLUE 方法

GLUE 方法根据似然度对参数进行区分. Beven 等人建议的似然度的表达形式比较复杂<sup>[13,14]</sup>,这里只采用比较简单的形式,即似然

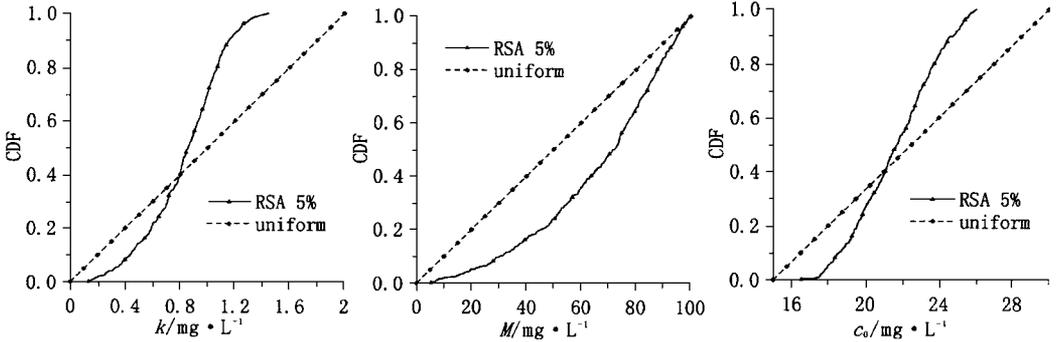


图 6 RSA方法的灵敏度分析

Fig. 6 Global sensitivity analysis by RSA

度与观测值与模拟值差值平方和的倒数成正比. 识别的似然度经过均一化处理, 即所有参数似然度之和等于 1.

图 7 是 GLUE 方法识别结果的似然度直方图. 可以看出,  $k$ 、 $M$  和  $c_0$  最大似然度的取值范围依次是:  $k = [0.8, 1]$ ,  $M = [80, 100]$ ,  $c_0 =$

[22.5, 23.5]. 与 RSA 所得出的结果非常的接近, 而  $k$  和  $M$  离真实值比较远, 造成这种偏离的原因也与 RSA 方法类似, 见图 8. 图 8 是 GLUE 方法似然度的分布, 为清晰起见, 在结果中只随机选取了 1000 个点. 图中气泡大小与对应该点的似然度成正比.

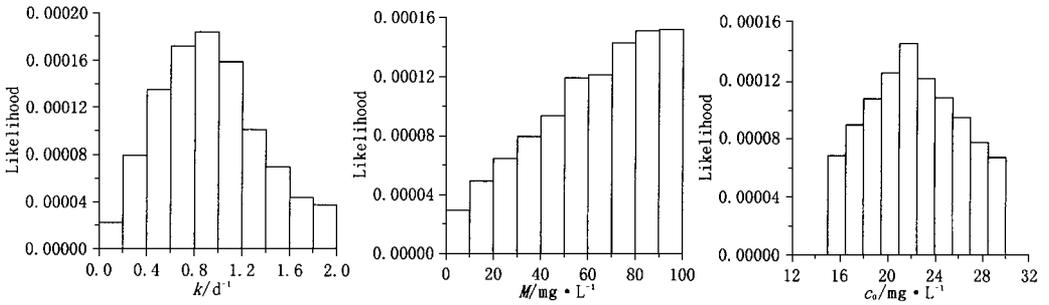


图 7 GLUE方法识别似然度的分布

Fig. 7 Marginal distribution of the likelihood by GLUE

采用 GLUE 方法进行预测, 各个参数预测结果的可信程度或者概率与似然度成正比, 由此可以求得预测结果的分布特征. 图 9 是当系统初值  $c_0 = 100 \text{ mg/L}$  时的模型验证. 这里只考虑 5% 和 95% 概率区间. 与 RSA 方法相比, 预测结果的范围略窄, 但其范围仍然能够覆盖实测值. 总的来说, 精度有所提高, 但是预测风险增大.

类似于 RSA 方法可以利用可接受参数的分布与原始分布的对比进行灵敏度分析, GLUE 方法也可利用累积似然度进行全局性灵敏度分析. 如果参数对目标函数没有显著影响,

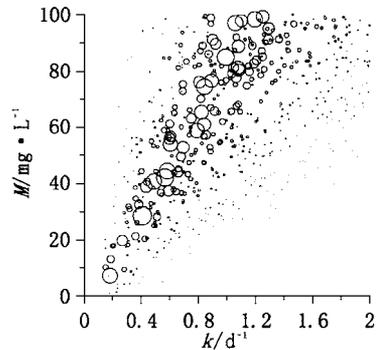


图 8 GLUE方法似然度气泡图

Fig. 8 Bubbles of the  $k$ - $M$  likelihood by GLUE

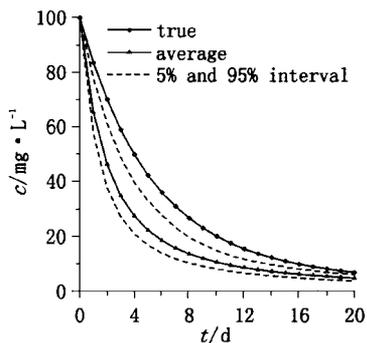


图 9 GLUE 方法的模型检验

Fig. 9 Validation of GLUE

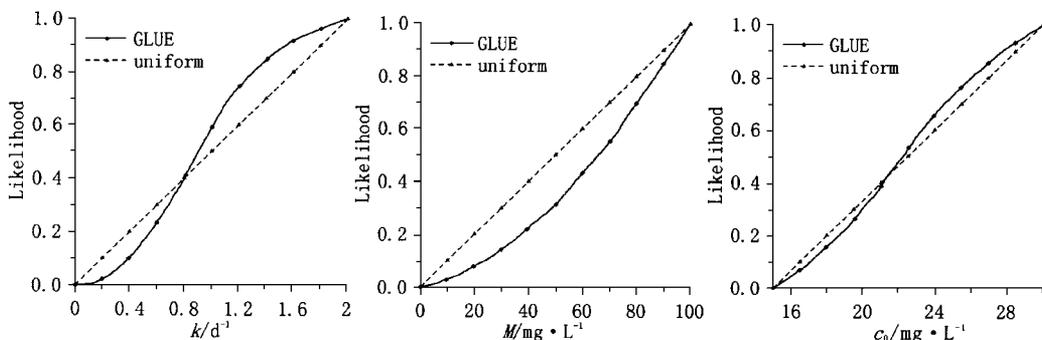


图 10 GLUE 方法的灵敏度分析

Fig. 10 Global sensitivity analysis by GLUE

来源的情况下,人们宁愿相信观测数据是系统真值的最好估计,使模拟值最逼近观测值的参数是最好的参数,这是优化方法的哲学基础。

RSA 方法将优化条件的放宽,改变了寻找最优单点参数的传统思维,从而避免了模型可识别性的问题。在某种意义上,RSA 方法是以牺牲精度换取可靠性,这使得 RSA 方法在降低精度的情况下,尽可能地估计到潜在的预测风险。RSA 方法体现了在目标函数有少量变化的情况下模型预测结果的变化,实际上这是对预测结果的稳定性分析。另外,RSA 方法还可进行全局灵敏度分析,以判断参数的相对重要性。

GLUE 方法对参数通过似然度进行区别。在理论上,GLUE 方法接受所有的参数,各个参数的可信程度由它所导致的模拟值与观测值的对比来确定,因此 GLUE 方法不是要接受或者拒绝某一参数值,而是确定各个参数值的可信

那么参数似然度的分布应接近于原始分布,即均匀分布;如果参数取值对似然度的影响较大,则参数累积似然度的分布与原始分布相差较大。图 10 中,参数似然度的累积分布与均匀分布的差别分别是: $k$  0.15,  $M$  0.19,  $c_0$  0.067,因此参数的重要性排序是: $M > k > c_0$ ,这与各个参数识别精度的排序一致。可见,GLUE 方法的全局灵敏度分析结果优于 RSA 方法。

## 6 分析与结论

优化方法形式简单,结果直观,容易在工程实践中采用。按照选定的准则,在没有其它信息

程度。另外,GLUE 方法也是一种有效的全局灵敏度分析方法,并且更少受到参数相关性的影响。在本例中,GLUE 方法显示了比 RSA 方法更强的全局灵敏度分析能力。

虽然 RSA 和 GLUE 方法各有特点,但都深深地打上了优化思想的烙印。RSA 方法所选择的可行参数是在最优附近的可行参数,最优值必然包含其中;而 GLUE 方法似然度最大、最可信的参数就是最优的参数。在这个意义上,RSA 和 GLUE 方法都不是对优化方法的否定,而是对优化方法的发展。

非线性环境模型的参数识别不是一个完全客观的过程,它受到人们主观判断的影响。优化方法的目标函数、RSA 可接受行为的准则、GLUE 方法似然度的确定等都没有唯一最佳的标准方法。虽然主观性在非线性环境模型的参数识别中是不可避免的,但只要这些主观设定

是合理的而不是武断的,对决策都是有益的,而这正是数学模型的意义所在。

参考文献:

- 1 邓义祥. 稀疏数据条件下河流水质模型的参数识别. 北京:清华大学博士论文, 2003. 1 ~ 39.
- 2 Chu W S. An evaluation of data requirements for groundwater contaminant transport modeling. *Wat. Res.*, 1987, **23** (3): 408 ~ 424.
- 3 Watermark Numerical Computing. PEST model-independent parameter estimation. *PEST Manual*, 2002. 1 ~ 27.
- 4 邢文训, 谢金星. 现代优化计算方法. 北京:清华大学出版社, 2001. 1 ~ 245.
- 5 刘毅, 陈吉宁, 杜鹏飞. 环境模型参数优化方法的比较研究. *环境科学*, 2002, **23**(2): 1 ~ 6.
- 6 Q Y Duan, V K Gupta, S Sorooshian. Shuffled complex evolution approach for effective and efficient global minimization. *Journal of optimization theory and applications*, 1993, **76**(3): 501 ~ 520.
- 7 Soroosh Sorooshian, Qingyun Duan, Vijai Kumar Gupta. Calibration of rainfall-runoff models: application of global optimization to the sacramento soil moisture accounting model. *Wat. Res.*, 1993, **29**(4): 1185 ~ 1197.
- 8 Qingyun Duan, Soroosh Sorooshian, Vijai Gupta. Effective and efficient global optimization for conceptual rainfall-runoff models. *Wat. Res.*, 1992, **23**(4): 1015 ~ 1031.
- 9 Qingyun Duan, Soroosh Sorooshian, Vijai K Gupta. Optimal use of the SCE-UA global optimization method for calibration watershed model. *Journal of Hydrology*, 1994, **158**(3 ~ 4): 265 ~ 284.
- 10 M B Beck. Water quality modeling: a review of the analysis of uncertainty. *Wat. Res. Res.*, 1987, **23**(3): 1393 ~ 1442.
- 11 邓义祥, 陈吉宁, 杜鹏飞. HSY 算法在水质模型参数识别中的应用. *上海环境科学*, 2002, **21**(8): 497 ~ 500.
- 12 Macros von Sperling. Calibration of poorly identifiable systems: application to activated sludge model. *J. of Envir. Eng.*, 1994, **120**(3): 625 ~ 644.
- 13 Keith Beven, Andrew Binley. The future of distributed models: model calibration and uncertainty prediction. *Hydrological Process*, 1992, **6**(3): 279 ~ 298.
- 14 Aronica G, Hankin B, Beven B. Uncertainty and equifinality in calibrating distributed roughness coefficients in flood propagation model with limited data. *Advances in Water Resources*, 1998, **22**(4): 349 ~ 365.
- 15 Ratto M, Tarantola S, Saltelli A. Sensitivity analysis in model calibration: GSA-GLUE approach. *Computer Physics Communication*, 2001, **136**(3): 212 ~ 224.
- 16 Jining Chen. Modeling and control of the activated sludge process: towards a systematic Framework. PhD Thesis, the University of London, 1993. 130 ~ 165.
- 17 徐士良. Fortran 常用算法程序集. 北京:清华大学出版社, 1995. 123 ~ 125.