

环境模型参数识别与不确定性分析

刘毅, 陈吉宁, 杜鹏飞(清华大学环境科学与工程系环境模拟与污染控制国家重点联合实验室, 北京 100084, E-mail: jchenl@mail.tsinghua.edu.cn)

摘要:在对水文模型实例的参数不确定性分析基础上, 分别采用传统灵敏度分析方法、HSY 算法、线性回归等方法对模型参数特性进行了识别与比较研究。结果表明参数优化算法与传统灵敏度分析方法不能解释模型结构复杂性特征, 采用不确定性分析方法对环境模型参数进行识别提供了深入分析与理解模型系统的有效途径。

关键词:参数优化; 不确定性; 模型结构; 参数识别; 灵敏度; HSY 算法

中图分类号: X11 文献标识码: A 文章编号: 0250-3301(2002)06-05-0006

Parameters Identification and Uncertainty Analysis for Environmental Model

Liu Yi, Chen Jining, Du Pengfei(Environmental Simulation and Pollution Control State Key Joint Laboratory, Department of Environmental Science and Engineering, Tsinghua University, Beijing 100084, China E-mail: Jchenl@mail.tsinghua.edu.cn)

Abstract: This paper examined a case study of hydrological model for identifying parameter uncertainty by using three sensitivity analysis methods: HSY algorithm, linear regression method and coupling analysis method. The results showed that optimal algorithms cannot give a sound explanation for complexity of model structure and identifying model parameters via uncertainty analysis methods presented an effective alternative to understand model system.

Key words: parameter optimization; uncertainty; model structure; parameter identification; sensitivity; HSY algorithm

随着环境模型结构复杂性的急剧增长, 模型参数在高维空间表现出了复杂的相关性结构并直接导致了优化后验参数的识别问题^[1, 2]。因此仅仅局限于参数优化算法效率和精度等方面的研究已经不能满足理论与实践的需要。相对于观测数据和模型参数而言, 基于现有科学认知体系构建的模型结构是建模过程中不确定性的根本来源^[3]。然而由于缺乏深入研究结构不确定性的理论基础和有效技术手段, 模型结构往往只能通过识别参数后验统计分布规律间接地得到验证。

参数不确定性依赖于模型结构, 并直接导致了灵敏度问题^[4]。模型参数不确定性包括参数可识别性和参数灵敏度 2 个基本方面。参数不确定性分析提供了后验地识别模型结构的可行途径^[5, 6]。本文以一个经典的水文箱式模型为实例, 从优化参数的不确定性和参数识别问题出发, 对复杂模型的参数灵敏度分析方法进

行了比较研究。

1 环境模型参数不确定性与识别

由于参数不确定性普遍存在, 根据经验估计或者观测值优化得到的参数并不能保证模型应用的精度和预测结果的可靠性。在追求更加高效和稳定优化算法的同时, 所得到优化参数的后验分布具有本质上的不确定性; 对于复杂模型来说, 甚至无法判断优化结果是否达到了全局最优, 也无法预测“最优”参数对于模型预测的影响^[7]。引入参数不确定性分析可以更为深刻地理解和认识到现实世界和模型系统特征之间的本质区别。

直接参数优化算法可以理解为在一定控制性准则约束下的空间随机搜索过程。算法根据新点产生器和判定准则在高维参数空间持续地

基金项目: 高等学校优秀青年教师教学科研奖励计划资助项目
作者简介: 刘毅(1975~), 男, 博士研究生, 主要从事环境系统分析方向的研究。

收稿日期: 2001-10-12; 修订日期: 2002-05-08

进行“产生新点-判断-接受/舍弃”的迭代过程,直至满足终止准则.为了说明参数不确定性及其识别问题,本文采用 4 种直接参数优化算法进行比较分析^[8].控制随机搜索算法(CRS)引入了几何学中“重心”的概念,即考虑了新点产生的随机性,又在一定程度上保证了搜索的整体性.复合形混合演化算法(SCE-UA)是将生物自然演化过程引入到数值计算中,模拟了生物进化的过程,提高了计算效率和全局搜索整体最优的能力.模拟退火算法(SA)则假设优化问题的解及其目标函数分别与固体物质的微观状态及其能量所对应,采用随机方法模拟固体稳定“退火”的过程.退火单纯形算法(AS)综合了下山单纯形方法和模拟退火法 2 种优化算法,更加充分地利用了单纯形的形变信息,从而提高了计算效率和算法稳定性.

随着环境模型的不断开发和广泛应用,环境模型的种类和数量日益丰富,模型本身所表现出的结构特征也日趋复杂.本研究仅以一个两箱式水文模型为例^[8],表 1 给出了模型中 11 个参数的先验取值范围及其物理意义.

表 1 水文模型参数取值及其物理意义

Table 1 Estimated values of parameters and theirs physical meaning for a case of hydrological model			
模型参数	最大取值	最小取值	物理意义
k_1	0.05	0.3	不同箱体出口处的径流流速/ h^{-1}
k_2	0.008	0.1	
k_3	0.01	0.1	
k_4	0.005	0.1	
k_5	0.0005	0.002	
h_1	5.0	20.0	不同箱体出口处的高度/ mm
h_2	1.0	10.0	
h_3	700.0	1500.0	
h_4	500	1000.0	
H_1	0.1	20.0	上层和下层水箱高度/ mm
H_2	5.0	200.0	

优化算法的基本思想认为在特定模型结构下只有唯一一组最佳参数与之对应.然而由于模型结构复杂性与输入数据的不确定性,优化算法通常不能寻优到环境模型的唯一真值,并且也无法判断算法是否达到全局最优.这是导致优化参数不确定性的根本原因.根据 4 种直

接优化算法得到的优化参数结果可知,在目标函数值无显著性差异的条件下($\max \delta < 1.8\%$),“最佳”估计参数之间具有较大差异(见表 2).由于本文采用的水文模型输入数据序列非常完整,可以忽略其不确定性,并且几种优化算法也是当前最为稳定和可靠的全局搜索算法^[8],因此上述优化结果直接证实了优化算法不能为深入研究复杂环境模型提供有效途径.

进一步研究表明,不同算法寻优进程中的参数收敛轨迹也具有较大差异^[8].以 CRS 算法运用于水文模型为例,图 1 是 500 个最佳优化参数样本点分布图,其中菱形符号代表最佳的 10 个参数值分布(采样总数 $n = 20000$),虚线对应于最佳目标函数值.

可以由图直观地看到最佳 10 组参数对应的目标函数值没有显著区别,但这 10 组参数值并没有非常一致地收敛到“最佳”参数估计值.

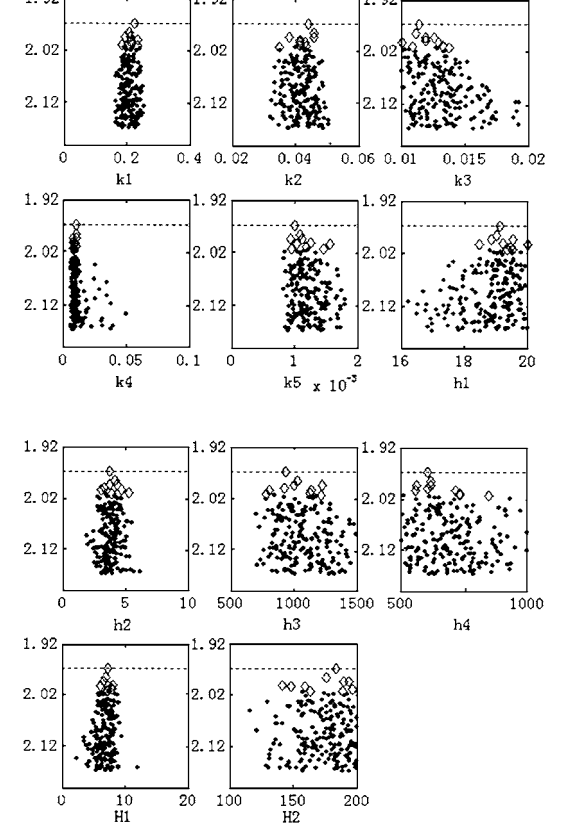


图 1 CRS 算法模拟的参数分布

Fig.1 Distribution of optimized parameters resulted from CRS algorithm

这种优化结果通常称为优化参数的等效性或可置换性^[7, 9]. 优化参数可置换性的产生是由于模型参数在高维空间具有的复杂相关性, 是模型结构复杂性和参数不确定性的集中体现. 这种与优化算法结构设计、新点产生器、空间搜索方法和接受/舍弃判定准则等高度相关的后验参数空间分布上的差异性, 是导致优化算法不能解释模型结构及其所产生的参数不确定性的主要原因.

2 环境模型的参数灵敏度

如前所述, 优化方法不能解释模型结构复杂性与参数不确定性, 由此产生了对于模型参

数的识别问题. 参数识别很重要的一个方面就是研究参数变化所引起的模型响应, 即参数灵敏度问题^[5, 9]. 研究参数灵敏度有助于深入理解并改进模型结构的稳定性.

传统参数灵敏度分析方法是在某个参数最佳估计值附近给定一个人工干扰, 并计算参数在这一很小范围内产生波动所导致模型输出的变化率, 即扰动分析方法. 从表 2 的计算结果中可以得出基于 4 种直接优化算法最佳参数估计值的灵敏度数值($\delta = \pm 10\%$) 具有较大差异, 其结果强烈依赖于优化算法的选择.

如前所述, 由于优化算法自身结构设计决

表 2 水文模型参数优化与局部灵敏度计算 ($\delta = \pm 10\%$)

Table 2 Local sensitivities related to the optimized values of parameters for the case of a hydrologic model

参数	CRS 算法		SCE ₂ UA 算法		SA 算法		AS 算法	
	最优值	灵敏度(+ / -)	最优值	灵敏度(+ / -)	最优值	灵敏度(+ / -)	最优值	灵敏度(+ / -)
k_1 (1/h)	0.2236	0.6044/0.0655	0.2150	0.3008/ - 0.2410	0.2151	0.3077/ - 0.2526	0.2136	0.2834/ - 0.2778
k_2 (1/h)	0.0439	0.5942/ - 0.5331	0.0427	0.5334/ - 0.4966	0.0431	0.7367/ - 0.4325	0.0434	0.7366/ - 0.4105
k_3 (1/h)	0.0114	0.1626/ - 0.2965	0.0101	0.2146/ - 0.1686	0.0100	0.1930/ - 0.1634	0.0100	0.2212/ - 0.1533
k_4 (1/h)	0.0100	0/0	0.0613	0/0	0.0494	0/0	0.0988	0/0
k_5 (1/h)	0.0010	0.2431/ - 0.2028	0.0010	0.1194/ - 0.2554	0.0010	0.1774/ - 0.1526	0.0008	0.1698/ - 0.1828
h_1 (mm)	19.1208	3.6064/ - 4.2128	19.4627	4.0179/ - 3.4003	18.6718	3.9311/ - 3.0903	18.8697	4.0154/ - 3.1122
h_2 (mm)	3.7686	0.2250/ - 0.0297	3.5547	0.1567/ - 0.0542	2.5141	0.0528/ - 0.0683	2.8560	0.0674/ - 0.0919
h_3 (mm)	930.9930	0/0	927.7306	0/0	1136.6304	0/0	1330.9256	0/0
h_4 (mm)	603.9370	0.0021/0.0022	545.0963	- 0.0012/ - 0.0020	977.1397	- 0.0001/ - 0.0003	680.0279	- 0.0007/ - 0.0013
H_1 (mm)	7.2466	0.0242/ - 0.0645	7.6458	0.0082/ - 0.0875	7.0637	0.0433/ - 0.0354	7.3351	0.0461/ - 0.0405
H_2 (mm)	183.6478	0.2388/ - 0.2005	171.9362	0.1152/ - 0.2496	159.3208	0.1774/ - 0.1526	199.7192	0.1657/ - 0.1773
目标函数值	1.9661		1.9377		1.9341		1.9332	

定了最佳参数估计值并得到具有特定收敛特征的参数样本, 因此基于“最佳”估计参数值的灵敏度分析不能完整地描述模型参数的空间分布形态; 另一方面更为本质的是, 由模型结构复杂性导致的参数相关性要求在灵敏度分析过程中必须考虑参数之间的相互影响, 而不是仅靠变动某一个参数得到模型响应. 考虑到模型参数之间的高度相关性, 现代环境系统研究在不确定性分析思想框架下提出了更为有效的参数灵敏度分析方法, 即区域灵敏度分析方法(Regional Sensitivity)^[1, 3]. 与传统分析方法不同, 区域灵敏度分析抛弃了“寻优”思想, 承认参数空间分布的复杂性与相关性, 是一个对模型结构在一定准则下对随机参数大样本发生响应的统计分析过程. 本文采用线性回归和 HSY 2 种区域灵敏度方法.

线性回归方法是通过随机采样产生参数样

本序列, 计算每个样本对应的模型响应, 然后进行线性回归(LR), 建立如下形式的线性模型^[6]:

$$y = \beta_0 + \sum_{i=1}^n \beta_i p_i \tag{1}$$

(1) 式中, β_i 为对应于参数 p_i 的线性回归模型系数, 表征了参数 p_i 对模型响应 y 的贡献率(权重), 即为基于线性回归方法得到的参数 p_i 的绝对灵敏度. 由于模型参数量纲不同, 通常采用(2)式计算参数相对灵敏度 $\beta_i^{(s)}$, 其中 σ_{p_i} 和 σ_y 分别表示参数样本和模型输出的方差:

$$\beta_i^{(s)} = \beta_i \frac{\sigma_{p_i}}{\sigma_y} \tag{2}$$

HSY 算法用于区域灵敏度分析是基于模型某种形式的判定准则, 通过随机采样过程产生 2 组参数, 并根据统计检验方法计算参数灵敏度^[3]. 这里的某一准则是根据已有经验、数据或是在一定置信区间下的模拟准则, 用来判断随机

采样的参数样本是否被接受,体现了不确定性分析中不承认“最优”的基本思想.可接受与不可接受 2 组参数样本共同表征了模型本质特征^[1, 5].本文中采用(3)式定义可接受的系统行为:

$$Y_{\text{obv}}[1 - \alpha\beta(n)] \leq Y_{\text{sim}} \leq Y_{\text{obv}}[1 + \alpha\beta(n)]$$

(3)

式中, n 是模型优化结果与观测值之间的误差向量, $\delta(n)$ 是 n 维向量的样本方差, α 表示可接受系统偏差的倍数. Y_{obv} 和 Y_{sim} 分别表示观测值和模型输出值.这样就形成了一个系统输出的控制“廊道”,凡是超出这个廊道边界的样本,都认为是不可接受的采样(见图 2).

两样本非参数假设检验的方法有 χ^2 -拟合

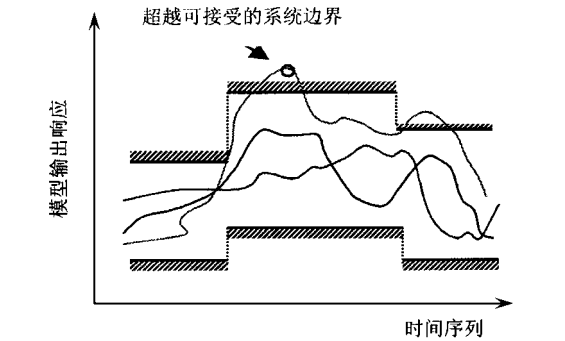


图 2 HSY 算法示意图^[3]

Fig. 2 Demonstration of HSY algorithm^[3]

优度检验, K-S 检验, 秩检验, 游程检验等, 不同方法对灵敏度计算结果会产生一定影响^[10]. 本文采用 K-S 检验方法^[1, 5, 10]. K-S 检验数值越大, 说明可接受参数的概率分布与不可接受参数的概率分布之间的差别越大, 从而参数灵敏性越大, 具有明确物理意义的参数也越容易被识别. 当涉及到大样本采样过程时, 可以采用下面的准则来判断参数灵敏度分析算法是否达到收敛^[1]: ①重复采样过程, 样本均值和方差不发生变化; ②重复采用过程, 模型参数的灵敏度排序不发生变化.

表 3 给出了采用 LR 与 HSY 算法得到水文模型实例中的 11 个参数的区域灵敏度数值及排序, 作为比较同时列出了参数的局部灵敏度排序. 由于 HSY 算法考虑了模型结构导致的参数相关性, 并且采用完全随机采样方法, 因此其区域灵敏度分析结果具有可靠性. 而 LR 算法是在整个参数可能空间上采样, 参数灵敏度中不加区分地包含了可接受与不可接受参数的特征信息; 另外, 尽管算法也采用了在整个参数空间上的随机过程, 但是从建立参数与模型响应之间的线性模型中得到的参数灵敏度又不可避免地抵消了模型结构自身复杂性的影响, 因此其灵敏度结果是不可靠的.

表 3 水文模型实例的区域参数灵敏度分析结果 (d/d^* , $\alpha=8.0$, $\xi=3.0$)¹⁾

模型参数	区域灵敏度				局部灵敏度							
	LR		HSY		CRS		SCE-UA		SA		AS	
	$\beta^{(s)}$	排序	d/d^*	排序	排序+	排序-	排序+	排序-	排序+	排序-	排序+	排序-
k_1 (1/hr)	0.0390	7	6.1755	6	2	6	3	5	3	3	3	3
k_2 (1/hr)	-0.2461	5	7.0886	5	3	2	2	2	2	2	2	2
k_3 (1/hr)	0.5802	1	8.3860	4	7	3	4	6	4	4	4	6
k_4 (1/hr)	-0.0076	9	0.5990	10	10	10	10	10	10	10	10	10
k_5 (1/hr)	-0.2901	3	3.8152	7	4	4	6	3	6	6	5	4
h_1 (mm)	0.0612	6	11.658	1	1	1	1	1	1	1	1	1
h_2 (mm)	0.2471	4	2.2472	8	6	8	5	8	7	7	7	7
h_3 (mm)	0.0070	11	0.6799	9	10	10	10	10	10	10	10	10
h_4 (mm)	-0.0073	10	0.3445	11	9	9	9	9	9	9	9	9
H_1 (mm)	0.0164	8	9.3208	3	8	7	8	7	8	8	8	8
H_2 (mm)	-0.3348	2	10.831	2	5	5	7	4	5	5	6	5

1) d 为可接受参数概率分布与不可接受参数概率分布的最大距离, d^* 为 90 %置信区间的 K-S 检验距离.

尽管局部灵敏度数值之间存在较大差异, 但某些参数的局部灵敏度排序却出现一定程度上的一致性, 例如局部最灵敏参数均为 h_1 . 如前所述, 由于仅考虑了单一参数在“最优点”附近沿特定方向变化所引起的模型输出响应, 参数局部灵敏度不能反映出模型结构特征与参数相关性. 而 HSY 算法提出将参数划分为系统可接受与不可接受样本的思想, 从根本上避免了

传统“寻优”思想带来的模型结构和参数不确定性分析中的困惑,并通过计算 2 参数样本之间的统计距离,得到了包含参数空间分布特征整体的区域灵敏度,为优化模型结构和改进模型稳定性提供了有力的分析基础和技术支持.

以部分参数为例,图 3 进一步给出了 HSY 算法可接受参数的最佳 500 个估计值与 4 种优化算法的最佳估计参数的空间分布比较.取代唯一“最优”的参数估计值,HSY 算法以降低计算效率为代价保证了参数空间搜索的整体性,避免了优化算法定向搜索对参数特性的片面理解;并以降低计算精度为代价得到的系统可接受的参数空间,对应着可接受的模型输出,其意义在于定量地给出参数的可识别性和模型预测结果的可靠性.特别是在模型应用于情景分析中,传统方法只能根据参数局部灵敏度绝对数值大小来计算未来各种可能情景下的模型响应,由于实际环境系统复杂性和参数之间的实际相关性,优化算法并不能保证模型预测的可靠性.而采用区域灵敏度方法得到的可接受参数样本,定量地给出了可接受模型输出的参数空间分布,从而保证了模型整体上的预测可靠性.正如有学者指出,与其面对没有可靠性的优化算法,不如接受一个更为合理的不确定性^[3].

进一步地,采用 K-S 统计检验方法所得到的区域灵敏度数值表征了参数可识别性,其为今后模型结构不确定性研究以及模型结构的简化和改进提供了科学分析依据^[5].

图 4 是基于各种区域灵敏度分析方法的参数灵敏度排序分布图.基于前文分析,认为只有 HSY 算法得到的参数灵敏度是无偏的.无论是参数优化还是参数灵敏度分析,优化算法都不能很好地解释环境模型复杂性和参数识别问题.基于不确定性分析思想的 HSY 算法提供了识别复杂模型参数和深入理解模型结构的有效方法.

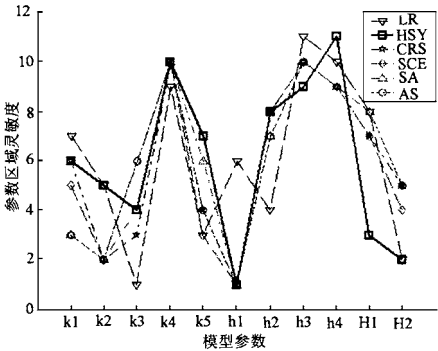


图 4 水文模型参数灵敏度排序结果

Fig.4 Distribution of the rankings of parameter sensitivities for the case of a hydrological model

参考文献:

- 1 Chen J. The Modeling and Control of activated sludge process: Towards a systematic framework. Ph.D. thesis. Imperial College, London, Engl., U.K. 1993. 54 ~ 94.
- 2 Chen J, Wheeler H S. Identification and uncertainty analysis of soil water retention models using lysimeter data. Water Resources Research, 1999, 35(8): 2401 ~ 2414.
- 3 Beck M B. Water quality modeling: A review of the analysis of uncertainty. Water Resources Research, 1987, 23(8): 1393 ~ 1442.
- 4 Spear R C, Grieb T M, Shang N. Parameter uncertainty and interaction in complex environmental models. Water Resources Research, 1994, 30(11): 3159 ~ 3169.
- 5 Chen J, Beck M B. Quality Assurance of Multi-Media Model for Predictive Screening Tasks. USEPA, Report. 1998.
- 6 Klepper O. Multivariate aspects of model uncertainty analysis: Subjective calibration of incorrect models. Report. National Institute of Public Health and Environmental Protection, The Netherlands. 1999.
- 7 Aalderink R H, Jovin J. Identification of the parameters describing primary production from continuous oxygen signals. Water Science and Technology, 1997, 36(5): 43 ~ 51.
- 8 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.
- 9 Beven K. Prophecy, reality and uncertainty in distributed hydrological modeling. Advances in Water Resources, 1993, 16: 41 ~ 51.
- 10 Reckhow K H, Clements J T, Dodd R C. Statistical evaluation of mechanistic water-quality models. Journal of Environmental Engineering, 1990, 116(2): 250 ~ 268.

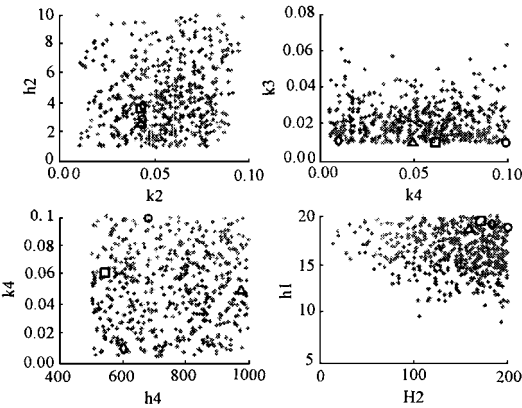


图 3 部分参数的二维空间相关性

(菱形、正方形、三角形与圆形分别表示 CRS、SCE、UA、SA 和 AS 算法的最佳参数估计值,空间点为 HSY 算法可接受参数的最佳 500 个估计值)

Fig.3 Parameters correlations under optimal algorithms and HSY algorithm