

多目标遗传单纯形算法在概念性水文模型 参数优化中的应用

陈炯烽^{1,3}, 张万昌^{1,2}, 吴波^{1,3}

(1. 南京大学 国际地球系统科学研究所, 江苏 南京 210093;

2. 中国科学院 大气物理所 东亚区域气候—环境重点实验室 全球变化

东亚区域研究中心, 北京 100029; 3. 南京大学 地理与海洋科学学院, 江苏 南京 210093)

摘要: 参数识别是水文模型应用的前提, 在参数识别的同时需要充分考虑模型自身及参数的不确定性。引入了一种具有较强全局与局部搜索能力的多目标优化算法(MOCOM-UA), 探讨了该算法在多目标基础上融合遗传算法和单纯形法两类算法不同搜索机制的模型优化方案, 并以汉江上游江口流域新安江模型降水—径流模拟实践为例, 将 MOCOM-UA 算法应用于新安江模型的参数识别, 得到了 4 目标函数情形下的 Pareto 参数空间和模型的预测范围, 并根据该算法模拟计算的结果, 初步分析了参数和模型的不确定性。

关键词: 参数识别; 水文模型; 多目标; 全局优化; 遗传算法; 单纯形法; 新安江模型

文献标识码: B

文章编号: 1000—288X(2008)03—0107—06

中图分类号: P334+.92

Multi-objective Calibration with Predictive Uncertainty Analysis for Conceptual Hydrological Models

CHEN Jiong-feng^{1,3}, ZHANG Wan-chang^{1,2}, WU Bo^{1,3}

(1. International Institute for Earth System Science, Nanjing University, Nanjing, Jiangsu 210093, China;

2. START Regional Center for Temperate East Asia, Institute of Atmospheric Physics, CAS,

Beijing 100029, China; 3. School of Geography and Ocean Science, Nanjing University, Nanjing, Jiangsu 210093, China)

Abstract: Parameter determination is the prerequisite for the simulation of rainfall and runoff in a hydrological model. Model and parameter uncertainty should be considered during that procedure. This paper started with an introduction of the theory on the multi-objective complex evolution (MOCOM-UA) algorithm of excellent global and local optimization capability for dealing with the complex problem of hydrological model calibration. Based on the Xin'anjiang model, different searching mechanisms between the multi-objective approach based method and genetic and simplex algorithm were respectively discussed by rainfall-runoff simulations for the Baohe watershed in the upper reaches of the Hanjiang River basin. With the application of the multi-objective genetic algorithm coupled with the simplex method as an optimal solution, the Pareto space of parameters and the prediction extent of the model were calculated and analyzed under four objective function conditions. The uncertainties of the model and its parameters were preliminarily investigated.

Keywords: parameter determination; rainfall-runoff model; multi-objective; global optimization; genetic algorithm; simplex method; Xin'anjiang Model

随着计算技术的发展, 基于优化算法的计算机自动参数优选成为该领域的研究主流^[1]。传统的优化方法, 如 Rosenbroke 法和梯度法容易受初值影响而局部收敛, 往往无法提供全局最优解, 缺乏足够的稳

定性, 使得这些传统的自动优选方法无法真正地在实际应用中推广^[2]。近年来, 遗传算法 (genetic algorithm) 由于具有全局搜索特性、思想简单、易于实现等优点在概念性水文模型参数优选研究领域得到广

收稿日期: 2007-04-15 修回日期: 2007-12-28

资助项目: 国家重点基础研究发展规划项目 (2006CB400502; 2001CB309404); 中国科学院“百人计划”择优支持项目 (8-057493); 中国科学院大气物理研究所东亚区域气候—环境重点实验室开放基金

作者简介: 陈炯烽 (1982—), 男 (汉族), 福建省龙岩市人, 博士研究生, 主要从事遥感和 GIS 在水文学中的应用方面的研究。E-mail: woyuna@gmail.com。

泛应用^[3-7]。遗传算法为启发式搜索算法,全局搜索能力较强,但其局部搜索能力较弱,导致优化解质量不高。在实际应用中往往出现“过早”收敛或收敛缓慢等缺点。

Duan 等^[8]基于算法混合的思路,在进化算法中加入局部单纯形搜索算子开发了 SCE—UA 优化算法,成功应用于水文模型参数优选。陆桂华等^[9]通过在实编码遗传算法中加入单纯形搜索算子及加速搜索算子,逐步调整参数变化区间的方法,建立了单纯形混合加速遗传算法,也取得了较好的效果。单纯形法(simplex method, SM)^[10],也称可变多面体搜索法,是确定性下降方法,其局部搜索能力很强。将搜索机制上存在如此差异的两种算法进行融合,有利于增强全局和局部意义下的搜索能力和效率^[11]。

同时,为优化算法选择恰当的误差统计判别函数是参数识别的重要环节。由于传统的概念性水文模型只有径流量输出,参数识别优化也通常采用单一的目标函数,如模拟流量与实测值的均方差。但一个统计目标往往无法描述水文过程(如流量过程线)所有特征,例如洪水预报中除了要求准确模拟径流量以外,还要求准确模拟峰现、滞时和洪峰流量^[12]。而在实际应用中,单一的目标优选往往只能使该目标函数达到最优,而其它的结果却是不理想的^[13]。因此进行模型参数识别时,运用多目标方法约束模型的模拟行为是必要的^[14-15]。20 多年来参数识别领域研究的另一个重要方面就是模型和模型参数的不确定性及其对模型输出的影响^[16]。在人为设参优选实验中,通过优化算法能获得预先设定的参数结果。这也说明模型不确定性与数据误差是造成实际应用中无法获得最优参数的根本原因。据此 Beven^[17]认为应该用等价的观点认同一个可接受参数集合,这个集合中每一参数对模型来说效果是同样好的,因此都是可以接受的参数。而探讨可接受参数集,通过多目标优化求得 Pareto 解是可行途径。在模型参数识别中,多目标函数的间断性及多峰性的复杂问题需要使用强大的优化算子解决,而模型自身结构和数据误差导致模型预测和参数的不确定性问题需要运用多目标方法分析。因此将较好的优化算法如遗传单纯形算法和多目标方法结合起来使用是必须的。Yapo 等^[18]在 SCE—UA 算法基础上提出了多目标计算框架(multi-objective complex evolution, 简称 MOCOM—UA),并将其尝试应用于 Sacramento 模型的参数识别和预测不确定性分析。MOCOM—UA 是遗传单纯形算法和多目标结合方式之一,理论上可以用于复杂模型的参数寻优和预测不确定性分析,实践上仍需要进一步的验证。目前 MOCOM—UA 方法

用于水文模型参数识别的国内外文献还很少,模型参数识别研究的国内文献多集中于利用优化算法获得最优参数^[19]。鉴于此,本文引入多目标遗传单纯形算法 MOCOM—UA,以新安江模型为例,进一步讨论其应用于模型参数识别的可行性,利用其强大的全局优化的特性,求得了 Pareto 优解,并初步分析了模型预测和参数的不确定性。

1 资料与方法

1.1 研究方法

多目标参数优选参数识别可转化为如下的最优化问题

$$\text{Min} \{ F_1(\cdot), F_2(\cdot), \dots, F_m(\cdot) \}, \quad (1)$$

式中: \cdot ——模型参数; $F_i(\cdot), i=1, 2, \dots, m$ 表示 m 个目标函数; \cdot ——参数空间。

多目标优化问题一个重要特点就是往往不存在唯一的最优解。其中任何一个目标函数的减小可能以另一目标函数的增大为代价^[20]。由于这些目标函数相互独立,不能在它们之间进行比较,所以不能确定哪一个为最好的解。这样的解集被称为 Pareto 最优集(Pareto set)。

数学上,定义 Pareto 最优集为不劣于参数空间中任一参数的解的集合。参数 A 劣于(dominated by)参数解 B 当且仅当满足不存在 i ,使

$$F_i(B) < F_i(A), \quad \forall i=1, \dots, m \quad (2)$$

式中, $<$ 表示左式目标函数表现值差于右式。根据此定义, Pareto 集中的元素有如下性质: Pareto 集中任一参数解必然优于 Pareto 集之外的解; Pareto 集中的任意两个参数解无法比较其优劣。

1.1.1 优选参数和目标函数 本例中共选取 4 个敏感参数 K, S_M, K_G, C_G 和深层蒸散发系数 C 作为优选对象。由于模型中 $K_G + K_I = 0.7$ 的隐性关系,优选 K_G 可同时得到 K_I 。汇流参数 K_e, X_e 。根据流域的水力学特性求得^[21], U_H 由单独的一次或者多次降雨产生的日尺度的洪水过程流量数据求得^[21]。 $W_M, W_{UM}, W_{LM}, I_{MP}, B, E_x$ 参数不敏感对模型模拟影响不大^[22], 其值直接取范围的中间值。

在模型不确定性和数据误差存在的情况下,为了更好地约束模型的模拟行为,分析不确定性带来的参数和预测的不确定性,建立合适的评价标准即目标函数体系是多目标优选的关键环节。

文献中常见的基本目标函数是用于检验实测与模拟过程线吻合程度的均方误差 RMSE(root-mean-square error)。无论是传统的单目标函数参数优选还是多目标函数优选, RMSE 是一个出现频率较高的评价标准。Gupta 等^[23]对美国国家气象部门(Na-

tional Weather Service) 在 SAC—SMA 模型参数率定中常用的 9 个目标函数进行了独立性分析,发现均方误差(DRMS),日均水量平衡误差(BIAS),误差符号改变数 NSC(number of sign changes)之间的相关系数最小,建议将这 3 个目标函数作为模型优选的主要目标函数。Madsen^[24]通过对 MIKE11/NAM 模型多目标函数参数优选研究,认为总水量平衡误差(overall volume error),均方误差(RMSE),高水均方误差(average RMSE of peak flow event),低水均方误差(average RMSE of low flow event) 4 个目标函数能够对模拟径流曲线的不同特征进行描述,并利用优选参数取得了较好的模拟结果。综合考虑两篇文献所使用的目标函数,本例采用了如下 4 个目标函数作为新安江模型日模拟参数优选的评价标准。同以上文献有所不同的是,为了计算方便,评价高水和低水模拟情况的目标函数采用赵人俊^[22]先生提出的两个评价函数。

(1) 水量平衡模拟用总水量平衡误差(overall volume error)表示

$$F_1(\theta) = \frac{\sum_{i=1}^n [M_i - Q_i(\theta)]}{\sum_{i=1}^n M_i} \quad (3)$$

式中: M_i ——实测流量; $Q_i(\theta)$ ——在参数 θ 下模型的计算流量; n ——模拟时段长。

(2) 实测与模拟流量过程线的吻合程度用均方误差(RMSE)表示

$$F_2(\theta) = \left\{ \frac{1}{n} \sum_{i=1}^n [M_i - Q_i(\theta)]^2 \right\}^{1/2} \quad (4)$$

式中,符号意义同(3)式。

(3) F_3 的取值是低水大,高水小,主要反映低水部分的误差

$$F_3(\theta) = \frac{\sum_{i=1}^n \text{ABS}\{\lg[M_i / Q_i(\theta)]\}}{\sum_{i=1}^n \lg(M_i)} \quad (5)$$

式中,符号意义同(3)式。

(4) F_4 绝对值误差是高水大,低水小,主要反映高水部分的误差

$$F_4(\theta) = \frac{\sum_{i=1}^n \text{ABS}\{[M_i - Q_i(\theta)]\}}{\sum_{i=1}^n M_i} \quad (6)$$

式中,符号意义同(3)式。

1.1.2 多目标遗传单纯形算法 遗传算法是求解多目标问题的有效手段,将其引入到多目标优化中,可在一次优化过程中产生一组 Pareto 解^[20]。但遗传算法局部搜索能力较弱,在实际应用中往往出现“早熟”收敛或收敛缓慢等问题,优化解的质量不高。遗传单纯形算法以遗传算法的基本框架为基础,由单纯形法替代标准遗传算法的交叉操作和变异操作,解决了这一问题。遗传单纯形混合算法不仅具有遗传算法通

用、简单、全局随机搜索的优点,而且融入了局部搜索法的快速寻优并收敛的优点^[11]。Duan 等^[8]指出由启发式搜索算法和局部优化法组成的混合算法是求解非线性复杂模型的最佳途径之一。鉴于水文模型自身复杂的不确定性,Yapo^[18]提出了可用于模型的参数识别和预测不确定性分析的遗传单纯形算法的多目标计算框架(MOCOM—UA)。有关多目标遗传单纯形法的算法策略详见文献^[18],MOCOM—UA 流程如图 1 所示。

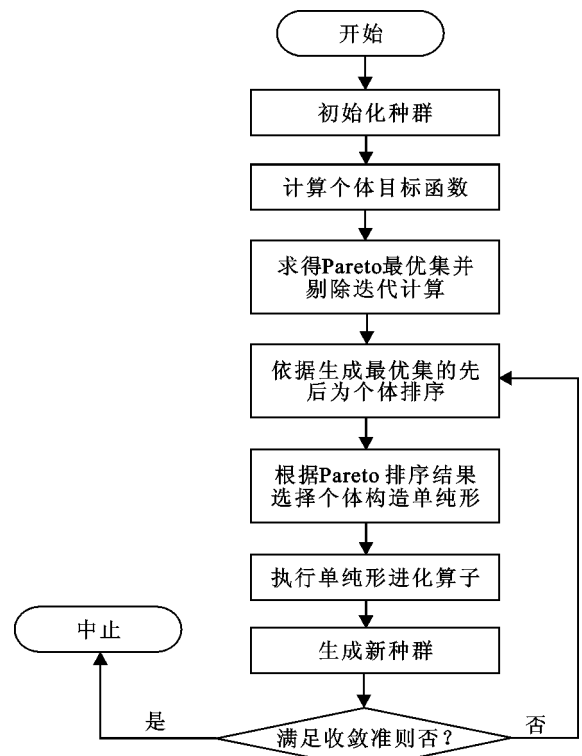


图 1 多目标遗传单纯形法策略流程

1.2 研究区域与资料

汉江是长江的第一大支流,地跨陕西、湖北、重庆等省市。本文的研究区为江口流域,位于汉江左岸支流褒河流域的上游,由江口水文站(107°01'E,33°42'N)控制,其集水面积为 2 341.6 km²,海拔高度变化在 950~3 400 m,年降水量变化在 600~900 mm 间,主要集中在夏季的 6—9 月。

研究区内土壤类型以黄棕壤和棕壤为主,主要有石灰岩黄棕壤,坡积洪积黄棕壤,花岗片麻岩黄褐土,坡积洪积黄褐土,以及石灰岩黄棕壤等。

该流域植被覆盖度高,与许多自然保护区毗邻,自然植被保护良好,以温带落叶阔叶林为主,如栓皮栎林,锐齿栎林,红桦林和杨、柳、槭等杂木林,以及一部分温带落叶灌丛,如胡枝子灌丛,白刺花灌丛,绣线菊灌丛等,还有少部分的温带和亚热带针叶植物,如油松林,华山松林等,另外还包括一些温带草丛和草甸等。

研究区江口地区有雨量站 10 个,蒸发站 1 个,本文选取了 1981—1983 年各测站的日降水与日蒸发资料以及江口站 1981—1983 年日流量数据进行模拟与检验。

2 结果与分析

2.1 目标函数空间

图 2(a—b) 给出了最终 Pareto 优集中任选的 24 组参数的目标函数空间分布。总体上看,参数在 F_1 与 F_2 的值域中分布相对均匀,无明显的相关。而图 2b 中主要反映高水误差的 F_4 的值则相对集中,多在 0.02~0.04 的较小范围内,只占值域的 1/3 大小,同时可以反映低水误差的 F_3 和 F_4 之间有一定的负相

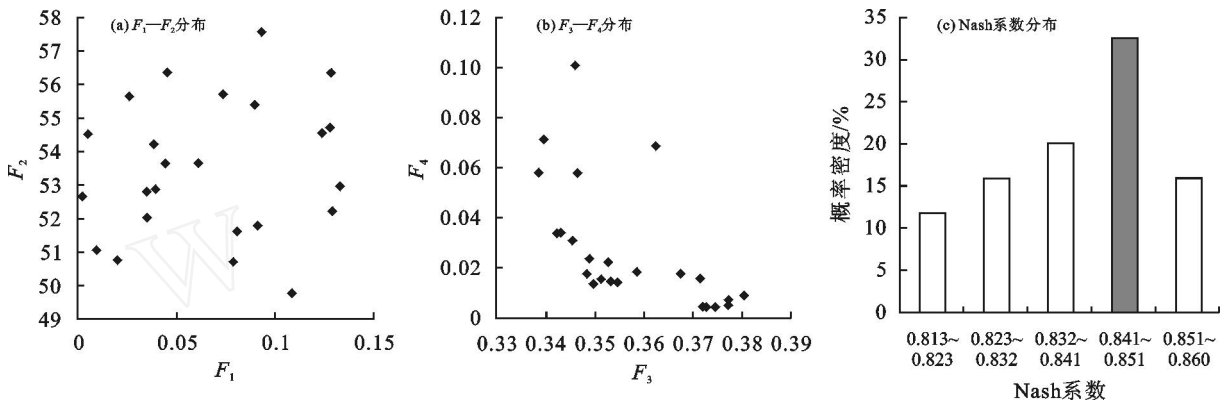


图 2 Pareto 优解的目标函数空间及 Nash 系数分布

2.2 参数空间

在模型不确定性和数据误差并存的情况下,多目标优化得到的 Pareto 参数可视为最优参数,Pareto 集中的所有参数同等重要且均可接受。图 3 给出了 Pareto 最优参数的分布,所有最优参数以其取值范围的边界作归一化处理。可以看出,Pareto 集中最优参数的变化幅度比较大,这是因为,参数的取值范围是根据流域气候土壤条件预估得到,已在其物理意义基础上做了取值限定,如 S_M 在本例中取值范围为 5~20 mm。另一方面,也反映了在现有数据条件下,所得到的优化参数的不确定性。参数 K 和 C_G 的最优分布都落在取值范围之内且没有越界的趋势,参数的范围设置合理。 S_M 的归一化值大部分集中于 0~0.8 之间, K_G 参数则分布于 0.6~1.0 之间,两参数的不确定性范围均接近边界,故可以考虑降低 S_M 取值范围的下界和提高参数 K_G 取值范围的上界。 S_M 虽是敏感参数,但在本例中可接受的范围较广,为 5~15.8 mm,其模型模拟效果仍然不错,表明在 S_M 参数的变化同时若调整其余参数同样可以使模型达到理想的模拟效果。在湿润地区,深层蒸散发系数 C 的取值对模型的影响不大,属于不敏感参数,因此正

相关性,即 F_3 目标函数值的减小将同时造成 F_4 值的增加, F_3 和 F_4 的互斥性使两个目标函数值很难同时取得优值。从模型角度上看,模型在低水模拟精度的提高将导致高水模拟精度的下降,这是模型系统性偏差的表现。因此改进模型以平衡高水和低水的模拟是提高模型效率的有效途径之一。在多目标优化的实际应用中,可根据不同的需求,通过设定目标函数的权值来控制模型的模拟行为。其余的目标函数组合结果类似 $F_1—F_2$ 组合,未一一列出。

图 2c 给出 24 组参数所对应的效率系数 (Nash 系数)。可以看出,模型效率普遍较高,且集中于相对高值区,分布较为理想。这表明算法通过多目标约束模型的模拟行为,可最大程度地提高模拟效率。

如图 3 所示其取值空间的离散程度是最高的。由图 3 可知,5 个参数的大致不确定范围依次是, K_G :0.34~0.5, S_M :5~15.8, K :0.62~0.93, C_G :0.982~0.9971, C :0.08~0.18。

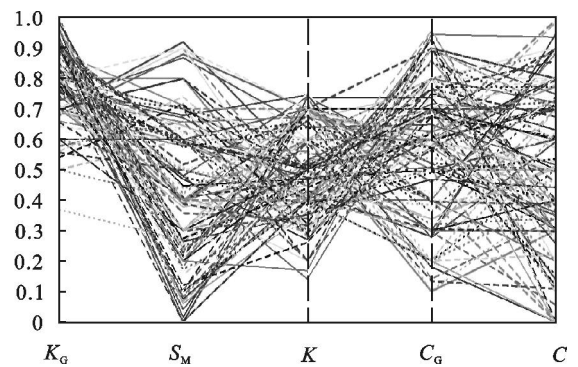


图 3 归一化的参数空间

2.3 流量过程线

如图 4—5 所示,虚线为实测流量,其余曲线为 Pareto 中任选的 24 组参数的模拟情况,为了在图中更直观显示流量过程线中占较大比例的低水过程,流量纵坐标采用对数刻度。

(1) Pareto 模拟的流量线趋势和实测吻合较好,

且绝大部分实测流量过程线落在模型预测范围内,从侧面验证了模型结构的合理性。

(2) 低水的模拟情况较好,个别小洪峰的模拟值稍微偏低,实测流量过程线位于的模拟区间的中央。

(3) 高水的模拟不尽人意,在两年的夏季中,均有2~3次大洪峰出现模拟滞后的情况,并且洪峰值偏大,而在1983年中,最大一次洪峰的模拟值偏小。这主要是由于在日模拟中未对汇流参数作进一步率定造成的。

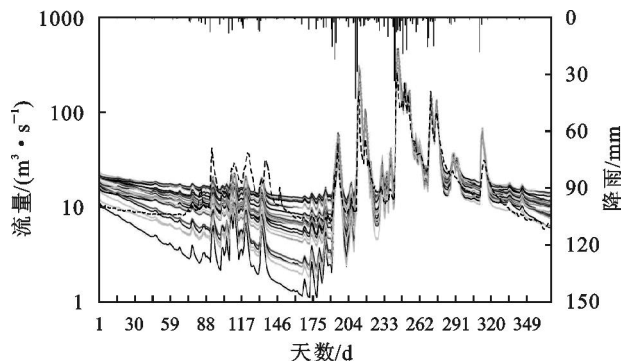


图4 1982年实测与 Pareto 模拟日流量过程线

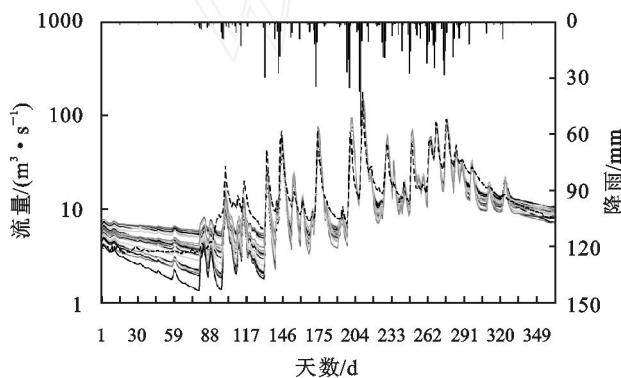


图5 1983年实测与 Pareto 模拟日流量过程线

(4) 通过1982年与1983年的径流拟合情况比较,可发现模型模拟误差在时间序列上具有较大的相似之处。首先,在降水稀少的1—3月,流域处于退水过程,此时流量拟合较为准确,实测值完全落在模拟区间中,4月底5月初出现一些少量降水,这一期间模型的模拟精度有所下降,模拟结果总体偏低,但在后续的5—6月份中,模拟预测范围重新覆盖实测流量过程线,此时模型模拟的拟合准确性再次提高。这一状况一直持续到6月底7月初,当降水量和频率逐渐增加,模型的模拟过程线范围与实测流量过程线的拟合精度再次下降,体现在洪峰模拟值总体偏大,并出现一定的滞后。而有趣的是,到了8月中旬后,在降水量较大和降水频次仍然较高的情况下,模型模拟径流量对实测流量线的拟合精度重新回到较高的水平,在图中表现为模拟流量过程与实测流量过程线具有很高的重合度。由图4—5明显看出,这两年的模

拟均为以上涨落趋势,实测流量过程线在模拟流量过程线中相对位置的分布极其相似,因此该误差为模型模拟的系统性偏差。由于多目标方法考虑了参数不确定性,并将不确定性最终转化为模型预测流量的范围,因此在实测流量过程线与预测不确定性范围之外的偏差,可视作是由模型自身结构性的偏差引起的,同时两年的模拟中均体现同样的偏差性质,可进一步推测上述误差为模型自身结构引起的系统性偏差。主要体现在汉江上游江口流域的日尺度模拟中,模型对季节性降水的模拟有一定滞后,初期的模拟有一定偏差。而造成这一偏差的具体模型结构原因及该偏差在其它流域是否具有同样的适应性,有待于下一步研究工作的开展,这些研究对进一步分析模型的不确定性,为水文预报提供可靠的预测结果是必要的。

3 结论

(1) 本文引入多目标优化算法 MOCOM—UA,将其应用于新安江模型的参数识别,以汉江上游江口流域降水—径流模拟为例,得到了4目标函数情形下的 Pareto 参数空间和模型的预测范围,并据此初步分析了参数和模型的不确定性。文中优化得到的 Pareto 参数分布范围较广,总体模拟效果良好,反映了由水文模拟相关误差引起的异参同效现象。不同年份的 Pareto 集合流量过程线呈现年际相关的结果。计算结果同时表明,本文所使用的多目标算法可使参数空间有效收敛,约束模型模拟行为并提供高效的模型参数集,该多目标方法同样适用于具有众多输出变量的分布式水文模型及复杂的环境模型。

(2) 影响模型模拟精度的主要因素包括模型结构和模型参数的选择,而多目标方法通过 Pareto 优集的计算一定程度上识别了模型参数和预测不确定性,实测数据和预测范围之间的系统性偏差可为模型不确定性分析提供有用的信息。如何在模型参数识别中考虑输入数据的误差,具体分析模型预测的系统性偏差与模型结构偏差的相关性,是下一步研究的重点,这些研究需要采用新的误差分析方法,如贝叶斯递归方法等,对这些问题的研究和解决将有助于提高模型的模拟效率和水文预报的可靠性。

[参 考 文 献]

- [1] Gupta V K, Sorooshian S. The automatic calibration of conceptual catchment models using derivative-based optimization algorithms. *Water Resources Research*, 1985, 21 (4): 473—485.
- [2] Boyle D P, Gupta H V, Sorooshian S. Toward improved calibration of hydrological models: combining the strengths of manual and automatic methods[J]. *Water*

- Resources Research, 2000, 36 (12) :3663—3674.
- [3] 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.
- [4] 武新宇, 程春田, 赵鸣雁. 基于并行遗传算法的新安江模型参数优化率定方法[J]. 水利学报, 2004(11) :85—90.
- [5] Khu S T, Madsen H. Multiobjective calibration with Pareto preference ordering: An application to rainfall-runoff model calibration[J]. Water Resources Research, 2005, 41 :W03004.
- [6] Jan Seibert. Multi-criteria calibration of a conceptual runoff model using a genetic algorithm[J]. Hydrology and Earth System Science, 2000, 4(2) :215—224.
- [7] 陈炯烽, 张万昌. 基于遗传算法的新安江模型日模拟参数优选研究[J]. 水文, 2006(4) :32—38.
- [8] Duan Q, Sorooshian S, Gupta V K. Effective and efficient global optimization for conceptual rainfall-runoff models[J]. Water Resources Research, 1992, 28(4) :1015—1031.
- [9] 陆桂华, 郦建强, 杨晓华. 水文模型参数优选遗传算法的应用[J]. 水利学报, 2004(2) :50—56.
- [10] Nelder J A, Mead R. A simplex method for function minimization [J]. Computer J, 1965, 7(2) :308—313.
- [11] 王建平, 程声通. 遗传单纯形混合算法在复杂环境模型参数识别中的应用[J]. 水利学报, 2005(6) :1—7.
- [12] 张洪刚, 郭生练, 刘攀, 等. 概念性水文模型多目标参数自动优选方法研究[J]. 水文, 2002(22) :12—16.
- [13] Brazil L E. Multilevel calibration strategy for complex hydrologic simulation models[D]. Fort Collins: Colorado State University, 1988.
- [14] Cheng, C T, Oub C P, Chauc K W. Combining a fuzzy optimal model with a genetic algorithm to solve multi-objective rainfall-runoff model calibration [J]. Journal of Hydrology, 2002, 268 : 72—86.
- [15] Beldring S. Multi-criteria validation of precipitation-runoff model [J]. Journal of Hydrology, 2002, 257 : 189—211.
- [16] Singh V P, Frevert D K. Mathematical Models of Large Watershed Hydrology [M]. U S: Water Resources Publications, 2002.
- [17] Beven K. A manifesto for the equifinality thesis[J]. Journal of Hydrology, 2006, 320:18—36.
- [18] Yapo P O, Gupta H V, Sorooshian S. Multi-objective global optimization for hydrological models[J]. Journal of Hydrology, 1998, 204:83—97.
- [19] 杨晓华, 杨志峰, 郦建强, 等. 水文模型参数识别算法研究及展望[J]. 自然科学进展, 2006(16) : 657—661.
- [20] 陈南祥, 李跃鹏, 徐晨光. 基于多目标遗传算法的水资源优化配置[J]. 水利学报, 2006, 37(3) :308—313.
- [21] 长江水利委员会. 水文预报方法(2版)[M]. 北京:水利电力出版社, 1993:126—134, 42—56.
- [22] 赵人俊, 王佩兰. 新安江模型参数的分析[J]. 水文, 1988(6) :2—9.
- [23] Gupta H V, Sorooshian S, Yapo P O. Toward improved calibration of hydrological models: multiple and noncommensurable measures of information[J]. Water Resources Research, 1998, 34 (4) :751—763.
- [24] Madsen H. Automatic calibration of a conceptual rainfall-runoff model using multiple objectives[J]. Journal of Hydrology, 2000, 235 :276—288.