

基于 IBUNE 方法的水文模型不确定性分析

贺颖庆¹,任立良²,李彬权²

(1.江苏省水文水资源勘测局,江苏 南京 210029; 2.河海大学水文水资源学院,江苏 南京 210098)

摘要:在贝叶斯理论框架下,根据一种可结合多个水文模型给出模拟或预报结果的 IBUNE 方法探讨了水文模型的输入、参数以及结构的不确定性问题。将 SCEM-UA 算法和 EM 算法嵌入新安江和 TOPMODEL 水文模型用于参数优化和模型平均,进而将输入与参数的综合不确定性处理后得到的预报量后验分布进行多模型综合,据此对水文模型的不确定性及其对水文模拟结果的影响进行评价。以湖南沅水流域龙家山水文站以上集水区域为例进行了应用研究,结果表明,IBUNE 方法能够有效估计水文模型的不确定性,并能给出合理的概率预报区间。

关键词:水文模型;不确定性;IBUNE;SCEM-UA;贝叶斯理论

中图分类号:P333

文献标识码:A

文章编号:1000-0852(2016)02-0023-05

流域水文模型是对自然水文物理变化过程从不同层面的概化和抽象,因此,自然水文过程的复杂性和人类认识水平的局限性必然会导致水文模型在水循环要素预测和模拟过程中的不确定性。利用水文模型进行模拟预报的不确定性一般可以归为三类:(1)模型输入的不确定性;(2)模型参数的不确定性;(3)模型结构的不确定性。在以往的研究中,人们大多仅仅关注模型参数的不确定性问题,特别地,传统的模型参数率定就是忽略其他两种不确定性,将模拟和预报误差都归结到模型参数的误差上。然而,模型参数间可能存在互补性和相关性,两套不同的参数组合往往取得相同(相当)的模拟或预测结果,即所谓的“异参同效”现象。近年来,人们也从不同角度综合分析了模型输入、参数和结构的不确定性。Ajami 等人^[1]在 2007 年提出的贝叶斯综合不确定性估计法(Integrated Bayesian Uncertainty Estimator,简称 IBUNE)就是其中一个有益的尝试。目前,该方法在国内还较少应用。

本文选用三水源新安江和 TOPMODEL 模型,将 IBUNE 方法应用于湖南沅水流域,在贝叶斯框架下综合分析了水文模型模拟和预报中的不确定性。

1 IBUNE 方法理论

1.1 贝叶斯方法

贝叶斯定理(贝叶斯公式)可表示为:

$$p(A|B) = \frac{p(A)p(B|A)}{p(B)} \quad (1)$$

式中: $p(A)$ 是事件 A 的先验概率; $p(A|B)$ 是已知事件 B 发生后 A 的条件概率,也称作 A 的后验概率; $p(B|A)$ 是已知 A 发生后 B 的条件概率,也叫相似度; $p(B)$ 是 B 的先验概率。

$p(B|A)/p(B)$ 也被称作标准相似度,则贝叶斯定理可表述为:后验概率=标准相似度×先验概率。即,

$$p(A|B) = L(B|A) \times p(A) \quad (2)$$

针对参数的不确定性,根据贝叶斯统计将模型参数视为随机变量,并服从联合后验概率分布 $p(\theta|\tilde{X}, \tilde{y})$,它表示在已知观测数据降水量 \tilde{X} , 流量 \tilde{y} 的情况下,模型参数 θ 的概率特征。应用 Shuffled Complex Evolution Metropolis (SCEM-UA) 算法可确定模型参数的后验分布。

针对模型结构的不确定性,利用多模型综合技术——贝叶斯模型平均法(BMA),将综合预报量 y_{bma}

收稿日期:2015-02-08

基金项目:国家自然科学基金青年项目(51509067)

作者简介:贺颖庆(1984-),女,黑龙江牡丹江人,工程师,主要从事水文测验、资料整编。E-mail:jsswhyq@foxmail.com

通讯作者:李彬权(1984-),男,江苏淮安人,讲师,主要从事流域水文过程及不确定性理论研究。E-mail:libinquan@hhu.edu.cn

视为随机变量,令 y 为实测流量, $M=[M_1, M_2, \dots, M_k]$ 是 k 个水文模型, y_k 是模型 M_k 的预报量,根据贝叶斯定理和全概率公式, BMA 综合预报量 y_{bma} 的后验分布为:

$$p(y_{bma} | M_1, \dots, M_k, \tilde{X}, \tilde{y}) = \sum_{k=1}^K p(M_k | \tilde{X}, \tilde{y}) \times p(y_k | M_k, \tilde{X}, \tilde{y}) \quad (3)$$

1.2 输入不确定性

在输入不确定性的处理上, IBUNE 方法引入了一个雨深乘数:

$$r_t = \phi_t \tilde{r}_t; \quad \phi_t \rightarrow N(a, v^2) \quad (4)$$

式中: \tilde{r}_t 为 t 时段的实测降水量; ϕ_t 表示 t 时段的随机乘数,服从正态分布 $N(a, v^2)$, 本文假定 $a \in [0.9, 1.1]$, $v \in [10^{-5}, 10^{-3}]$ 。

1.3 SCEM-UA 算法

SCEM-UA 算法是 Vrugt 等人对单纯多边形进化算法 (SCE-UA)^[2] 加以改进而提出的一个全局搜索算法, 它将马尔可夫链蒙特卡罗 (Markov chain Monte Carlo, MCMC) 采样方法与 SCE-UA 中复合型重组的概念结合, 形成一个不仅能搜索到一套具有最大似然性的参数组, 同时也能估计参数不确定性的算法^[3]。本文首先确定各模型的敏感参数, 假定敏感参数的先验分布为已知上下限的均匀分布, 利用 SCEM 算法分别优化各个模型敏感参数, 并估计参数的后验分布, 并利用产生的 20 000 组收敛到后验分布的随机参数变量来确定预报量的置信区间。

1.4 EM 算法

期望最大化 (Expectation-maximization, EM) 是 A. P. Dempster 等人于 1977 年首次提出的, 属于最大似然法中的一种^[4]。通过 EM 算法统计 IBUNE 方法中各个模型的后验分布, 即似然函数值。本文假定 BMA 预报量的先验分布为均匀分布, 赋予每个模型统一的初始权重 0.5; 获得实测数据信息后, 利用 EM 算法估计各个模型的最终权重, 将输入与参数的综合不确定性处理后得到的预报量后验分布进行多模型综合, 即可得到最终的预报结果和某一置信度下的预报量置信区间。

2 研究区域及水文模型

2.1 研究区域及资料情况

本文选择湘江一级支流沅水流域龙家山水文站上集水区作为研究区域, 流域面积 4 515 km², 边界内有 12 个雨量站, 1 个蒸发站。选用 2000~2006 年 7a 的日

径流资料进行日径流模拟, 同时摘取 12 场包含单峰和复式等不同形状的洪水过程进行洪水演算。采用空间分辨率 3" 的 DEM 数据提取集水区的流域边界, 并划分为 6 个子流域, 生成的水系与实际水系基本一致。

2.2 水文模型及其敏感参数

利用基于子流域单元的三水源新安江模型和 TOPMODEL 进行日径流和洪水模拟演算。分单元新安江模型的特点是分阶段 (坡地和河网) 汇流, 对每个单元流域作分水源产汇流计算, 得出单元流域的出口流量过程, 再进行出口以下的河道洪水演算, 求得流域出口的流量过程。半分布式流域水文模型 TOPMODEL 以地形空间变化为主要结构, 以地形指数 $\ln(\alpha/\tan U)$ 或土壤地形指数 $\ln(\alpha/T_0 \tan U)$ 描述水流趋势和由于重力排水作用径流沿坡向的运动, 以地形指数的空间分布来反映流域水分亏缺的空间分布^[5]。

采用抽样方法逐个验证参数的敏感性, 结果表明: 新安江模型参数中的 B、C 和 EX 灵敏度较低, 而敏感性参数为 KC、SM、CI、CG、KG 和 CS; TOPMODEL 模型的敏感性参数为 S_{rmax} 、 S_{zm} 、 T_0 、 Sr_0 、 R_v 、 T_d 。灵敏度较低的参数对水文预报不确定性的影响较小, 所以, 为减少计算量, 本文只对敏感性参数进行不确定性分析。

3 实例分析

3.1 参数不确定性分析

利用 2000~2006 年连续七年的日径流资料, 根据式 (3) 计算各参数后验概率密度, 应用 SCEM-UA 算法确定两个模型的参数后验分布, 算法设置为: 初始均匀分布样本个数 1 000, 复合型个数 5, 算法收敛后继续运行, 最大运行次数为 25 000 次。由 SCEM-UA 算法确定的分布可获得后验概率密度最大的参数值, 也就是使模型模拟的效果最优的参数值, 并估计出各参数的不确定边界。算法收敛后从参数后验分布中产生 20 000 组参数样本代入水文模型中, 获得包含有 20 000 个系列的模拟流量的集合。图 1 和图 2 是利用 SCEM-UA 算法从新安江模型和 TOPMODEL 的参数后验分布中抽取的 20 000 组参数作出的频数分布图。结果表明: SCEM-UA 算法可以在给定先验分布的情况下搜索到最优参数组, 同时估计出参数的不确定性范围。不确定程度越高的参数, 越难以推断其后验分布, 其分布形状越不规则。

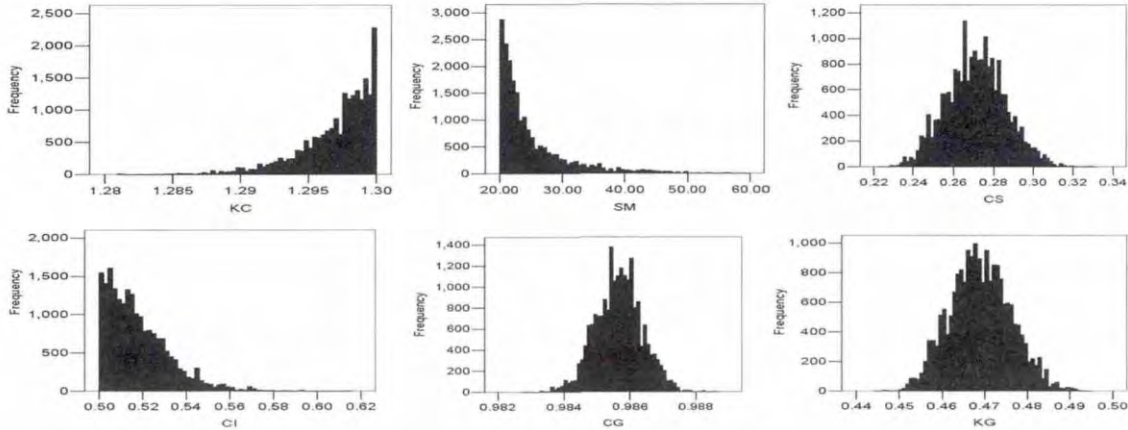


图1 新安江模型日模参数后验分布直方图

Fig.1 Posterior distribution histogram of the Xin'anjiang model parameters

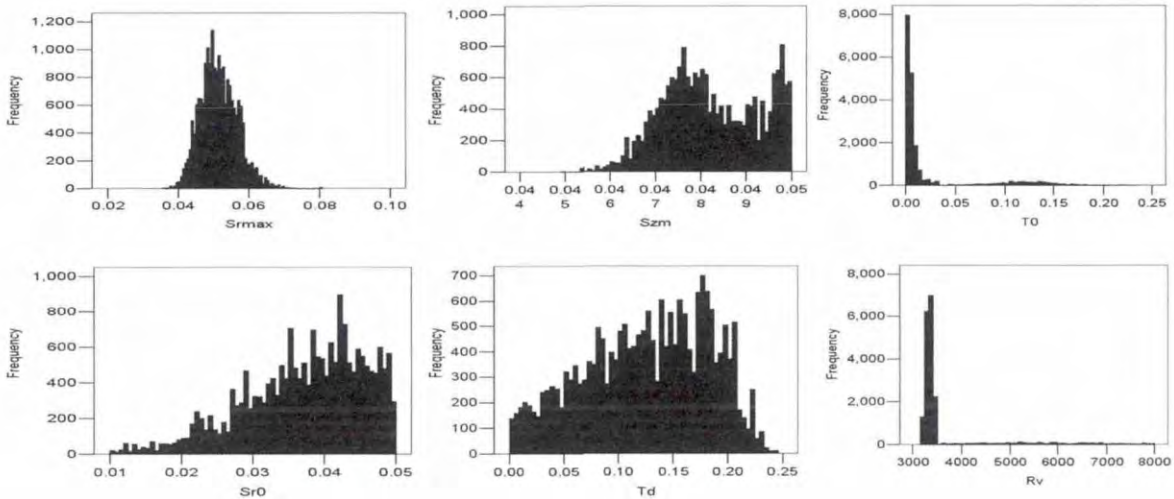


图2 TOPMODEL 日模参数后验分布直方图

Fig.2 Posterior distribution histogram of the TOPMODEL parameters

3.2 参数与输入的综合不确定性分析

考虑输入不确定性,将式(4)雨深乘数分布 $N(a,v)$ 中的 a 和 v 加入到模型敏感参数中,再次运用 SCEM-UA 进行优化,各参数后验分布边界向更合理的方向移动了,分布形状更加接近正态分布,同时输入误差模型的方差参数 v 难以向单一数值集中。由此可以判定,模型输入存在着不确定性。考虑输入不确定性后日径流模拟精度与仅考虑参数不确定性,未考虑输入不确定性的 20 000 组日径流模拟精度对比情况见表1。结果表明,综合考虑模型参数和输入不确定性的模拟结果要优于仅考虑参数不确定性的结果。次洪过程模拟 Nash 系数、洪峰流量、峰现时间同样。

3.3 参数、输入与模型结构的综合不确定性分析

综合预报量的后验分布由各模型的权重和各模型的预报流量后验分布两部分组成,其中,预报流量的后验分布可以通过前文的输入和参数不确定性估

计推断出来,而模型的权重需要通过 EM 算法进行估计,通过迭代最终可以找到两模型权重的估计值,日径流模拟新安江模型权重 0.69, TOPMODEL 权重 0.31; 次洪模拟新安江模型权重 0.61, TOPMODEL 权重 0.39。对于本文的研究区域,新安江模型的权重系数大于 TOPMODEL,与精度统计的结果相符,因此,新安江模型对最后的结果贡献量更大。根据 20 000 组来自 BMA 方法后验分布的模拟流量样本可以得到概率预报区间,并通过计算平均值等到单一值预报结果。以 2002 年日径流为例,图 3 中绘制的置信概率 95% 的 IBUNE 预报流量区间能够将大部分的实测流量包含在内。

对三种来源的不确定性作不同处理情况下实测流量散点落在 95% 置信区间内的百分比统计见表 2; IBUNE 方法日径流和次洪过程的模拟精度统计见表3,表 4。

表1 日径流模拟精度统计表
Table1 The daily runoff simulation precision statistics

年份	实测径流深 / mm	新安江模型				TOPMODEL			
		Nash 系数平均值		径流深平均相对误差 / %		Nash 系数平均值		径流深平均相对误差 / %	
		para	para+input	para	para+input	para	para+input	para	para+input
2000	1003.48	0.90	0.91	5.95	0.60	0.76	0.81	15.11	9.55
2001	1041.61	0.88	0.88	-2.62	-10.30	0.80	0.83	1.68	-3.75
2002	1426.87	0.92	0.93	-2.33	-7.62	0.85	0.87	6.59	-0.61
2003	730.90	0.86	0.87	12.75	1.43	0.84	0.88	14.17	6.88
2004	809.60	0.77	0.75	6.78	0.52	0.64	0.71	20.89	17.39
2005	905.38	0.86	0.87	13.11	5.53	0.74	0.81	19.54	13.37
2006	1122.55	0.87	0.89	21.30	13.20	0.76	0.81	26.17	17.88
00~06	7040.39	0.87	0.87	7.18	0.04	0.77	0.82	14.30	7.97

注:para 表示考虑参数不确定性;input 表示考虑输入不确定性。

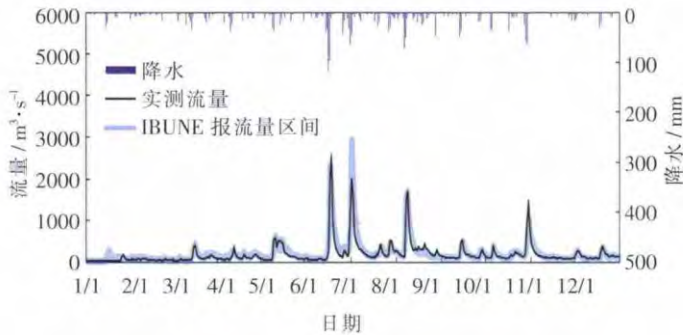


图3 2002年日径流 IBUNE 方法预报流量区间

Fig. 3 The daily runoff flow interval in 2002 with IBUNE method

表2 置信区间准确度统计表

Table2 The confidence interval accuracy statistics

百分比 / %	SCEM-UA (para)		SCEM-UA (para+input)		两模型综合 (para+input)	
	TOP-新安江	MOD-EL	TOP-新安江	MOD-EL	简单平均	IBUNE
7年日径流	34.45	52.87	38.68	55.22	79.11	80.22
12场洪水过程	54.82	63.35	55.56	65.34	65.66	74.83

表3 日径流模拟精度统计表

Table3 The runoff simulation accuracy statistics

年份	日径流模拟		
	Nash 系数平均值	计算径流深平均值/mm	径流深平均相对误差/%
2000	0.8972	1036.88	3.33
2001	0.8974	955.46	-8.27
2002	0.9104	1349.13	-5.45
2003	0.9031	753.73	3.12
2004	0.9004	856.15	5.75
2005	0.8980	955.44	5.53
2006	0.8984	977.44	-12.93
00~06	0.9007	6884.22	-2.22

表4 次洪过程模拟精度统计表

Table4 The flood simulation accuracy statistics

次洪过程模拟		
洪号	Nash 系数平均值	洪峰相对误差平均值/%
20000408	0.8193	-10.30
20000605	0.8195	7.99
20000901	0.9778	-1.80
20010405	0.8363	-1.12
20010515	0.8199	3.24
20010608	0.9413	-9.77
20020806	0.9415	0.92
20020910	0.8521	5.27
20030514	0.9265	-4.83
20030602	0.8448	-8.87
20040515	0.8499	-8.51
20040813	0.8257	10.20

结果显示,IBUNE 方法的模拟流量平均确定性系数比任何一个模型的确定性系数都要高,且原来相对误差较大的年份和洪水场次在综合两模型的模拟结果后相对误差都有所下降。这里采用绝对误差 ABS 精度评价指标,其计算公式为:

$$ABS = \frac{\sum_{t=1}^T |y_t^{obs} - y_t'|}{T} \quad (5)$$

式中:T 为时段数; y_t^{obs} 为 t 时刻的实测流量; y_t' 为 t 时刻的模拟流量。

20 000 组日径流模拟的 ABS 概率分布见图4, IBUNE 方法与单个水文模型相比,单个模型考虑参数和输入不确定性与仅考虑参数不确定性相比,ABS 值

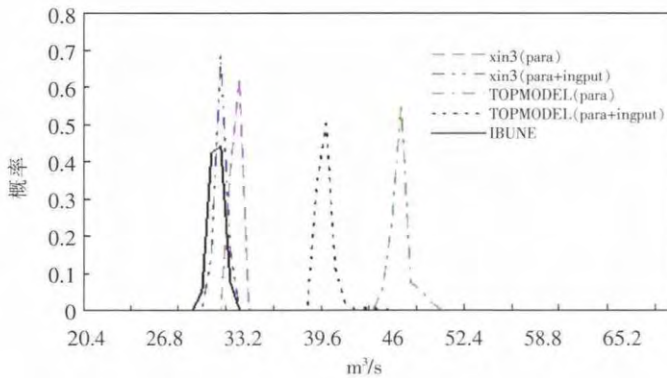


图4 不同情况下 ABS 概率分布图

Fig.4 The probability distribution of ABS under different conditions

均有明显减小。

4 结论

IBUNE 方法能综合考虑输入、参数和模型结构的不确定性,在利用 SCEM-UA 算法分别估计各模型的参数和输入不确定性的基础上,结合多个模型的模拟结果,进而利用 BMA 方法获得综合预报结果。本文选择新安江模型和 TOPMODEL 两个水文模型对湖南_γ 流域龙家山水文站以上集水区域的日径流和次洪过程进行模拟,分别得到了单个模型只考虑参数不确定性、单个模型考虑参数和输入不确定性,以及两模型考虑参数和输入不确定性后经 BMA 方法综合,这三种情况下的水文模拟结果。对结果进行对比分析可知 IBUNE 方法在不确定性估计和多模型综合预报方面具有一定的先进性:

(1)IBUNE 不仅能够综合不同模型给出精度较高的预报流量确定值,而且还可以较准确地给出一定置

信度下的预报流量区间。

(2)加入输入不确定性与只考虑参数不确定性相比,置信区间包含的实测流量百分比有所增加。两模型综合与单个模型相比,置信区间包含的实测流量百分比有大幅度增加。

(3)与简单地取两模型的算术平均值相比,IBUNE 更能体现在研究区域内某一模型的优劣,综合置信区间更加准确可靠。

综上,IBUNE 方法对于提高水文预报精度、研究水文模型不确定性有重要意义。该方法具有综合、完整、算法先进等优点。

参考文献:

- [1] Ajami N.K., Duan Q., Sorooshian S.. An integrated hydrologic Bayesian multimodel combination framework: confronting input, parameter, and model structural uncertainty in hydrologic prediction[J]. *Water Resources Research*, 2007,43:W014303.
- [2] Qingyun Duan, Sorooshi Sorooshian S, Vuai Gupta. Effective and efficient global optimization for conceptual rainfall-runoff models[J]. *Water Resources Research*, 1992,28(4):1015-1031.
- [3] Vrugt J.A., Gupta W., Bouten W., et al. Shuffled complex evolution metropolis algorithm for optimization and uncertainty assessment of hydrological model parameters[J]. *Water Resource Research*, 2003,39(8):1.
- [4] Q Duan, Ajami N K, X Gao, et al. Multi-model ensemble hydrologic prediction using Bayesian model averaging [J]. *Advances in Water Resources*, 2007,30:1371-1386.
- [5] 郭方,刘新仁,任立良.以地形指数为基础的流域水文模型 TOPMODEL 及其拓宽应用[J]. *水科学进展*, 2000,11(3):296-301. (GUO Fang, LIU Xinren, REN Liliang. A topography based hydrological model: TOPMODEL and its widened application [J]. *Advances in Water Science*, 2000,11(3):296-301. (in Chinese))

Uncertainty Analysis of Hydrologic Model with IBUNE Method

HE Yingqing¹, REN Liliang², LI Binquan²

(1. *Hydrology and Water Resources Survey Bureau of Jiangsu Province, Nanjing 210029, China;*

2. *College of Hydrology and Water Resources, Hohai University, Nanjing 210098, China)*

Abstract: In order to get more accurate modeling or forecasting results, a Bayesian multi-model comprehensive framework-IBUNE (Integrated Bayesian Uncertainty Estimator) was used in this paper to analyze the uncertainties from observation data, model parameter and structure based on probability and statistics methods. Research based on Bayesian theory, SCEM-UA algorithm and EM algorithm was embedded in XAJ model and TOPMODEL to optimize parameters and average models. Example shows that IBUNE method can efficiently estimate the uncertainty of hydrologic model, and give a reasonable interval of probability forecast.

Key words: hydrologic model; uncertainty; IBUNE; SCEM-UA; Bayesian