

DOI :10.3876/j.issn.1000-1980.2009.02.001

基于贝叶斯理论的 TOPMODEL 参数不确定性分析

梁忠民 ,李彬权 ,余钟波 ,华家鹏 ,刘金涛

(河海大学水文水资源与水利工程科学国家重点实验室 ,江苏 南京 210098)

摘要 :应用贝叶斯理论探讨了流域水文模型参数及预报不确定性问题 .通过 Monte-Carlo 途径确定 TOPMODEL 模型的敏感参数 ,采用 MCMC 抽样技术估计敏感参数的后验概率密度分布 ,并根据参数的抽样系列构造水文模型预报值的经验分布 ,据此对模型参数的不确定性及其对水文预报结果的影响进行评价 .以浙江密赛流域为例进行了应用研究 ,提供了模型参数及预报结果不确定性的定量分析结果 .

关键词 :贝叶斯理论 ;TOPMODEL ;流域水文模型 ;不确定性分析 ;MCMC

中图分类号 :P333 文献标识码 :A 文章编号 :1000-1980(2009)02-0129-04

水文模型参数反映的是流域水文特征 ,往往需要通过实测资料率定一套“最佳”的模型参数^[1-5] .实际应用中 ,模型参数的率定不可避免地存在着不确定性 ,必将导致对水文过程模拟和预测结果的不确定性 .目前 ,关于这种不确定性的定量描述及其对水文预报不确定性的影响评价 ,在国内外已成为研究热点 ,其中 Beven 等^[6]提出的 GLUE 方法在国内外水文领域广泛应用 .

本文通过模型参数变差系数与确定性系数或 Nash-Sutcliffe(NS)系数的关系 ,确定水文模型敏感参数 .对敏感参数 ,采用马尔科夫链蒙特卡罗法(Markov Chain Monte Carlo ,MCMC)抽样技术^[7] ,从参数后验分布进行随机抽样 ,进而得到对应预报值的经验分布 .根据参数的后验分布及预报值的经验分布即可对参数的不确定性及其对水文预报不确定性的影响进行定量分析 .本文以浙江密赛流域为例 ,选择 TOPMODEL 参数的不确定性问题进行研究 .

1 模型参数敏感性分析

在水文模型的识别中 ,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 ,本文只对灵敏度较高的模型参数进行不确定性分析 ,因此首先分析参数的敏感性 .对第 i 个模型参数 θ_i 分析的具体步骤如下 (a)根据模型识别方法得到模型参数 θ_i 的率定值 $\hat{\theta}_i$ ($i = 1, 2, \dots, n$) , n 为模型参数个数 .(b)将 θ_i 作为随机变量并假设其概率分布 ,本文假设其服从对数正态分布 :

$$f(\theta_i) = \frac{1}{\sqrt{2\pi}\theta_i\sigma} e^{-\frac{(\ln\theta_i - a)^2}{2\sigma^2}} \tag{1}$$

其中 $a = \ln\hat{\theta}_i$ $\sigma = C_V a$

式中 C_V 为 θ_i 的变差系数 .从该分布中随机抽取 θ_i ,代入到水文模型中(其他参数采用 $\hat{\theta}_j$ ($j \neq i$)) ,得到预报流量过程 ,计算确定性系数 ,记为 N_i .(c)改变 C_V 值 ,重复步骤(b) ,则可以建立第 i 个模型参数的 $N_i \sim C_V$ 关系 .(d)重复步骤(b)和(c) ,得到所有模型参数的 $N_i \sim C_V$ 关系 .(e)分析模型各参数的确定性系数 N_i 随 C_V 值的变化情况 ,给出衡量模型参数是否敏感的阈值 λ .若 N_i 的变化区间大于 λ ,则认为模型参数 θ_i 为敏感性参数 .

收稿日期 :2008-03-19

基金项目 :国家自然科学基金(50779013) ;教育部博士点基金(20070294018) ;长江学者和创新团队发展计划(IRT0717) ;高等学校学科创新引智计划(B08048)

作者简介 :梁忠民(1962 —) ,男 ,辽宁凤城人 ,教授 ,主要从事水文不确定性研究 .

2 模型参数不确定性分析

2.1 贝叶斯方法

贝叶斯理论的基本原则是利用数据获得变量的条件分布,即变量的后验概率密度函数,公式为

$$h(\theta/x)=\frac{L(x/\theta)g(\theta)}{\int L(x/\theta)g(\theta)d\theta}$$

(2)

式中: θ,x ——随机变量和与之相关的统计数据; $h(\theta/x)$ ——后验概率密度函数; $L(Q/\theta)$ ——似然函数; $g(\theta)$ ——先验分布。

对于本文研究应用的 TOPMODEL, θ 为模型的敏感性参数, x 为模型的输出变量,即流量 Q 。模拟或预报所需要的其他要素,如实测降雨、流量等次洪资料记为 w 。则式(2)改写为

$$h(\theta/Q,w)=\frac{L(Q/\theta,w)g(\theta/w)}{\int L(Q/\theta,w)g(\theta)d\theta}\propto L(Q/\theta,w)g(\theta)$$

(3)

似然度函数 $L(Q/\theta,w)$ 的度量有多种方法^[6],本文取如下形式:

$$L(Q/\theta,w)=\left(\prod_{j=1}^N\sigma_j^2\right)^{-1}\quad\sigma_j^2=\frac{1}{m}\sum_{k=1}^m(Q_{cal,k}-Q_{obs,k})^2$$

(4)

式中: $Q_{cal,k},Q_{obs,k}$ ——计算与实测的流量值; σ_j^2 ——第 k 场洪水误差的方差; m ——第 i 场洪水的总时段数; N ——次洪个数。

2.2 AM-MCMC 方法

由公式(3)(4)可知,参数后验分布一般难以解析计算,可以通过 MCMC 抽样技术研究模型参数的不确定性问题。本文采用一种改进的 MCMC 采样器,即自适应的 Metropolis 算法^[8](adaptive metropolis,AM)估计参数的后验分布。采样序列是否收敛于后验分布,可以采用 Gelman 等^[9]提出的一种定量的比例缩小得分(scale reduction score) $R^{1/2}$ 进行评判。若 $R^{1/2}$ 接近于 1,表示参数收敛于后验分布。详细理论及方法见文献[9]。

3 TOPMODEL

TOPMODEL^[10]是 1979 年由 Beven 和 Kirkby 提出的一个以地形为基础的流域水文模型^[10]。本文采用 TOPMODEL95.02 版本。模型共有 7 个参数,分别是: S_{zm} 为土壤下渗率呈指数衰减的速率; T_0 为土壤刚达饱和时的传导度; T_d 为重力排水的时间滞时参数; S_{max} 为根带最大蓄水能力; S_{i0} 为根带土壤的初始缺水量; R_V 为地表坡面汇流的有效速度; CH_V 为主河道汇流的有效速度。

4 算 例

4.1 流域及资料情况

选择浙江省开化县密赛流域作为分析实例。该流域面积为 700.6 km²,年降雨量在 1 500~2 000 mm 之间,属湿润地区。共选用了 1982~1988 年的 9 场次洪资料。

4.2 参数敏感性分析

本文研究中,计算时段 $\Delta t=1\text{ h}$,并将马氏京根法参数 k_e 取为 Δt ,所以需要进行敏感性分析的参数共 7 个,表 1 给出了这些参数的取值范围。

以确定性系数 N_i 为目标函数,敏感性阈值 λ 取 0.2。分析表明,无论模型参数 T_0,T_d,CH_V 和 S_{max} 在取值范围内如何变化,目标函数 N_i 变化幅度都小于敏感性阈值 λ ,而参数 S_{zm} 和 R_V 取值的微小改变都会对模拟结果造成

较大影响,如图 1(a)所示。图 1(b)使用了 9 场次洪对模型参数 S_{i0} 进行敏感性分析,从图中可以看出 S_{i0} 并不敏感。结果表明,土壤下渗率呈指数衰减的速率参数 S_{zm} 和地表坡面汇流的有效速度 R_V 为最敏感的 2 个参数。

表 1 TOPMODEL 模型参数取值范围

Table 1 Upper and lower boundaries for parameters of TOPMODEL

参 数	S_{zm}/m	$\ln(T_0/(\text{m}^2\cdot\text{h}^{-1}))$	T_d	S_{max}/m	S_{i0}/m	$R_V/(\text{m}\cdot\text{h}^{-1})$	$CH_V/(\text{m}\cdot\text{h}^{-1})$
最小值	0	1	0	0.001	0	4 000	100
最大值	0.05	10	30	0.100	0.05	8 000	10 000

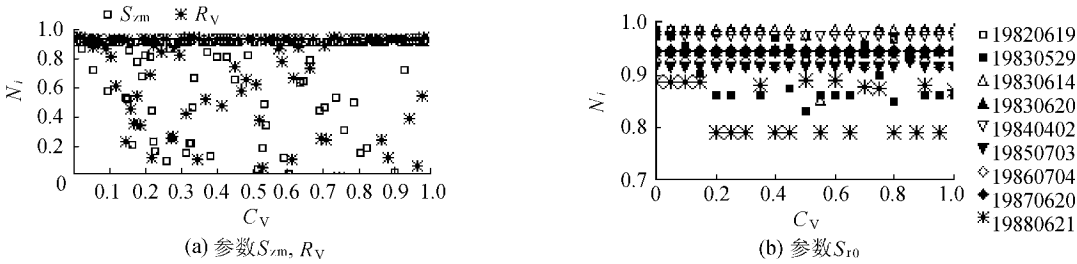


图 1 模型参数的 $N_i \sim C_V$ 关系

Fig. 1 Relationship between N_i and C_V for model parameters

4.3 MCMC 算法配置

作为示例, 本文参数先验分布采用均匀分布.

对 S_{zm} 和 R_V 采用 AM 算法对其后验分布进行估计. AM 算法的配置为: 参数初始协方差矩阵 C_0 为对角矩阵, 方差取参数取值范围(表 1)的 $1/20$; 初始迭代次数 $i_0 = 1000$. 算法平行运行 3 次, 每次采样 5000 个样本.

以参数 S_{zm} 的分析为例, 结果表明单一采样序列的参数 S_{zm} 的均值和方差基本达到稳定, 因此单一序列是收敛的. 计算的 $R^{1/2}$ 演化过程如图 2 所示. 由图 2 可知, 在迭代初期 ($i < 1000$), $R^{1/2}$ 变化较剧烈; 当 $i > 1000$ 后, $R^{1/2}$ 趋于稳定, 并接近于 1.0. 这说明参数 S_{zm} 和 R_V 的 MCMC 采样序列均能稳定收敛到其参数的后验分布. 综合考虑上述单序列和多序列参数的收敛情况, 将每一序列的前 1000 次舍去, 这样 3 次平行试验共采集了 12000 个样本用于参数后验分布的统计分析.

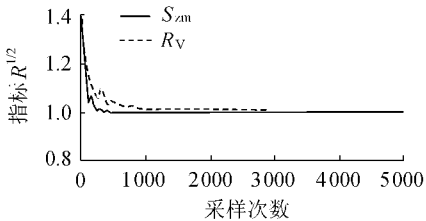


图 2 收敛指标值的变化过程

Fig. 2 Evolution of scale reduction scores

4.4 不确定性分析结果

根据 MCMC 抽样得到参数 S_{zm} 和 R_V 的后验分布见图 3. 由图 3 可知, R_V 比 S_{zm} 具有更大的不确定性. 分别取后验分布的众数作为 S_{zm} 和 R_V 的点估值, 并传统的 Ronsenbrock 方法优化结果^[11]一同列于表 2. 由表 2 可知, S_{zm} 的贝叶斯点估值与优化值较为接近, 而 R_V 相差较大. 由此说明在参数的识别中, 具有较小不确定性的参数更可能得到合理的率定结果; 相反, 若参数不确定较大, 则更可能得到具有相同效果的参数率定值.

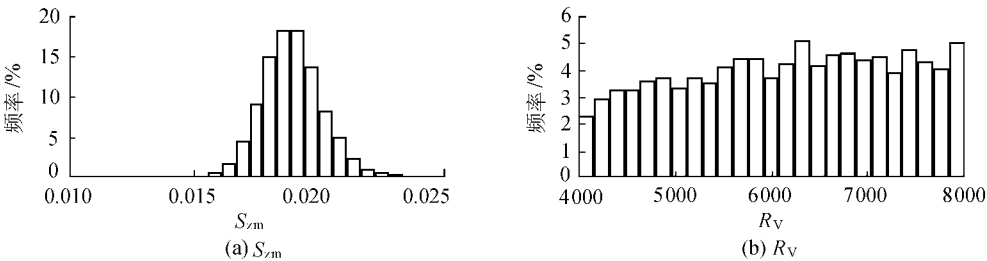


图 3 参数后验分布直方图

Fig. 3 Histogram of parametric posterior distribution

将 S_{zm} 和 R_V 的 MCMC 抽样值输入 TOPMODEL 中(其他输入变量及模型参数取自文献[11]), 得到预报变量的经验分布, 根据该分布可以构造不同置信水平的预报置信区间; 同时, 采用 S_{zm} 和 R_V 的贝叶斯点估值也可得到预报过程线. 图 4 提供了贝叶斯点估值、优化参数的预报结果及 90% 置信度的不确定性范围. 采用参数后验分布众数的贝叶斯点估值也可得到较好的预报结果(图 4 中贝叶斯预报值), 且可提供置信区间预报.

从图 4 可看出, 不确定性范围随流量而变, 在高流量区较大, 在低流量区较小. 此外, 大多数实测数据都包含于 90% 的置信区间内, 表明在 TOPMODEL 中, 考虑了敏感参数的不确定性后, 就可以较有效地定量描述预报结果的不确定性. 但是, 预报置信区间没能包含所有时段

表 2 MCMC 法和 Ronsenbrock 方法优化结果对比

Table 2 Comparison between results of MCMC and Ronsenbrock			
参数名称	MCMC 方法后验分布的众数	Ronsenbrock 方法优化值	相对差值/%
S_{zm}	0.019	0.02	5
R_V	6300	5000	26

的实测数据的事实也表明,仅通过敏感性参数的不确定性或仅考虑参数的不确定性,并不能对预报过程的不确定性做出全面的评价.

5 结 论

- a. 本文采用确定性系数 N_i 并结合随机抽样,建立 N_i 与参数 C_V 的变化关系,以此确定模型的敏感参数,结果表明, TOPMODEL 中的参数 S_{zm}, R_V 较为敏感.
- b. MCMC 算法可以用于模型参数后验分布的抽样,对参数的不确定性进行评价. 根据后验分布抽样,可以得到模型预报值的经验分布,由此可以提供诸如参数分布的众数、期望值等点估计值,也可以获得某一置信水平的区间预报,这为水文预报不确定性的定量评价及水文过程的概率预报提供了有效途径.
- c. 水文模型的应用包含着输入、模型参数和结构等诸多的不确定性因素,本文仅对参数的不确定性问题进行了初步探讨,更完整的结论需要进一步深入研究.

参考文献:

[1] 赵人俊. 流域水文模拟[M]. 北京: 水利电力出版社, 1984.

[2] 叶守泽, 夏军. 水文系统识别: 原理与方法[M]. 北京: 水利电力出版社, 1989.

[3] DUAN Q, SOROOSHIAN S, GUPA V K. Effective and efficient global optimization for conceptual rainfall-runoff models[J]. Water Resources Research, 1992, 28(4): 265-284.

[4] 马海波, 董增川, 张文明, 等. SCE-UA 算法在 TOPMODEL 参数优化中的应用[J]. 河海大学学报: 自然科学版, 2004, 34(4): 361-365.

[5] 芮孝芳, 朱庆平. 分布式流域水文模型研究中的几个问题[J]. 水利水电科技进展, 2002, 22(3): 56-58.

[6] BEBEN K, BINLEY A. The future of distributed models: model calibration and uncertainty prediction[J]. Hydrological Processes, 1992, 6: 279-298.

[7] 王建平, 程声通, 贾海峰. 基于 MCMC 法的水质模型参数不确定性研究[J]. 环境科学, 2006, 27(1): 24-30.

[8] HAARIO H, SAKSMAN E, TAMMINEN J. An adaptive Metropolis algorithm[J]. Bernoulli, 2001, 7(2): 223-242.

[9] GELMAN A, RUBIN D B. Inference from iterative simulation using multiple sequences[J]. Statistics Science, 1992, 7(4): 457-511.

[10] BEVEN K, KIRKBY M J. A physically based variable contributing area model of basin hydrology[J]. Hydrological Bulletin, 1979, 24: 43-69.

[11] 姚成. 基于栅格的分布式新安江模型构建与分析[D]. 南京: 河海大学, 2007.

Parametric uncertainty analysis for TOPMODEL based on Bayesian theory

LIANG Zhong-min, LI Bin-quan, YU Zhong-bo, HUA Jia-peng, LIU Jin-tao
(State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering,
Hohai University, Nanjing 210098, China)

Abstract: The Bayesian theory was employed to discuss uncertainties of parameter and predication of basin hydrological models. The sensitive parameters of TOPMODEL were determined using the Monte-Carlo method. The Markov Chain Monte-Carlo (MCMC) method was adopted to estimate the posterior distributions of these parameters. The empirical distribution of the model forecast was estimated by running TOPMODEL with parameters sampled from the posterior distributions. The uncertainties of model parameters and its effect on the hydrological forecast results were evaluated. The present procedure was applied to the Misai basin, providing quantitative results of uncertainties of parameters and model forecasts.

Key words: Bayesian theory; TOPMODEL; basin hydrological model; uncertainty analysis; Markov Chain Monte-Carlo

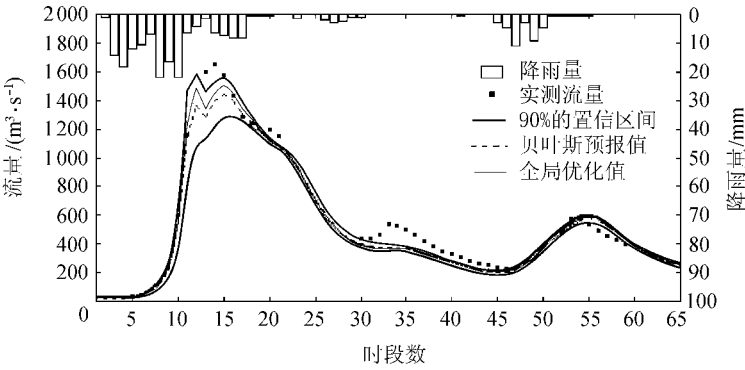


图 4 19820619 次洪水预报及不确定性分析结果
Fig. 4 Results of mode forecast and uncertainty analysis
for 19820619 flood event