

基于 Beta-PERT 分布的水文模型参数不确定性分析*

林凯荣^{1,2}, 陈晓宏², 江涛²

(1. 中国水利水电科学研究院, 北京 100044;
2. 中山大学地理科学与规划学院, 广东 广州 510275)

摘要: 以汉江中流域为例, 引进 Beta-PERT 分布来描述模型参数的分布, 结合蒙特卡罗模拟试验来研究 TOP-MODEL 模型参数分布特征对水文模拟不确定性的影响。结果表明: 参数分布特征对模拟结果的不确定性有很大的影响; 随着参数变差系数从 7.6% 变化到 37.8%, 模拟结果的不确定性区间在增大, 尤其是参数变差系数从 7.6% 变化到 15.1%, 之后变化趋于缓慢; 当偏态系数为 0.61 时, 得到的模拟结果的不确定区间最大; 参数分布特征的变化对信度水平为 5% 的不确定度的上边界的影响比较明显, 对下边界的影响则较小。

关键词: Beta-PERT 分布; 水文模型; 参数分布; 不确定性

中图分类号: P331 文献标志码: A 文章编号: 0529-6579 (2010) 04-0139-05

Parameter Uncertainty Analysis of Hydrological Model Based on Beta-PERT Distribution

LIN Kairong^{1,2}, CHEN Xiaohong², JIANG Tao²

(1. China Institute of Water Resources and Hydropower Research, Beijing 100044, China;
2. School of Geographical Science and Planning, Sun Yat-sen University, Guangzhou 510275, China)

Abstract: Parameter uncertainty is one of main sources of the uncertainty in hydrological modeling and forecast. In this paper, Beta-PERT distribution is employed to describe the distribution of parameters, using the Monte Carlo simulation method to study the effect of the distribution pattern of parameters on the uncertainty in hydrological modeling and forecast in the Hanzhong basin. The results show that the distribution of parameters had a significant effect on model output uncertainty. With the change of the coefficients of parameter variance from 7.6% to 37.8%, particularly from 7.6% to 15.1%, the width of the model output uncertainty interval markedly increases. The maximum uncertainty occurs when the coefficient of skewness is 0.61. The distribution pattern of parameters has more effect on the upper bound discharge at the significance level of 5% than the lower bound discharge.

Key words: Beta-PERT distribution; hydrological model; distribution of parameter; uncertainty

水文复杂性和不确定性是水文科学研究最为重要和棘手的两个方面, 特别是在全球水文及气候变化中, 水文不确定性成为水文科学发展中一个新的前沿课题^[1-2]。目前, 常用的参数不确定性估计方法是假设参数服从均匀分布或正态分布^[3-6], 为了

研究模型参数分布特征对水文模拟与预报不确定性的影响, 本文引进了 Beta-PERT 分布^[7]。Beta-PERT 分布已经用于其他具有很大不确定性的随机变量的分布的估计。这是因为该分布具有一些良好的性质: 首先, 该分布有上下限, 它表示的是一个

* 收稿日期: 2009-04-14

基金项目: 国家自然科学基金资助项目 (50809078, 50839005); 中国水利水电科学研究院开放基金资助项目 (IWHR02009003); 水资源与水电工程科学国家重点实验室开放基金资助项目 (2008B043); 中央高校基本科研业务费专项基金资助项目 (3161395)

作者简介: 林凯荣 (1980 年生), 男, 博士, 讲师; E-mail: linkr@mail.sysu.edu.cn

闭区间内的随机变量的概率分布。而其他的分布中随机变量的取值范围往往都是开区间,如正态分布中随机变量的取值范围是 $(-\infty, +\infty)$,这就与实际情况有一些出入。另外,该分布非常灵活,它可能有很多种不同的形状,可以是对称的,也可以是不对称的。Beta-PERT 分布不仅只要求提供变量的最大值、最小值和最可能值就可以对变量的概率分布的参数进行估计;而且可以根据需要,较好地拟合均匀分布、正态分布和对数正态分布。因此,本文旨在利用 Beta-PERT 分布来描述模型参数的分布,结合蒙特卡罗模拟法进一步探讨水文模型参数的不确定性问题。

1 Beta-PERT 分布

Beta-PERT 分布是 Beta 分布的一种,仅要求提供变量的最大值、最小值和最可能值就可以对变量的概率分布的参数进行估计。在本文讨论的问题中,水文模型参数的 Beta-PERT 分布概率密度函数为

$$f(x) = \begin{cases} \frac{x^{a-1}(1-x)^{b-1}}{B(a,b)} & 0 \leq x \leq 1; a, b > 0 \\ 0 & \end{cases} \quad (1)$$

式中, $B(a, b)$ 为标准的 Beta 函数,其表达式如下

$$B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} \quad (2)$$

其中, $\Gamma(q)$ 为伽玛函数,其表达式如下

$$\Gamma(q) = \int_0^{\infty} t^{q-1} e^{-t} dt \quad (3)$$

式(1)中参数 a, b 的计算式为

$$a = \frac{(x_{\text{mean}} - x_{\text{min}})(2x_{\text{mode}} - x_{\text{min}} - x_{\text{max}})}{(x_{\text{mode}} - x_{\text{mean}})(x_{\text{max}} - x_{\text{min}})} \quad (4)$$

$$b = \frac{a(x_{\text{max}} - x_{\text{mean}})}{(x_{\text{mean}} - x_{\text{min}})} \quad (5)$$

式中: $x_{\text{min}}, x_{\text{max}}, x_{\text{mode}}, x_{\text{mean}}$ 分别为水文模型参数的最小值、最大值、最可能值和平均值,其中平均值 x_{mean} 的计算式为

$$x_{\text{mean}} = x_{\text{min}} + 4x_{\text{mode}} + x_{\text{max}}/6 \quad (6)$$

2 研究流域和模型

2.1 流域概况

汉江是长江中游最大的支流,发源于秦岭南麓,干流经陕西、湖北两省,于武汉市汇入长江,全长 1 570 余 km,流域面积 15.9 万 km²。汉江流域在防洪和供水等方面在我国具有十分重要的地位,丹江口水库是南水北调中线工程的水源地,江汉平原是我国主要的商品粮生产基地之一。本文选

取汉江上游的汉中流域为研究区域,该区域面积 9 329 km²,流域位置在北纬 32°35' - 34°10' 和东经 106°10' - 107°30'。流域内有 50 个雨量站,3 个蒸发站(铁锁关、武侯镇和红庙塘),和 7 个水文站(茶店子、铁锁关、武侯镇、元墩、江口、马道和汉中)。汉江上游的暴雨具有量大、分配集中和笼罩面积广等特点,其洪水主要由暴雨形成,由于流域内山高坡陡,洪水汇流速度快,具有猛涨猛落、峰型尖瘦的特点。数据选取 1980 - 1986 年的逐日降雨量和水文站流量资料。

2.2 模型原理

TOPMODEL (TOP graphy based hydrological MODEL) 作为一个以地形为基础的半分布式流域水文模型,自 Beven 和 Kirkby 于 1979 年提出以来,已在水文领域获得了广泛的应用^[8]。其主要是利用地貌指数来反映流域水文现象,特别是径流运动的分布规律。模型结构简单,优选参数相对较少,物理概念明确,在集总式和分布式流域水文模型之间起到了一个承上启下的作用。TOPMODEL 主要通过流域含水量(或缺水量)来确定源面积的大小。而含水量的大小可由地形指数计算,因此 TOPMODEL 被称为以地形为基础的流域水文模型。

2.3 评价指标

因为传统的水文学的大多数研究都是尽力找到最好的估计值,所以水文模拟和预测结果都被设定提供点估计,这时一般采用 Nash-Sutcliffe 确定性系数作为评价指标。从数理统计学的角度看,点估计只是给出了模拟和预测内容的近似值,并没有给出这个近似值的误差范围和估计的可信程度。那么水文模拟与预测的不确定性估计引进了数理统计学区间估计的思想,就是为了弥补点估计的这个不足,这时只能用信度水平和置信区间以及统计特征进行评价。因此,为了评价模型参数分布对模拟结果的不确定性的影响,本文选取统计量和信度水平为 5% 的不确定度作为评价指标。其中统计量有变差系数 C_v 、偏态系数 C_s 和峰值系数 C_k 。变差系数 C_v 越大表示系列的离散程度越大;偏态系数 C_s 越大表示系列的不对称性越大;峰值系数 C_k 越大表示系列越平缓。

3 不确定性分析

3.1 参数分布区间对结果不确定性的影响

本文通过对参数的一系列扰动,利用 Beta-PERT 分布来描述模型参数的分布,结合蒙特卡罗法进行模拟试验,研究参数的不确定性对模拟的结

果不确定性的影响。选取 TOPMODEL 模型中的 4 个主要参数 S_{zm} 、 T_0 、 T_d 和 SR_{max} 进行研究；其中 S_{zm} 表示非饱和区最大蓄水深度； T_0 是饱和导水率； T_d 是时间参数； SR_{max} 则是根系区最大容水量。在模拟试验中，首先参数的取值范围和最可能值的确定是根据 GLUE 方法得到有效参数组确定的（似然函数阈值设为 0.7），得到的 4 个参数参数的取值范围和最有可能取值见表 1；然后，参数分布的区间按照每个参数最可能值的 $\pm 20\%$ ， $\pm 40\%$ ， $\pm 60\%$ ， $\pm 80\%$ ， $\pm 100\%$ 逐步地增加，即相应的参数的变差系数 C_v 分别为：7.6%，15.1%，22.7%，30.2%，37.8%。蒙特卡罗模拟的次数先取为 10 000。

表 1 模型参数的取值范围和最可能取值
Table1 Ranges and most likely values of parameters used in TOPMODEL model

参数	最小值 (x_{min})	最大值 (x_{max})	最可能值 (x_{mode})
S_{zm}/m	0.1	1	0.36
$T_0/(m^2 \cdot h^{-1})$	0.1	3	1.23
T_d/h	1	30	11.17
SR_{max}/m	0.1	0.5	0.19

表 2 参数分布区间对结果不确定性的影响统计
Table 2 Statistics of effect of parameter distribution interval on output uncertainty

参数统 序号	结果统计值				置信区间		
	$C_v/\%$	$C_v/\%$	C_s	C_k	5%	95%	90% CI
1	7.60	1.70	0.02	2.94	124.23	109.6	14.63
2	15.10	2.80	0.14	2.98	163.99	92.60	71.39
3	22.70	4.34	0.40	3.25	181.01	89.19	91.82
4	30.20	6.13	0.74	3.98	203.25	87.42	115.83
5	37.80	7.48	0.97	4.78	219.02	87.42	131.6
6	39.25	8.58	0.98	4.63	223.65	86.80	136.85
7	40.94	9.05	1.00	4.46	218.38	86.53	131.85
8	41.94	9.89	0.92	3.94	216.63	85.85	130.78
9	42.03	9.91	0.86	3.90	219.59	85.04	134.55
10	42.07	9.91	0.93	3.99	231.64	84.87	146.77
11	42.13	9.87	0.94	4.26	233.70	84.72	148.98

根据上面的试验，可以得到 5 次参数的扰动对应的结果，如表 2 中序号为 1-5 行所示。表 2 中置信区间为信度水平为 5% 的不确定度的上边界、下边界流量过程的平均值，以及两者的差值（称为 90% CI）。图 1 显示的是参数变差系数 C_v （%）与信度水平为 5% 置信区间关系。从中可以看出随

着参数变差系数从 7.6% 变化到 37.8%，模拟结果的不确定性区间在增大，尤其是参数变差系数从 7.6% 变化到 15.1%，之后变化趋于缓慢；参数变差系数的变化对信度水平为 5% 的不确定度的上边界的影响比较明显，对下边界的影响则较小。图 2 显示的是参数变差系数与模拟结果统计值（包括变差系数 C_v 、偏态系数 C_s 和峰度系数 C_k ）关系。从中可以得到随着参数变差系数增加，模拟结果的变差系数、偏态系数和峰度系数也在增加，但变化速度明显慢于参数；另一方面说明不仅对模拟结果的离散程度有影响，而且对模拟结果的对称性和峰度都有一定的影响。

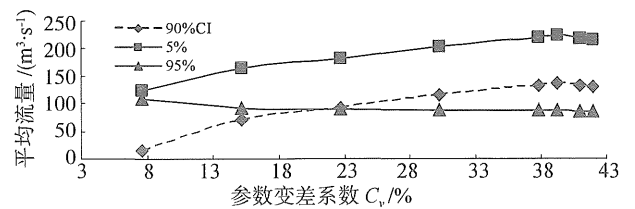


图 1 参数变差系数与信度水平为 5% 置信区间关系
Fig. 1 The relationship between C_v and prediction limits at the significance level of 5%

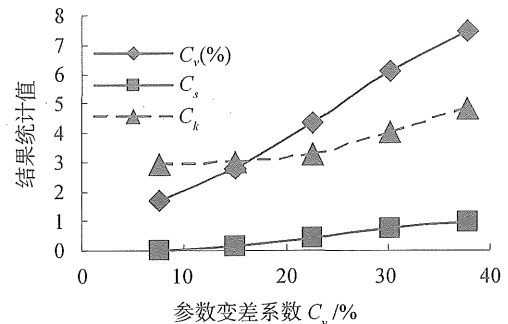


图 2 参数变差系数与模拟结果统计值关系
Fig. 2 The relationship between C_v and output distribution statistics at the significance level of 5%

接下来，我们尝试着继续增大参数的取值范围，参数变差系数从 39.25% 变化到 41.94%，结果如表 2 中序号为 6-8 行所示。图 1 中还显示了这次试验参数变差系数 C_v （%）与信度水平为 5% 置信区间关系。从中可以发现这次试验，随着参数变差系数增大，模拟结果的不确定性反而减小。为了寻找原因，我们在参数变差系数 C_v 为 41.94% 时，加大蒙特卡罗模拟试验的次数，从 10 000 变化到 40 000，结果如表 2 中序号为 8-11 行所示。图 3 显示了模拟试验次数与信度水平为

5% 置信区间关系。从中可以发现随着模拟次数的增加, 模拟结果的不确定区间增大了。这说明模拟结果的不确定性不仅跟参数的分布有关系, 而且和蒙特卡罗模拟的次数有很大的关系。随着参数取值范围的进一步扩大, 原有的模拟次数 (这里为 10 000) 已经不能够选取足够多的样本, 因此, 为了更加全面地分析模拟结果的不确定性, 就要相应地增加蒙特卡罗模拟的次数。

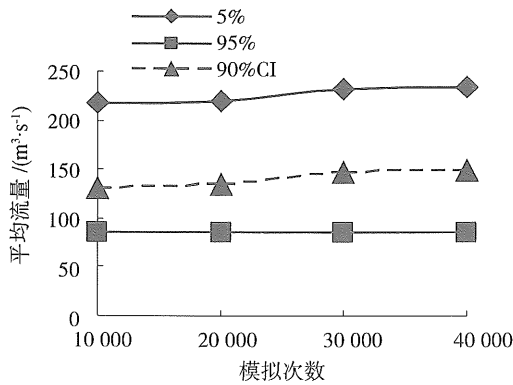


图 3 模拟试验次数与信度水平为 5% 置信区间关系
Fig. 3 The number of sample between C_0 and prediction limits at the significance level of 5%

3.2 参数分布不对称性对结果不确定性的影响

上面分析表明, 参数分布的区间对模拟结果的不确定性有很大的影响。同样, 为了研究参数分布的不对称性对结果不确定性的影响, 我们设定了不同的情景。在参数变差系数 C_v 为 41.94% 时, 也就是在分布区间足够大的情况下, 变动最可能值, 使得参数分布的偏态系数 C_s 从 -0.48 变化到 0.95, 然后进行蒙特卡罗模拟试验, 试验次数为 20 000, 结果如表 3 所示。图 4 显示了参数偏态系数与信度水平为 5% 置信区间关系。从中可以得到参数分布的不对称性对模拟结果的不确定性有一定

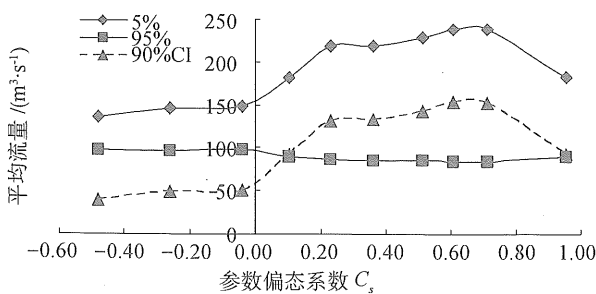


图 4 参数偏态系数与信度水平为 5% 置信区间关系
Fig. 4 The relationship between C_s and prediction limits at the significance level of 5%

的影响; 当偏态系数为 0.61 时 (负偏), 得到的模拟结果的不确定区间最大。与参数分布区间一样, 参数分布的不对称性变化对信度水平为 5% 的不确定度的上边界的影响比较明显, 对下边界的影响则较小。

表 3 参数分布的不对称性对结果不确定性的影响统计

Table 3 Statistics of effects of parameter distribution skewness on output uncertainty

参数偏 序号	态系数 C_s	结果统计值		置信区间			
		$C_v/\%$	C_s	C_k	5%	95%	90% CI
1	-0.48	9.71	0.49	2.14	136.28	97.31	38.97
2	-0.26	8.81	0.36	1.95	145.60	96.24	49.36
3	-0.04	4.34	0.40	3.25	147.43	97.36	50.07
4	0.10	8.14	1.36	5.78	182.42	90.31	92.11
5	0.23	8.86	1.14	4.99	219.02	87.42	131.60
6	0.36	8.58	0.98	4.63	218.52	85.24	133.28
7	0.51	10.07	0.93	4.16	227.66	84.97	142.69
8	0.61	10.33	0.91	4.14	237.83	84.42	153.42
9	0.71	10.89	0.86	3.93	237.12	84.53	152.60
10	0.95	8.14	1.36	5.78	182.42	90.31	92.11

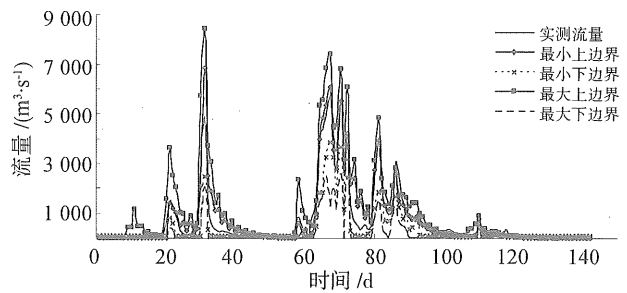


图 5 参数偏态系数为 -0.48 和 0.61 时的信度水平为 5% 的不确定区间 (1981 年 6 月 14 日 - 11 月 2 日)
Fig. 5 Prediction limits for a selected period of 1981 data set (14th June - 2th Nov) when $C_s = -0.48$ and $C_s = -0.61$

根据模拟试验的结果, 当参数偏态系数为 -0.48 和 0.61 时分别得到最小和最大的信度水平为 5% 的不确定区间, 如图 5 所示。另一方面, 从得到的结果, 可以看出模拟得到的流量过程线的上、下边界并不能完全包含实测流量过程线, 总有一些实测流量落在信度水平为 5% 的不确定区间之外, 如图 5 所示。这说明 TOPMODEL 模型并不能完全模拟出该流域的流量过程, 很多研究也已经表明^[6-8], 这是由于模型结构等其他的不确定性因素而引起的。

4 结 论

虽然遥感技术、雷达技术以及同位素示踪技术等逐渐被引入到水文模拟中, 为水文研究提供了更多的信息, 并在一定程度上减少了不确定性, 但是水文循环的复杂非线性特征使得模型的结构和参数都不可避免存在不确定性, 所以必须认真考虑这种不确定性。本文引进了 Beta-PERT 分布研究模型参数分布特征对水文模拟不确定性的影响, 结果表明: ①参数分布的区间对模拟结果的不确定性有很大的影响, 其不仅对模拟结果的离散程度有影响, 而且对模拟结果的对称性和峰度都有一定的影响。随着参数分布区间的增大, 模拟结果的不确定性区间在增大; 随着参数取值范围的进一步扩大, 为了更加全面地分析模拟结果的不确定性, 就要相应地增加蒙特卡罗模拟的次数。②参数分布的不对称性对模拟结果的不确定性有一定的影响; 当偏态系数为 0.61 时 (负偏), 得到的模拟结果的不确定区间最大。③参数分布特征的变化对信度水平为 5% 的不确定度的上边界的影响比较明显, 对下边界的影响则较小。对于复杂环境系统下的水文模拟和预测而言, 造成不确定性的因素有很多, 本文的研究也只是其中的一个方面, 今后仍需要进一步深入的研究。

参考文献:

- [1] 叶守泽, 夏军. 水文科学研究的世纪回眸与展望[J]. 水科学进展, 2002, 13(1): 93 - 104.
- [2] BEVEN K J, BINLEY A. The future of distributed models: model calibration and uncertainty prediction [J]. Hydrological Processes, 1992, 6: 279 - 298.
- [3] FREER J, BEVEN K J, AMBROISE B. Bayesian estimation of uncertainty in runoff prediction and the value of data: an application of the GLUE approach [J]. Water Resources Research, 1996, 32(7): 2161 - 2173.
- [4] BEVEN K J, FREER J. Equifinality, data assimilation, and uncertainty estimation in mechanistic modeling of complex environmental systems using the GLUE methodology [J]. Journal of Hydrology, 2001, 249: 11 - 29.
- [5] 黄国如, 解河海. 基于 GLUE 方法的流域水文模型的不确定性分析 [J]. 华南理工大学学报: 自然科学版, 2007, 35(3): 137 - 142.
- [6] 林凯荣, 陈晓宏, 江涛. 基于 Copula-Glue 的水文模型参数不确定性研究 [J]. 中山大学学报: 自然科学版, 2009, 48(3): 109 - 115.
- [7] 王建军, 徐伟宣, 张勇. 基于 Beta-PERT 分布的单项不良资产定价决策 [J]. 数理统计与管理, 2007, 26(3): 495 - 502.
- [8] 熊立华, 郭生练. 分布式流域水文模型 [M]. 北京: 中国水利水电出版社, 2004.
- [10] 赵永平, 陈永利. 南海 - 热带印度洋海温年际变化与南海夏季风爆发关系的初步分析 [J]. 热带气象学报, 2000, 16(2): 115 - 122.
- [11] 江静, 钱永甫. 南海表面海温异常对南海季风影响的数值模拟 [J]. 南京大学学报: 自然科学, 2002, 38(4): 556 - 564.
- [12] 梁建茵, 吴尚森. 南海西南季风爆发日期及其影响因素 [J]. 大气科学, 2002, 26(6): 829 - 844.
- [13] 黄荣辉, 顾雷, 等. 东亚夏季风爆发和北进的年际变化特征及其与热带西太平洋热状态的关系 [J]. 大气科学, 2005, 29(1): 20 - 36.
- [14] 王丽娟, 王辉, 闫俊岳, 等. 南海海气界面潜热通量的分布特征及其对西南季风爆发影响的初步分析 [J]. 海洋学报, 2008, 30(1): 20 - 30.
- [15] 穆明权, 李崇银. 1998 年南海夏季风的爆发与大气季节内振荡的活动 [J]. 气候与环境研究, 2000, 5(5): 375 - 386.
- [16] 林爱兰. 南海夏季风的低频特征 [J]. 热带气象学报, 1998, 14(2): 113 - 118.
- [17] CHANG C P, CHEN G T J. Tropical circulation associated with southeast monsoon onset and westerly surge over the South China Sea [J]. Mon Wea Rev, 1995, 123: 3254 - 3267.
- [18] DING Y H, LIU Y J. Onset and the evolution of the summer monsoon over the South China Sea during SCS-MEXfield experiment in 1998 [J]. J Meteor Soc, Japan, 2001, 79: 255 - 276.
- [19] CHAN J C L, WANG Y G, XU X L. Dynamic and thermodynamic characteristics associated with the onset of the 1998 South China Sea summer monsoon [J]. J Meteor Soc Japan, 2000, 78: 367 - 380.
- [20] 毛江玉, 吴国雄, 刘屹岷. 季节转换期间副热带高压带形态变异及其机制的研究 II: 亚洲季风区季节转换指数 [J]. 气象学报, 2002, 60(4): 409 - 420.
- [21] 何金海, 丁一汇, 高辉, 等. 南海夏季风建立日期的确定与季风指数 [M]. 北京: 气象出版社, 2001.
- [22] 毛江玉, 吴国雄, 刘屹岷. 季节转换期间副热带高压带形态变异及其机制的研究 I: 副热带高压结构的气候学特征 [J]. 气象学报, 2002, 60(4): 400 - 408.

(上接第 138 页)