

# 基于参数优化的集对分析月径流预测

刘冀,徐刚,彭涛,明波

(三峡大学水利与环境学院,湖北宜昌 443002)

**摘要:**采用集对分析法进行月径流预测时,针对级别划分较多时不易确定差异度分量系数的问题,建立了基于 SCEM-UA 算法优化该系数的月径流预测方法。研究实例表明,本方法能够有效区分集对间不同差异度的影响,优化所得的差异度分量系数是有效的、合理的,能够提高月径流预测精度并发布概率预报。此外,分析表明集对分析预报中的参数不确定性在模型不确定性中占主导地位。

**关键词:**集对分析;径流预测;SCEM-UA;差异度分量系数

**中图分类号:**P338.2

**文献标识码:**A

**文章编号:**1000-0852(2013)01-0008-04

月径流预报因预见期较长,受天文、气象、自然地理特性和人类活动等众多因素的影响,径流变化极其复杂,具有很大的不确定性。由赵克勤<sup>[1]</sup>先生提出的不确定性关系分析理论-集对分析(Set Pair Analysis, SPA)理论为水文中长期预报提供了一条新途径,丰富了现有的水文预报方法,已取得了一些重要研究成果<sup>[2-4]</sup>。应用集对分析法进行径流预报时,首先需划分径流级别,构造径流集合并进行集对分析,然后计算联系度,选取联系度较大的历史相似样本的后续值作为预报结果。上述步骤的关键是如何确定适宜的差异度分量系数用以计算联系度,而目前对于多个差异度分量系数的确定还没有较为完善的方法,针对此问题,本文尝试在一定的优化目标准则下采用 SCEM-UA 算法<sup>[5]</sup>优化率定 SPA 中的差异度分量系数,验证该方法的可行性,并进一步分析月径流的概率区间预报成果。

## 1 模型原理

### 1.1 集对分析

基于哲学中的对立统一和普遍联系的观点,赵克勤提出了集对分析理论,为分析解决水文与水资源中的不确定性问题提供了一条新思路。集对分析的核心思想是对不确定性系统的两个有关联的集合构建集对,再对集对的特性做同一性、差异性、对立性分析,然后用联系度描述集对的同、异、反关系。设有联系的集

合  $A$  和  $B$  构成集对  $H(A, B)$ , 联系度定义为  $\mu_{A-B}$ :

$$\mu_{A-B} = \frac{S}{n} + \frac{F}{n}I + \frac{P}{n}J \quad (1)$$

式中: $n$  为集合特性总数,  $n=S+F+P$ ;  $S$  为同一性的个数;  $F$  为差异性的个数;  $P$  为对立性的个数;  $I$  为差异不确定系数,在  $(-1, 1)$  区间视不同情况取值;  $J$  为对立系数,且  $J \equiv -1$ 。令  $a=S/n$ ,  $b=F/n$ ,  $c=P/n$  分别称为集对  $H(A, B)$  的同一度、差异度和对立度,则式(1)可写为  $\mu_{A-B}=a+bI+cJ$ 。当  $a$  越接近于 1 时,表明这两个集合的关系越倾向于同一;当  $c$  越接近于 1 时,表明这两个集合的关系越倾向于对立;当  $b$  越接近于 1 时,表明这两个集合的关系越倾向于差异。

将式(1)进行拓展,对差异度进行细分可得  $k$  元联系度

$$\mu_{A-B}=a+b_1I_1+b_2I_2+\dots+b_{k-2}I_{k-2}+cJ \quad (2)$$

式中: $a+b_1+b_2+\dots+b_{k-2}+c=1$ ;  $b_1, b_2, \dots, b_{k-2}$  称为差异度分量;  $I_1, I_2, \dots, I_{k-2}$  称为差异不确定分量系数,其取值合理性对联系度影响较大。

对于差异不确定分量系数的确定常采用的方法有经验取值法、均匀取值法<sup>[3]</sup>等。前者需要与研究对象的变化特性和研究人员的经验结合起来进行取值,经验性较强,不易掌握。均匀取值法将差异度分量  $b_k$  均匀划分,方法简单,易于操作,但缺乏理论依据。本文尝试采用优化方法率定参数  $b_k$ , 以期提高径流预报精度。

收稿日期:2012-03-23

基金项目:国家自然科学基金项目(41101511)

作者简介:刘冀(1980-),男,河北保定人,讲师,博士,主要从事水文预报与防洪调度研究。E-mail: liuji@ctgu.edu.cn

## 1.2 SCEM-UA 优化算法

SCEM-UA 算法由 Vrugt<sup>[5]</sup>等人为优化和评估模型参数而提出的模拟搜索优化算法。SCEM-UA 算法是在 SCE-UA 算法的基础上,根据马尔可夫链蒙特卡罗(MCMC)理论,以 Metropolis-Hastings 算法取代 SCE-UA 中的坡降算法(Downhill Simplex Method),估计出最有可能的参数集及后验概率分布,并尽量避免算法陷入局部极值点。SCEM-UA 算法优化步骤如下:

- (1)选择样本群大小参数  $s$  和复合形数目  $q$ ,那么每个复合形中样本数目  $m=s/q$ ;
- (2)产生  $s$  个样本,并根据所设定的目标函数计算每个样本点的后验概率密度;
- (3)将样本点按后验概率密度递减方式排序,存储在数组  $D[1:s, 1:n+1]$ 中,其中  $n$  为估计参数个数,数组中最右一列存储各样本点的后验概率密度;
- (4)初始化  $q$  个并行序列  $S^1, S^2, \dots, S^q$  的起始点,即  $S^k$  为  $D[k, 1:n+1]$ ,此处  $k=1, 2, \dots, q$ 。将  $D[1:s, 1:n+1]$  划分为  $q$  个复合形  $C^1, C^2, \dots, C^q$ ,每个复合形含有  $m$  个样本点,使得第一个复合形包含次序为  $q(j-1)+1$  的点,第二个复合形包含次序为  $q(j-1)+2$  的点等,此外  $j=1, 2, \dots, m$ ;
- (5)选择参数  $L, T, AR_{\min}, c_n$ 。对于每个复合形  $C^k$ ,调用 SEM 算法,运行  $L$  次;
- (6)将所有复合形放入数组  $D[1:s, 1:n+1]$ ,并将各样本点依后验概率密度递减排列;
- (7)检查 Gelman - Rubin (GR)收敛准则,如果符合收敛条件则计算结束,否则转第(4)步。

## 2 计算步骤

已知月径流序列  $\{X_t\}_N$ ,对其进行预测依赖于前  $m$  个相邻历史值  $X_{t-1}, X_{t-2}, \dots, X_{t-m}$ 。首先对径流序列  $X$  划分流量级别得到序列  $\{S_t\}_N$ ,定义  $A_t=(S_{t-m}, S_{t-m+1}, \dots, S_{t-1})$  为一个径流集合,  $X_t$  为  $A_t$  的后续值( $t=m+1, m+2, \dots, N$ )。这样,径流序列  $\{X_t\}_N$  就可以构成  $N-m+1$  个径流集合。要预报后续值  $X_{t+1}$ ,可将第  $N-m+1$  个径流集合分别与其余  $N-m$  个径流集合  $A_t$  组成  $N-m$  个集对,对这  $N-m$  个集对进行同一性、差异性和对立性分析,并计算相应的联系度。然后将联系度按降序排列,从中选择与当前预测状态  $A_{t+1}$  最接近的前  $k$  个径流集合  $A_k$ ,  $k$  为相似集合样本数,建立  $A_k$  与其后续值  $X_k$  的多元线性回归方程,最后将  $A_{t+1}$  代入该多元线性回归方程中即可得到径流预报值  $X_{t+1}$ 。

采用 SCEM-UA 算法对集对分析的差异不确定系数  $I_k$  与相似集合样本数  $k$  进行优化率定,优化目标函数为平均相对误差(MAE)最小化。

$$\text{Min } MAE = \frac{1}{N} \sum_{i=1}^N \frac{|y_i - y_{sim,i}|}{y_i} \times 100\% \quad (3)$$

式中:  $y_i$  为第  $i$  时段实测流量;  $y_{sim,i}$  为第  $i$  时段的预报流量。

此外,利用该算法得到的“异参同效”(不同的参数取值,但相应的 MAE 相差较小)参数可进行月径流的概率预报,从而更好的量化参数不确定性。具体方法为:以 SCEM-UA 算法的采样参数进行月径流模拟预报,计算各组参数的似然值(目标函数值),设定一个临界值,低于该临界值的参数似然值被赋为零。对高于临界值的所有参数似然值重新归一化,然后将预报流量按大小排序,计算累积似然权重,从而估算出一定置信度下的模型参数不确定性区间。第  $t$  时段的模型不确定性区间(UI)按式(4)计算。

$$UI = y_{obs} \pm 1.96RMSE \quad (4)$$

式中:  $RMSE$  为均方根误差,  $RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_t - y_{sim,i})^2}$ , 其余符号同前。

## 3 实例分析

采用某水电站 1958~2005 年共 48a 的实测月径流资料作为研究对象,流域降雨主要集中于每年的 6~9 月,此阶段为丰水期,12 月至次年 4 月为枯水期,该时段径流主要由地下水补给,5 月与 11 月为过渡期。以月为时段,取前 38a 为训练期,余留 10a 为检验期。首先对径流进行分级,考虑到年内径流的分期特性及径流变幅的差异,将各分期(枯水期、过渡期与丰水期)的径流均划分为 10 级,其中枯水期级别区间为  $50\text{m}^3/\text{s}$ ,过渡期为  $100\text{m}^3/\text{s}$ ,丰水期为  $500\text{m}^3/\text{s}$ ,若对应级别处于同一状态定义为同,对应级别相差 9 级则定义为相反,其它情况为相异。采用 SCEM-UA 算法对集对分析的差异不确定系数  $I$  与相似集合样本数  $k$  进行优化率定,算法参数设置为:样本群大小  $s=500$ ,复合型个数  $q=10$ ,迭代次数设为 1500。所得优化后的参数称为优化参数。由于流量分级数目较多,采用一般方法不易确定  $I$  值,为便于比较,采用均匀取值法确定差异度分量系数  $I$ ,所得参数称为经验参数,参数取值见表 1。

表1 优化参数与经验参数取值表  
Table1 The values of optimal parameters and empirical parameters

参数	$I_1$	$I_2$	$I_3$	$I_4$	$I_5$	$I_6$	$I_7$	$I_8$	$k$
优化参数	0.99	0.80	0.79	0.64	0.09	-0.35	-0.44	-0.78	60
经验参数	0.78	0.56	0.33	0.11	-0.11	-0.33	-0.56	-0.78	60

由表1可知,两种方法所得参数差别还是较为明显的,特别是参数 $I_1 \sim I_4$ ,即相异状态差别不大时(相差1~4级),SCEM-UA算法所得到的优化参数相差不大(0.99~0.64),从而便于更加合理的选取相似历史样本进行预报,而当相异状态差别较大时( $I_5 \sim I_8$ ),差异度分量系数变化较大,由0.09减小至-0.78,特别是 $I_4$ 到 $I_5$ ,表明优化算法能够通过改变差异度分量系数达到有效识别相似历史径流集合的目的。对比均匀取值法所得的经验参数,由于是均匀取值,只能线性分配各差异度分量系数,无法有效描绘不同级别差异度的影响。取SCEM-UA算法参数收敛后所得的“异参同效”参数进行分析,各参数的取值范围、优化参数及经验参数见图1。

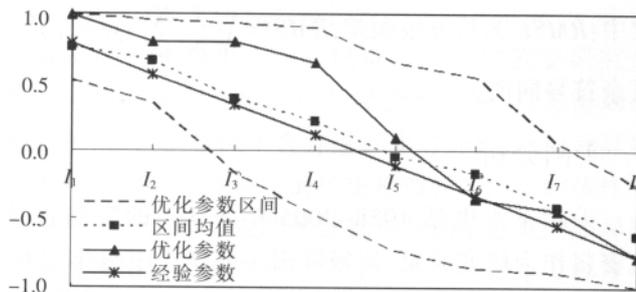


图1 差异度分量系数变化曲线  
Fig.1 The variation curve of difference coefficient

由图1可见,经验参数与区间均值(“异参同效”参数集的相应参数均值)相差较小,最大相差为0.17。“异参同效”参数集中各参数的变幅较大,最大变幅达1.4,表明参数的不确定性较大。通过比较各“异参同效”参数集中的各组参数, $I$ 值均呈单调递减,表明SCEM-UA算法能够有效优化率定差异度分量系数 $I$ 值。表2为两类参数的径流模拟精度统计。

表2 径流模拟精度对比  
Table2 The results of monthly runoff simulation

参数	MRE/%	
	训练期	检验期
优化参数	19.6	16.3
经验参数	24.1	22.1

由表2可知,采用SCEM-UA算法的优化参数进行径流预报的精度明显高于经验参数的预报精度。在训练期,利用优化参数的径流模拟MRE高于经验参数4.5%,检验期则为5.8%,表明利用集对分析法进行径流预报时,采用SCEM-UA算法优化差异度分量系数是必要的,能够较大幅度提高模拟与预报精度。进一步分析精度提高的原因,其一为在一定的优化目标函数下对参数进行率定,必然比采用经验的均匀取值法的预报精度有一定的提高,特别在需确定较多差异度分量系数的情况下,应尽量采用优化方法确定参数;其二为采用SCEM-UA算法时,该算法是一种全局搜索算法,采用拉丁超立方体采样,并根据马尔可夫链自适应的调整转移概率,从而保证参数后验概率密度的连续更新与进化,最终达到识别非劣参数及后验分布的目的。

图2为检验期径流区间预报结果,参数不确定性区间为采用“异参同效”参数集进行径流预报的区间范围。模型不确定性区间按式(4)计算得到。由图2可见,参数不确定性区间并没有完全覆盖全部的实测流量,特别是对于高流量,且模型不确定性区间变幅过大,参数不确定性占全部不确定性的70%左右,但对于高流量该比例相对较小,表明对于高流量预报,单独依赖径流序列自身的单要素预报方法,其精度很难得到进一步提高,需考虑多因素预报。

#### 4 结论

采用集对分析法进行径流预报时,针对流量分级较多情况下差异度分量系数不易确定的问题,本文采用SCEM-UA算法对参数进行优化率定,从而提高径流模拟精度。实例计算表明:

(1) 采用SCEM-UA算法优化差异度分量系数,以此进行径流预报,模拟精度较采用均匀取值法有较大提高,基于优化参数的集对分析预测模型是有效的;

(2) 参数不确定性在径流预报的全部不确定性中占有较大比例,但在高流量预报中该比例显著减小;

(3) 仅采用径流时间序列进行单因素预报的预报

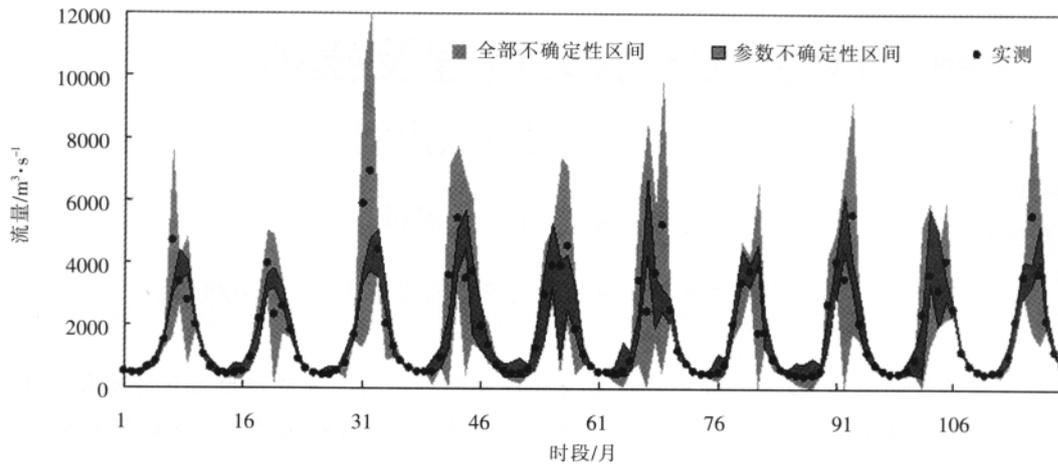


图2 基于优化参数的集对分析径流区间预报结果(检验期)

Fig.2 Hydrograph prediction uncertainty derived using optimal parameters (Validation period)

精度还较低,且预报不确定性较大,应尽量采用具有物理成因联系的因素进行多因素径流预报,以期进一步提高预报精度。

参考文献:

- [1] 赵克勤.集对分析及其初步应用[M].杭州:浙江科学技术出版社,2000.(ZHAO Keqin. Set Pair Analysis and Its Preliminary Application [M].Hangzhou: Zhejiang Science and Technology Press,2000.(in Chinese))
- [2] 邓红霞,李存军,赵太想,等.基于SPA的水文预测模型评估[J].四川大学学报(工程科学版),2006,38(6):34-37.(DENG Hongxia, LI Cunjun, ZHAO Taixiang, et al. Evaluation of hydrologic forecast model based on SPA [J]. Journal of Sichuan University(Engineering Science Edition), 2006, 38(6):34-37. (in Chinese))
- [3] 王文圣,李跃清,金菊良,等.水文水资源集对分析[M].北京:科学出版社,2010.(WANG Wensheng, LI Yueqing, JIN Juliang, et al. Set Pair Analysis for Hydrology and Water Resources Systems[M]. Beijing: Science Press, 2010. (in Chinese))
- [4] 李佳,王黎,马光文,等.基于SPA-ANN耦合模型的年径流预测[J].水力发电学报,2009,28(1):41-44.(LI Jia, WANG Li, MA Guangwen, et al. Annual runoff forecasting based on SPA-ANN coupling model[J]. Journal of Hydroelectric Engineering, 2009, 28(1):41-44. (in Chinese))
- [5] Vrugi J A, Gupta H V, Bouten W, et al. A shuffled complex evolution metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters[J]. Water Resources Research. 2003, 39(8): 18.

## Monthly Runoff Prediction Based on Set Pair Analysis with Parameters Optimization

LIU Ji, XU Gang, PENG Tao, MING Bo

(College of Hydraulic and Environmental Engineering, China Three Gorges University, Yichang 443002, China)

**Abstract:** When predicting monthly runoff with set pair analysis (SPA) method, it will become much difficult to determine the difference degree coefficients as the number of runoff levels increased. To solve this problem, the SCEM-UA algorithm was employed for optimizing the difference degree coefficients of SPA. This method was applied to forecast the monthly runoff. The results show that the proposed method can effectively distinguish the effects of difference degree among set pairs, and the difference degree coefficient optimized by the method is effective and reasonable, the runoff prediction accuracy is improved greatly. Also, the parameter uncertainty occupies the largest proportion in the model uncertainty.

**Key words:** set pair analysis; runoff prediction; SCEM-UA; difference degree coefficient