

基于递推最小二乘改进算法的洪水预报模型研究

周 轶¹, 菅浩然², 李致家¹, 李志龙³

(1. 河海大学水资源环境学院, 江苏 南京 210098 ; 2. 黄河水利职业技术学院, 河南 开封 475003 ;
3. 山东电力工程咨询院, 山东 济南 250013)

摘要 : 由递推最小二乘算法估算出的自回归系数在一定条件下具有最佳的统计特性 , 但在实际应用中 , 这种方法往往难以动态地把握水文现象的动态特性 . 为提高自回归洪水预报模型的精度 , 分别用衰减记忆、有限记忆及 2 种算法相结合的方法对基本的递推最小二乘算法进行改进 , 并利用这几种改进算法对白马寺水文站的实测径流序列进行了模拟演算 . 结果表明 , 这 3 种改进的递推最小二乘算法 , 都可以使自回归洪水预报模型取得较好的预报效果 , 但实际应用时应根据不同预报的侧重点选择相应的算法 .

关键词 : 洪水预报 ; 递推最小二乘算法 ; 衰减记忆 ; 有限记忆

中图分类号 : P338 文献标识码 : A 文章编号 : 1000-1980(2007)01-0077-04

自回归模型作为一种随机数学模型 , 在水文时间序列的预报中得到了广泛应用 . 这种模型可以直观地描述时间序列之间的关系 , 其精度取决于模型的结构与参数 . 为了探讨模型结构与参数的确定方法 , 本文首先通过 AIC 与 BIC 2 种准则的对比分析确定模型结构 , 然后采用 3 种改进的递推最小二乘算法估算模型参数 , 最后利用实测径流序列进行了模拟洪水预报 .

1 洪水预报自回归模型

洪水预报的自回归模型为^[1]

$$y_K = \sum_{i=1}^n a_i y_{K-i} + e_K \tag{1}$$

式中 : y_K —— K 时刻的流量或水位 ; y_{K-i} —— $K-i$ 时刻的流量或水位 ; a_i ——自回归系数 ; n ——模型阶数 ; e_K ——模型噪声或残差 .

在预报中 , 将 K 时刻以前 n 个时段的流量或水位分别乘以对应的回归系数 a_i 即可得到 K 时刻流量或水位的预报值 y_K^S . 残差 e_K 为预报值 y_K^S 与实测值 y_K 的差值 . 显然 , e_K 越小 , 预报精度越高 . 实际运算中往往利用递推最小二乘、卡尔曼滤波这样的递推算法实时地计算出自回归系数 a_i , 本文选用改进后的递推最小二乘算法来估算自回归系数 .

应用自回归模型进行洪水预报前必须首先确定模型的阶数 (结构) , 而自回归阶数通常根据经验来选择 . 这缺乏科学依据 . 因此 , 本文采用最小信息准则 (AIC) 与贝叶斯信息准则 (BIC) 相比较的方法确定了三花间流域中洛河白马寺水文站 1958 ~ 1982 年 5 场代表性次洪的自回归阶数^[2] . 其中 1982072910 场次洪的 AIC 与 BIC 准则函数值的计算结果如图 1 所示 .

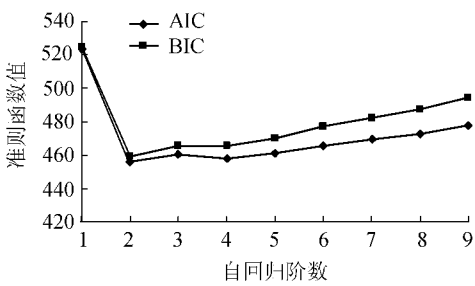


图 1 1982072910 场次洪的 AIC 与 BIC 准则函数值的计算曲线

Fig.1 Calculation curve of AIC and BIC criterion functions for flood No. 1982072910

由图1可以看出,当阶数 $n=2$ 时,1982072910场次洪的AIC与BIC准则函数具有极小值.因此,对于该场次洪,式(1)的阶数为2.利用此方法可以确定出其余4场次洪的自回归阶数.

2 递推最小二乘算法的改进

2.1 基本算法

洪水预报在线参数估计可以采用递推最小二乘算法.递推最小二乘算法(RLS)的基本思想是在有新的系统输入、输出量的情况下,不断对原有系统参数的估计量加以修正,从而计算出新的更能反映模型现状的参数估计量.

递推最小二乘基本算法可表示为^[3-4]

$$\theta_{N+1}^S = \theta_N^S + G_{N+1}(y_{N+1} - \phi_{N+1}^T \theta_N^S) \quad (2)$$

$$G_{N+1} = P_N \phi_{N+1} (1 + \phi_{N+1}^T P_N \phi_{N+1})^{-1} \quad (3)$$

$$P_{N+1} = (1 - G_{N+1} \phi_{N+1}^T) P_N \quad (4)$$

式中: $\theta_{N+1}^S, \theta_N^S$ ——模型的参数向量; y_{N+1} ——模型新的输出向量; ϕ_{N+1} ——模型新的输入向量; P_N, P_{N+1} ——第 N 步与第 $N+1$ 步误差协方差阵; G_{N+1} ——第 $N+1$ 步增益阵.

式(2)~式(4)为递推最小二乘的基本算法.其意义是,可以利用“新息”即预报误差($y_{N+1} - \phi_{N+1}^T \theta_N^S$)对原有的模型回归系数的估计量 θ_N^S 进行修正,并且可以得出改善后的参数估计量 θ_{N+1}^S .在利用递推最小二乘运算前必须先给出模型参数的初值 θ_0^S 及初始误差协方差阵 P_0 .实际运算时可以先选用 m 个数据,根据下列公式直接求出初值,然后再根据式(2)~式(4)进行递推运算.

$$\theta_0^S = (\phi_m^T \phi_m)^{-1} \phi_m^T y_m \quad (5)$$

$$P_0 = (\phi_m^T \phi_m)^{-1} \quad (6)$$

式中: θ_0^S ——模型参数初值向量; P_0 ——初始误差协方差阵; ϕ_m ——模型输入向量; y_m ——模型输出向量.

2.2 基本算法的改进

在基本的递推最小二乘算法中,所有实测数据对自回归模型参数的估算起相同的作用,即每次的预报误差 $e_K = (y_{K+1} - \phi_{K+1}^T \theta_N^S)$ 在回归系数的估计中加权相等.如果系统是线性定常的,那么这种处理方式是合理的.由于水文系统是非线性、时变系统,用这种方法就估算不出真正能反映系统特性的参数值.

对于时变系统,越新的数据越能反映当前系统的特征,也就越应对模型当前的参数估计产生较大影响.因此,为了提高参数估计的准确性和预报精度,本文分别对实测数据采用衰减记忆、有限记忆以及这2种算法相结合的方法,对递推最小二乘的基本算法进行了改进.

2.2.1 衰减记忆递推最小二乘算法

基本思想就是强化新数据在系统参数估计中的作用.其具体做法是,给所有实测数据都加上一个权值.加权原则是:距离当前时刻越远的数据,加权越小;距离当前时刻越近的数据,加权越大.这种算法相当于在所有观测数据上加上了一个滑动的指数加权窗.由此可导出衰减记忆递推最小二乘算法^[4-7],即

$$\theta_{N+1}^S = \theta_N^S + G_{N+1}(y_{N+1} - \phi_{N+1}^T \theta_N^S) \quad (7)$$

$$G_{N+1} = P_N \phi_{N+1} (\mu + \phi_{N+1}^T P_N \phi_{N+1})^{-1} \quad (0 < \mu < 1) \quad (8)$$

$$P_{N+1} = (1 - G_{N+1} \phi_{N+1}^T) P_N / \mu \quad (0 < \mu < 1) \quad (9)$$

式中: μ 为衰减因子,是一个介于0与1之间的正常数;其他符号含义同前.

可以发现,当衰减因子 $\mu=1$ 时,式(7)~式(9)就转变为基本型的递推最小二乘法.由于衰减因子的有机融入,该算法有了更大的适用范围与灵活性,但也带来了一些问题.选择较小的 μ 值时,算法能够较好地把握系统的时变特性,但此时算法平滑能力差;而选择较大的 μ 值时,算法平滑能力较强,但不能有效地跟踪系统的时变特性.在实际应用中,通常采用试错法来筛选衰减因子.一般来说,对于水文变量的预报问题,衰减因子 μ 的取值在0.9~1.0之间.

2.2.2 有限记忆递推最小二乘算法

无论是基本型的递推最小二乘算法还是它的衰减记忆改进形式,都属于无限增长记忆模式,即在实际演算中,这2种算法都利用了当前时刻之前的所有实测数据.尽管在衰减记忆递推最小二乘算法中利用衰减因

子人为地降低了旧数据对当前参数估计的影响,但过去的数据并没有被彻底“忘掉”。而有限记忆递推最小二乘算法通过设定一个记忆长度 K ,使得每次计算只利用最近的 K 个数据的方法来彻底地消除旧数据的影响^[6-7]。这种算法相当于给所有观测数据都加上一个滑动的矩形窗。

根据上述思路可以导出有限记忆递推最小二乘的具体算法。

- 步骤 1 当 $N \leq K$ 时,按基本的递推最小二乘算法进行模型参数估计,即按照式(2)~式(4)进行演算;
- 步骤 2 当 $N > K$ 时,按下列式子进行推导:

$$G_{K+1} = P_K \phi_{K+1} (1 + \phi_{K+1}^T P_K \phi_{K+1})^{-1}$$

(10)

$$P_{K+1} = (1 - G_{K+1} \phi_{K+1}^T) P_K$$

(11)

$$P_{K+1}^* = P_{K+1} + P_{K+1} \phi_{K+1} (1 - \phi_{K+1}^T P_{K+1} \phi_{K+1})^{-1} \phi_{K+1}^T P_{K+1}$$

(12)

$$\theta_{K+1}^S = \theta_K^S + P_{K+1}^* (\phi_{K+1} \Delta y_{K+1} - \phi_{K+1}^T \Delta y_1)$$

(13)

式中: P_{K+1}^* ——修正后的第 $K+1$ 步误差协方差阵; $\Delta y_1 = y_1 - \phi_1^T \theta_K^S$; $\Delta y_{K+1} = y_{K+1} - \phi_{K+1}^T \theta_K^S$;其他符号含义同前。

步骤 1 和步骤 2 即为记忆长度为 K 的有限记忆递推最小二乘算法。该算法属于有限增长记忆模式的递推算法,减少了运算量与运行时间,因而有利于式(1)在洪水预报中的应用。与衰减因子产生的问题相似:当选择较小的记忆长度时,算法能够准确反映新息对参数估计产生的影响,但易产生震荡;当记忆长度较大时,算法平滑,但对系统的时变特性难以把握。

2.2.3 衰减记忆与有限记忆相结合的递推最小二乘算法

从以上 2 种算法的推导过程可以看出:尽管衰减因子递推最小二乘算法能够体现新数据对参数估计的作用,但是它不能完全消除旧数据的影响,而有限记忆递推最小二乘算法虽然能彻底消除旧数据的影响,但在记忆长度内所有数据加权相等。将这 2 种算法相结合,在算法中同时加入衰减因子与记忆长度,可以有机地综合这 2 种算法的优点^[8]。因此,对原有有限记忆递推最小二乘算法进行改进,即在算法中加入衰减因子 μ ,可得

$$G_{K+1} = P_K \phi_{K+1} (\mu + \phi_{K+1}^T P_K \phi_{K+1})^{-1}$$

(0 < μ < 1)

(14)

$$P_{K+1} = (1 - G_{K+1} \phi_{K+1}^T) P_K / \mu$$

(0 < μ < 1)

(15)

这种改进后的算法即为衰减记忆与有限记忆相结合的递推最小二乘算法。该算法相当于给所有观测数据都加上了一个有限宽度的滑动的指数窗。

3 实例计算

分别利用上述 3 种递推最小二乘算法的改进算法对三花间流域中洛河上的白马寺水文站 1983~1996 年共 6 场洪水进行了模拟演算,结果如表 1、图 2 所示。

表 1 3 种递推最小二乘改进算法的演算结果比较

Table 1 Comparison of calculated results of three improved RLS procedures

洪号	流量相对误差/%			峰值相对误差/%			峰现时间误差/h			确定性系数		
	方法 1	方法 2	方法 3	方法 1	方法 2	方法 3	方法 1	方法 2	方法 3	方法 1	方法 2	方法 3
1983072810	-1.03	-1.43	-1.4	8.99	-3.67	-3.65	7	-1	-1	0.97	0.98	0.98
1984070410	-0.46	2.3	2.02	0.93	1.37	0.15	3	-2	-3	1	0.99	0.99
1985091410	-0.42	0.74	0.51	1.51	-1.39	-1.68	3	3	3	1	1	1
1988080810	-2.75	3.93	0.74	-1.91	0.63	0.63	-1	-1	-1	0.98	0.98	0.96
1989070910	-0.77	-0.45	1.94	4.24	2.15	2.15	1	-1	-1	1	0.90	0.99
1996080210	-7.29	4.68	4.54	-7.59	0.23	0.23	1	-4	-4	0.98	0.97	0.97

注:方法 1 为衰减记忆递推最小二乘算法;方法 2 为有限记忆最小二乘算法;方法 3 为衰减记忆与有限记忆相结合的递推最小二乘算法。

从流量误差的比较来看,方法 3 优于方法 1,方法 1 优于方法 2;从峰值误差的比较来看,方法 3 优于方法 2,方法 2 优于方法 1;从峰现时间误差的比较来看,方法 2 与方法 3 效果接近,但都优于方法 1;从确定性系数的比较来看,方法 1 优于方法 3,方法 3 优于方法 2。

4 结 论

本文利用 AIC 和 BIC 2 种准则相比较的方法确定了自回归模型的结构,采用衰减记忆递推最小二乘、有

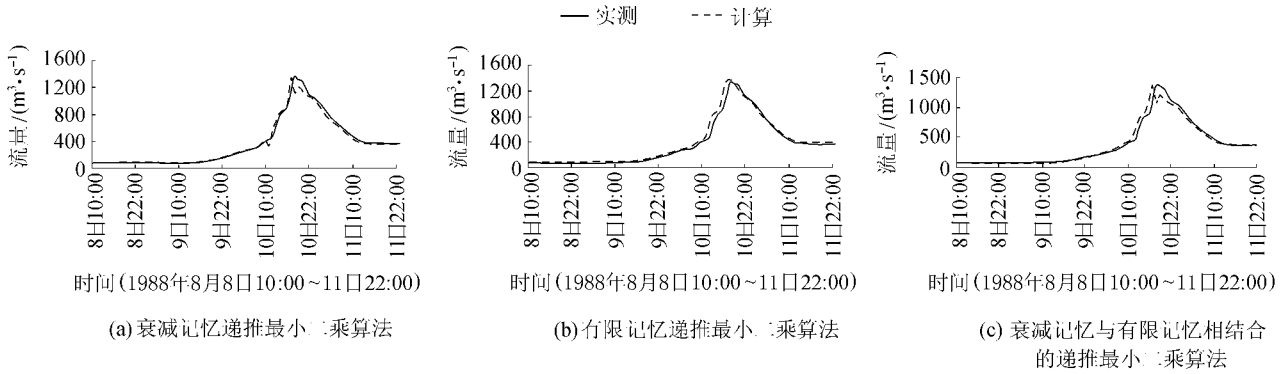


图 2 1988080810 场洪水 3 种算法计算流量过程线与实测流量过程线的比较

Fig.2 Measured and calculated flow graphs of flood No.1988080810 by using the RLS procedures of the faded-memory , the fixed-memory , and their combination

限记忆递推最小二乘以及这 2 种算法相结合的方法实时地估算了自回归系数,并利用上述方法对三花间流域中洛河上的白马寺水文站 1983 ~ 1996 年共 6 场洪水进行了模拟演算.虽然这 3 种算法都可取得较为理想的成果,但这 3 种算法又各有特点,实际应用时应根据不同预报的侧重点选择相应的算法.

参考文献：

[1] 刘东旭,何新林,刘华梅,等.新疆马纳斯河流域洪水的实时预报研究[J].石河子大学学报:自然科学版,2003,7(4):312-314.
[2] 张显扬,董增川,王建群,等.雅鲁藏布江尼洋河流域洪水预报方法研究[J].河海大学学报:自然科学版,2005,33(5):530-533.
[3] 包为民,王浩,赵超,等.AR 模型参数的抗差估计研究[J].河海大学学报:自然科学版,2006,34(3):258-261.
[4] 郭磊,赵英林.基于误差自回归的洪水实时预报校正算法的研究[J].水电能源科学,2002,20(3):25-27.
[5] 梁忠民,董增川,王建群,等.西辽河实时洪水统计预报模型[J].水利水电技术,2004,35(8):8-10.
[6] 徐宁寿.随机信号估计与系统控制[M].北京:北京工业大学出版社,2001:182-248-252.
[7] 王秀峰,卢桂章.系统建模与辨识[M].北京:电子工业出版社,2004:27-37.
[8] 刘整社.时变参数的一种快速跟踪最小二乘估计算法[J].控制理论与应用,1995,12(2):260-264.

Flood forecasting model based on improved recursive least square method

ZHOU Yi¹, JIAN Hao-ran², LI Zhi-jia¹, LI Zhi-long³

(1. College of Water Resources and Environment , Hohai University , Nanjing 210098 , China ;
2. Yellow River Conservancy Technical Institute , Kaifeng 475003 , China ;
3. Shandong Electric Power Consulting Institute , Ji ' nan 250013 , China)

Abstract :The auto-regression (AR) flood forecasting model parameters estimated by the recursive least square (RLS) method have optimum statistical characteristics under certain condition ; however , it is difficult for the method to reflect the dynamic character of hydrological phenomenon in practical application . For improvement of the precision of the AR model , the basic RLS procedure was improved , and three improved forms , i . e . the faded-memory RLS procedure , the fixed-memory RLS procedure , and a combined form of the above two procedures , were deduced . The three improved RLS procedures were applied to simulate the measured runoff series of the Baimasi hydrological station . The results show that the accuracy of flood forecasting by the AR model based on three improved RLS procedures is satisfactory , but rational procedure should be selected according to the practical situation .

Key words :flood forecasting ; recursive least square method ; faded-memory ; fixed-memory