

AR 模式误差修正方程参数抗差估计

瞿思敏,包为民,石朋,赵超

(河海大学水资源环境学院,江苏 南京 210098)

摘要 把抗差理论引入自回归(Auto-Regressive,简称 AR)模型的参数估计中,利用抗差系统具有的抗差能力,可以阻止非正常因素进入系统,保证洪水预报精度.为此,首先介绍几种常用估计方案(Huber 估计、IGG 估计)的抗差特征函数,然后与传统的最小二乘法进行比较,最后给出算例.算例表明:对于正常观测值,采用 Huber 法、IGG 法及 LSM 法均能取得比较满意的结果,如观测值中存在粗差(非正常值),则用 LSM 法估计的结果就很不合理,而用 Huber 法及 IGG 法能取得较好结果.

关键词 洪水预报;抗差估计;AR 模型;最小二乘法

中图分类号:P338 文献标识码:A 文章编号:1000-1980(2003)05-0497-04

实时洪水预报,正常情况下预报精度较高,能满足防汛要求,若受实测资料误差、人类活动等非自然因素的影响,洪水预报精度会大大降低,或正常洪水预报的环境会被破坏.越是发大洪水的时候,越容易产生非正常因素.因此,如何有效地抗御各种实时误差的影响,使预测系统遇非正常因素也能保证正常作业和洪水预报的精度,成为实时洪水预报重要的研究课题.

传统的处理方法是,根据模型的计算误差来实时修正估计结果、模型参数或模型输入等.这些方法的共同点是,非正常因素都毫无阻挡地进入了预报系统,且对计算结果产生了影响,进而影响计算误差.如按这些计算误差来进行修正,则修正效果不会很理想.

本文提出的 AR 模式误差修正方程参数的动态抗差估计,就是把发展较成熟的抗差估计理论^[1]引入 AR 模型的参数估计中,利用抗差系统具有的抗差能力,使许多严重的非正常因素误差影响不可能进入系统.这样,可以减少非正常因素的影响,大大提高系统的稳定性和洪水预报的精度.

1 最小二乘法

设有一组相互独立的观测值 $\{l_i\} (i = 1, 2, \dots, n)$,观测值的权为 $\{p_i\}$,观测方程为

$$l + \Delta = Ax \tag{1}$$

相应的误差方程为

$$v = A\hat{x} - l \tag{2}$$

式中: l —— $n \times 1$ 阶观测向量; A —— $n \times m$ 阶系数矩阵; $\hat{x}$ —— $m \times 1$ 维未知向量, $\hat{x} = (x_1, x_2, \dots, x_m)^T$; Δ —— $n \times 1$ 阶观测误差向量; v ——观测值余差向量.最小二乘估计的准则函数为

$$\sum_{i=1}^n p_i v_i^2 = \min \tag{3}$$

求解 x 的最小二乘估值 $\hat{x}$ 的法方程式为

$$A^T P A \hat{x} - A^T P l = 0 \tag{4}$$

$$\hat{x} = (A^T P A)^{-1} A^T P l \tag{5}$$

其中 P 是对角元素为 p_i 的权矩阵.

2 M估计

所谓抗差估计,就是在粗差不可避免的情况下选择适当的估计方法使未知量估值尽可能地减少粗差的影响,以得出正常模式下的最佳估值.其最常用的方法就是M估计,即极大似然估计.

2.1 等价权

设有观测子样 $\{l_i\}$ 相互独立,观测值的权为 $\{p_i\}$ ($i=1, 2, \dots, n$).化为单位权误差,其经验概率密度可写为

$$f_n(l) = \frac{\sum_{i=1}^n p_i \delta(l - l_i)}{\sum_{i=1}^n p_i} \quad (6)$$

式中 $\delta(l - l_i)$ 为密度集中于 l_i 的Dirac函数.

M估计通过观测值 $\{l_i\}$ 求参数 $\{x_j\}$ 的估值($j=1, 2, \dots, m$),余差为 $\{v_i\}$.其准则函数为

$$\int \rho(v) f(l) dl = \min \quad (7)$$

$$\text{或} \quad \sum_{i=1}^n p_i \left[\frac{\partial}{\partial v_i} \rho(v) \right] \frac{\partial v_i}{\partial x_j} = \sum_{i=1}^n p_i \psi(v) a_{ij} = 0 \quad (j=1, 2, \dots, m) \quad (8)$$

式中 $\rho(\cdot)$ ——适当选择的极值函数; $\frac{\partial}{\partial v_i} \rho(v) = \psi(v)$.把 m 个 x_j 相应的方程联立,并记

$$A = \frac{\partial v}{\partial x} = \begin{bmatrix} \frac{\partial v_1}{\partial x_1} & \cdots & \frac{\partial v_1}{\partial x_m} \\ \vdots & & \vdots \\ \frac{\partial v_n}{\partial x_1} & \cdots & \frac{\partial v_n}{\partial x_m} \end{bmatrix}$$

$$\bar{p}_i = p_i \left(\frac{\partial \rho}{\partial v} \frac{1}{v} \right)_i = p_i \frac{\psi(v)}{v_i} = p_i \omega_i \quad (9)$$

称 $\bar{p}_i$ 为等价权, ω_i 为权因子.于是求极值问题可转化为求解观测方程组

$$A\hat{x} = l + v \quad (10)$$

式(8)整理后可表示成

$$A^T \bar{P} v = 0 \quad (11)$$

由此得到抗差最小二乘估值

$$\hat{x} = (A^T \bar{P} A)^{-1} A^T \bar{P} l \quad (12)$$

由于观测值相互独立,故 $\bar{P}$ 为对角阵,其对角元素为 $p_i \omega_i$.

2.2 几种常用估计方法

a. 最小二乘法(LSM)

$$\omega(u) = 1 \quad (13)$$

b. Huber法

$$\omega(u) = \begin{cases} 1 & |u| \leq k \\ k \cdot \text{sign}(u)/u & |u| > k \end{cases} \quad (14)$$

c. IGG法

$$\omega(u) = \begin{cases} 1 & |u| \leq 1.5 \\ k/|u| & 1.5 < |u| \leq 2.5 \\ 0 & |u| > 2.5 \end{cases} \quad (15)$$

2.3 算法

a. 取权函数 $\omega_i^{(0)}=1$,用最小二乘法求系数,然后计算各点残差 $e_i^{(0)}$;

b. 根据残差的不同取值范围,按不同估计方法赋予不同的权,然后按式(12)重新计算系数,再重新计算

- 各点残差 $e_i^{(1)}$;
- c. 重复步骤 b ,如此不断地计算 e_i ;
- d. 选取适当的精度控制指标 ε ,当 $|e_i^{(j+1)} - e_i^{(j)}| \leq \varepsilon$ 时 ,对所有的 i 均成立 ,迭代结束 ,输出结果.

3 AR 模型

把误差系列用一个 p 阶自回归模型来模拟 :

$$\varepsilon_t = C_1\varepsilon_{t-1} + C_2\varepsilon_{t-2} + \cdots + C_p\varepsilon_{t-p} + \xi_t \tag{16}$$

式中 : ε_t —— t 时刻的模型误差 ($\varepsilon_t = Q_t - QC_t$, Q_t 为实测流量 , QC_t 为计算流量) ; ξ_t —— t 时刻的残差 .将误差系列 $\varepsilon_1 , \varepsilon_2 , \cdots , \varepsilon_t$ 代入自回归式中 ,用向量矩阵表示可得最小二乘估计^[2] :

$$C = (X^T X)^{-1} X^T Y \tag{17}$$

$$Y = \begin{bmatrix} \varepsilon_{p+1} \\ \varepsilon_{p+2} \\ \vdots \\ \varepsilon_t \end{bmatrix} \quad X = \begin{bmatrix} \varepsilon_p & \varepsilon_{p-1} & \cdots & \varepsilon_1 \\ \varepsilon_{p+1} & \varepsilon_p & \cdots & \varepsilon_2 \\ \vdots & \vdots & \ddots & \vdots \\ \varepsilon_{t-1} & \varepsilon_{t-2} & \cdots & \varepsilon_{t-p} \end{bmatrix} \quad C = \begin{bmatrix} C_1 \\ C_2 \\ \vdots \\ C_p \end{bmatrix}$$

4 算 例

4.1 计算实例

假设一理想系统 ,采用马斯京根法^[3]由上断面入流推求下断面出流 .如已知上断面入流 ,河段长 , K , Δt , x ,则可以算出 C_0 , C_1 , C_2 ,再根据式 $Q_2 = C_0 I_2 + C_1 I_1 + C_2 Q_1$ (18) 算出理想的出流系列 $\{Q_i\}$ ($i = 1, 2, \cdots, n$) ,然后分别在此出流系列上加上服从正态分布 $N(0, 10^2)$ 的随机误差和粗差 ,采用上述几种方法率定 C_0 , C_1 , C_2 ,并与实际值进行比较.

现以长江上游干流寸滩至万县河段为例 ,选取 1999 年 12 场洪水进行计算 .该河段长为 321 km , $K = 33$ h , $\Delta t = 6$ h , $x = 0.46$,求得 $C_0 = -0.560$, $C_1 = 0.844$, $C_2 = 0.716$.据此对 990612 次洪水及所有洪水加上正态分布的随机误差以及加上粗差的参数进行了率定 ,结果见表 1.

表 1 参数估值及比较

Table 1 Estimation and comparison of parameters

方法	990612 次洪水参数估值						所有洪水参数估值平均值					
	加正态分布误差			加粗差			加正态分布误差			加粗差		
	C_0	C_1	C_2	C_0	C_1	C_2	C_0	C_1	C_2	C_0	C_1	C_2
LSM	-0.563	0.846	0.716	-0.864	1.275	0.591	-0.562	0.846	0.716	-0.763	1.234	0.532
Huber	-0.563	0.846	0.716	-0.560	0.844	0.716	-0.562	0.846	0.716	-0.560	0.844	0.716
IGG	-0.563	0.846	0.716	-0.561	0.844	0.716	-0.562	0.846	0.716	-0.589	0.901	0.687

又从危水水库流域洪水数据库中选取洪峰 ,前期土湿和降雨相近的 2 场洪水即 740512 和 750608 ,采用 2 阶 AR 模型率定 740512 参数(见表 2) ,然后用该参数对 750608 次洪水进行校正 ,结果见表 2.

表 2 危水水库流域洪水率定及校正结果

Table 2 Results of calibration and correction of flood in Weishui Reservoir basin

方法	正常值			含粗差			原模型效果	
	率定的参数		校正后效果	率定的参数		校正后效果		
	C_1	C_2		C_1	C_2			
LSM	1.601	-0.711	95.19	0.146	0.119	80.79	74.56	
Huber	1.706	-0.718	96.49	1.869	-0.875	96.31	74.56	
IGG	1.691	-0.726	96.29	1.167	-0.386	93.19	74.56	

注 (a) 原模型效果 = $(1 - EOS/QS) \times 100\%$;校正后效果 = $(1 - EES/QS) \times 100\%$. 其中 : $EOS = \sum_{i=1}^n EO(i)$; $QS = \sum_{i=1}^n Q(i)$; $EES = \sum_{i=1}^n EE(i)$ (i 为系列长度) . (b) $EO(i)$ 为原误差系列 , $EO(i) = Q(i) - QC(i)$; $EE(i)$ 为校正误差系列 , $EE(i) = QC(i) - QCC(i)$. 其中 $Q(i)$, $QC(i)$, $QCC(i)$ 分别为实测、计算和校正后的流量系列.

4.2 算例分析

从表 1 中可以看出 :如果误差服从正态分布 ,传统最小二乘法结果与抗差估计方法结果一致 ,并且与真值相差不大 ,如果观测流量中出现粗差 ,则采用抗差估计方法能取得比较令人满意的结果 .

从表 2 可以看出 :用 Huber 法及 IGG 法率定的参数比较稳定可靠 ,而用 LSM 法率定的参数变化较大 ,容易受误差的影响 ,抗差性较差 ;当观测值含有粗差时 ,率定的参数很不可靠 ,用于 AR 模型实时修正效果很差 .

5 结 论

对于实时洪水预报系统 ,AR 模型误差修正方程的参数估计 ,需要排除的干扰是实测流量观测误差 ,特别是对于用实测水位反推的入库流量作为误差修正依据的情形 ,如果不排除反推入库流量的“波动” ,那么误差修正 ,就会越修正结果越差 .

本文算例表明 :对于 AR 模型误差修正方程的参数估计 ,采用具有抗差性能的 Huber 估计法和 IGG 估计法能得到比较稳定、可靠的参数 ,校正效果也很好 .

参考文献 :

[1] 周江文 , 杨元喜 , 黄幼才 , 等 . 抗差最小二乘法 [M] . 武汉 : 华中理工大学出版社 , 1997 . 104—119 .
[2] 华东水利学院 . 水文学的概率统计基础 [M] . 北京 : 水利出版社 , 1981 . 309—331 .
[3] 庄一鸽 , 林三益 . 水文预报 [M] . 北京 : 水利电力出版社 , 1986 . 11—24 .

Robust estimation of parameters for AR model error correction equation

QU Si-min , BAO Wei-min , SHI Peng , ZHAO Chao

(College of Water Resources and Environment , Hohai Univ . , Nanjing 210098 , China)

Abstract :The theory of robust estimation is introduced into parameter estimation of the auto-regressive model . The robust system can prevent abnormal factors from entering the flood system , so as to ensure the accuracy of flood forecast . Furthermore , some estimation methods commonly used , including the Huber estimation and IGG estimation , are introduced and compared with the Least Square Method (LSM) . Examples show that , with normal observed data , the results obtained by the three methods are satisfactory , however , if the observed data are abnormal , the estimated result by LSM is not satisfactory , while the results from the other two methods are still satisfactory .

Key words :flood forecast ; robust estimation ; AR model ; LSM