半参数预报方法在水位预报中的应用*

夏学文
(湖南纺织专科学校·湘潭,411104)

摘要:本文用ARMA序列的半参数预报方法对湘江水位进行预报,结果表明,这一方法具有好的应用价值.

关键词:ARMA序列;新息定理;半参数预报

1 前　言

江河的水位预报在水文分析中起着重要作用,它对防洪抗旱及工农业生产等都有大的影响.怎样才能准确地预测未来水位是水文研究的一大课题.一般,水文站按月测量水位数据,即每个月测取一个水位数值.这样得到的数据列是一时间序列,从而可用时间序列分析来研究数据中包含的信息.从数据序列的特征来看,序列包含有线性趋势和周期趋势等确定性成分,因此,首先要提取出确定性趋势,剩下的随机成分用时序模型处理而进行预报.

周知,用ARMA序列模型进行预报,其计算比较繁琐,即使是新息预报[1]也免不了这一缺点.ARMA序列的半参数预报[2]是在新息定理的基础上,避开对ARMA模型中MA参数作估计这一复杂的工作,在只须知道AR参数的情况下直接得到序列的预报值,使计算量大为简化同时又保证有一定的精度.本文把这一方法应用于湘江水位预报中,取得了好的效果.

2 基本理论

设 $\{x_t\}$ 是平稳ARMA序列,即

$$x_t - \sum_{i=1}^{p}\varphi_i x_{t-i} = a_t - \sum_{i=1}^{q}\theta_i a_{t-i}. \tag{1}$$

其中 a_t 是均值为零方差为 σ_a^2 的正态白噪声.

定义

$$X_k \triangleq \{x : x = \sum_{i=-\infty}^{k} c_i x_i,\quad c_i \in \mathrm{R}_n,\quad \sum_{i=-\infty}^{k} c_i^2 < \infty\}, \tag{2}$$

$$X_k^1 \triangleq \{x : x = \sum_{i=1}^{k} c_i x_i,\quad c_i \in \mathrm{R}\}, \tag{3}$$

$$\hat{X}_k(m) \triangleq \{E(x_{k+m} \mid X_k)\}, \tag{4}$$

$$\hat{X}_{k+m,k} \triangleq \{E(x_{k+m} \mid X_k^1)\}. \tag{5}$$

$\hat{x}_{k+m,k}$ 的一步预报误差为 $\varepsilon_k \triangleq x_k - \hat{x}_{k,k-1}$, ε_k 被称为"新息", $\hat{x}_k(m)$ 和 $\hat{x}_{k+m,k}$ 分别表示用无限样

* 本文得到国家自然科学基金资助.
本文于1990年11月5日收到. 1992年11月25日收到修改稿.

本序列$\{x_k,x_{k-1},\cdots\}$和有限样本序列$\{x_k,x_{k-1},\cdots,x_1\}$对未来时刻的值$x_{k+m}$所做的线性最小方差预报的估计.

新息定理 设

$$y_k=\begin{cases}x_k, & k\leqslant Q,\\ x_k-\sum\limits_{i=1}^{p}\varphi_i x_{k-i}, & k>Q.\end{cases}\tag{6}$$

其中$Q=\max(p,q)$. 则

$$\varepsilon_k=\begin{cases}y_1=x_1, & k=1,\\ y_k-\sum\limits_{j=1}^{k-1}J_{kj}\varepsilon_j, & 1<k\leqslant Q,\\ y_k-\sum\limits_{j=k-p}^{k-1}J_{kj}\varepsilon_j, & k>Q.\end{cases}\tag{7}$$

其中

$$J_{kj}=\begin{cases}\left(E(y_ky_j)-\sum\limits_{i=1}^{j-1}J_{ki}J_{ji}E(\varepsilon_i^2)\right)/E(\varepsilon_j^2), & 1<k\leqslant Q,\ 1\leqslant j\leqslant k-1,\\ \left(E(y_ky_j)-\sum\limits_{i=k-q}^{j-1}J_{ki}J_{ji}E(\varepsilon_i^2)\right)/E(\varepsilon_j^2), & k>Q,\ k-q\leqslant j\leqslant k-1,\end{cases}\tag{8}$$

$$E(\varepsilon_k^2)=\begin{cases}E(y_1^2), & k=1,\\ E(y_k^2)-\sum\limits_{j=1}^{k-1}J_{kj}^2E(\varepsilon_j^2), & 1<k\leqslant Q,\\ E(y_k^2)-\sum\limits_{j=k-q}^{k-1}J_{kj}^2E(\varepsilon_j^2), & k>Q.\end{cases}\tag{9}$$

证 见[3].

m步的新息预报公式为

$$\hat{x}_{k+m,k}=\begin{cases}\sum\limits_{j=1}^{k}J_{k+m,j}\varepsilon_j, & k+m\leqslant Q,\\ \sum\limits_{j=1}^{p}\varphi_j\hat{x}_{k+m-j,k}+\sum\limits_{j=k+m-q}^{k}J_{k+m,j}\varepsilon_j, & k+m>Q.\end{cases}\tag{10}$$

在公式(10)中,要估计参数$\{\varphi_i\}$,$\{\theta_i\}$,σ_a^2,计算$E(y_ky_j)$,在此基础上由(7)~(9)式计算J_{kj},$E(\varepsilon_k^2)$,$\{\varepsilon_k\}$,最后由(10)得到预报值,在(10)中,$\{\varphi_i\}$可由 Yule-Walker 方程求得. 至于$\{\varepsilon_k\}$,由新息定理知关键是在$\{\varphi_i\}$,$\{\theta_i\}$和σ_a^2的基础上估计$E(y_ky_j)$. 由式(6)知$E(y_ky_j)$又可表为如下形式:

$$E(y_k y_j)=\begin{cases} r(k-j), & j\leqslant k\leqslant Q,\\ r(k-j)-\sum\limits_{s=1}^{p}\varphi_s r(k-j-s), & j\leqslant Q,k>Q,\\ \sum\limits_{s=0}^{p}\sum\limits_{t=0}^{p}\varphi_s\varphi_t r(k-j-(s-t)), & k\geqslant j>Q. \end{cases} \tag{11}$$

这里 $\varphi_0=-1,r(h)=E(x_n x_{n+h})$，即 $E(y_k y_j)$ 可在 $\{\varphi_i\}$ 和 $r(h)$ 的基础上估计出来. 显然 $r(h)$ 的估计比 $\{\theta_i\}$ 的估计简单得多，由此得到一种只要用到 $\{\varphi_i\}$ 和 $r(h)$ 的新的预报方法，即 ARMA 序列的半参数预报，具体步骤如下：

i) 已知时间序列 $\{x_i\}$，计算 $\{x_i\}$ 的样本自协方差函数如下

$$\hat{r}(h)=\frac{1}{N}\sum_{i=1}^{N-k}x_i x_{i+h},\quad h=1,2,\cdots,N.$$

ii) 用 Yule-Walker 方程估计 AR 参数 $\{\varphi_i\}$.

iii) 将 $\{\hat{\varphi}_i\}$ 和 $\hat{r}(h)$ 代入(11)式，并估计 $E(y_k y_j)$.

iv) 用式(7)～(9)估计 $\{\varepsilon_k\}$.

v) 由(10)得到预报值.

3 湘江水位预报分析

根据湘江水位数据，从中取 120 个(1958 年 1 月至 1968 年 12 月每月一个数据)，先去掉数据的趋势项，然后将余量(随机成分)用 ARMA(p,q) 模型拟合而进行半参数预报. 具体如下：

$$w_t=\beta_0+\beta_1 t+y_t,$$

$$y_t=\sum_{j=1}^{p}A_j\cos(\omega_j t+\Phi_j)+x_t,\quad (t=1,2,\cdots,120). \tag{12}$$

此处 w_t 是实际测量得到的序列，$\beta_0,\beta_1,A_j,\omega_j,\Phi_j$ 是待定常数，x_t 是随机序列. β_0,β_1 可用线性回归方法求得. 为了求出(12)式中的未知参数 A_j,ω_j,Φ_j 并估计出其阶数 p，将其改写为[4]

$$y_t=\sum_{j=1}^{2p}a_j e^{it\lambda_j}+x_t.$$

其中

$$\lambda_j=\begin{cases}\omega_j, & (j=1,2,\cdots,p),\\ -\omega_j, & (j=p+1,\cdots,2p),\end{cases}\qquad a_j=\begin{cases}\frac{1}{2}A_j e^{i\Phi_j}, & (j=1,2,\cdots,p),\\ -\frac{1}{2}A_{j-p}e^{-i\Phi_{j-p}}, & (j=p+1,\cdots,2p).\end{cases}$$

故只需计算值 p,λ_j,a_j. 令

$$J_{y,120}(\lambda)=120^{-31/16}\left|\sum_{j=1}^{120}y_j e^{-i\lambda_j}\right|.$$

因数据个数 $N=120$，再将$[-\pi,\pi]$分成 $2N=(240)$等分，在各分点上计算

$$D(k)=J_{y,120}\left(\frac{k\pi}{120}-\pi\right),\quad k=0,1,2,\cdots,240.$$

取
$$E = 120^{1/16}\frac{1}{240}\sum_{k=0}^{240}D(k).$$
为估计阶数 p，在 $D(k)(k=1,2,\cdots,240)$ 中选取 $D(k)>E$ 的 k，设其总个数为 q，记其全体为 $k_1,k_2,\cdots,k_q$. 令
$$A_j = k_{j+1} - k_j - \left[\frac{\sqrt{120}}{\pi}\right] - 1,\quad j = 1,2,\cdots,q-1,$$
$$A_q = 240 + k_1 - k_q - \left[\frac{\sqrt{120}}{\pi}\right] - 1.$$
则 p 的估计值 $\hat{p}$ 为 $A_j(j=1,2,\cdots,q)$ 中大于零的总数，记 $A_j(j=1,2,\cdots,q)$ 中满足 $A_j>0$ 的全体足标为 $j_1<j_2<\cdots<j_{\hat{p}}$，则 λ_j 的估计值为
$$\hat{\lambda}_j = \frac{\pi}{120}j_{k,\max},\quad j = 1,2,\cdots,\hat{p}.$$
其中 $j_{k,\max}$是$[j_{k-1},j_k](k=1,2,\cdots,\hat{p}$ 中使 $D(k)$取极大值的足标，且设 $j_0=0$，a_j 的估计值为
$$\hat{a}_j = \frac{1}{120}\sum_{t=1}^{120}y_t e^{-it\hat{\lambda}_j},\quad j = 1,2,\cdots,\hat{p}.$$
最后得到 A_j,ω_j,Φ_j 的估计值为 $\hat{\omega}_j=\hat{\lambda}_j$，$\hat{A}_j=2|\hat{a}_j|$，$\hat{\Phi}_j=\arg\hat{a}_j$，$(j=1,2,\cdots,\hat{p})$.

最后，对随机序列$\{x_t\}$进行半参数预报，其预报值与相应的趋势项的值相加后即得$\{w_t\}$的预报值. 表 1 是 1969 年 1 月至 1970 年 12 月的测量值(实际值)与预报值及相对误差表，表中 m 表示预报步数，$k=120$，$w_k(m)$表示 w_{k+m}，$\hat{w}_k(m)$为其预报值. 从表 1 可以看出，相对误差很小，即表明有好的预报效果. 另外，再对 1993 年 1 月至 12 月的水位进行预报，其结果列于表 2，从表 2 可知，1993 年水位变化比较平稳，不会出现大的洪灾旱灾.

表 1 实测值与预报值对照(单位：米)

m	$w_k(m)$	$\hat{w}_k(m)$	$\lvert w-\hat{w}\rvert/w$	m	$w_k(m)$	$\hat{w}_k(m)$	$\lvert w-\hat{w}\rvert/w$
1	28.95	28.90	0.00173	13	30.19	30.43	0.00795
2	29.64	29.13	0.01721	14	29.67	29.95	0.00270
3	31.00	29.82	0.03806	15	30.53	30.87	0.01114
4	33.79	33.13	0.01953	16	30.78	30.93	0.00487
5	31.91	31.08	0.02601	17	32.26	32.81	0.01705
6	34.64	34.60	0.00115	18	31.08	31.25	0.00547
7	35.52	35.51	0.00028	19	32.87	32.96	0.00274
8	32.05	32.02	0.00094	20	32.45	32.77	0.00986
9	31.57	31.41	0.00507	21	30.46	30.12	0.01116
10	29.96	29.97	0.00033	22	30.24	30.05	0.00628
11	29.64	29.71	0.00236	23	29.91	29.62	0.00970
12	29.17	29.85	0.02331	24	29.06	28.37	0.02374

表 2　1993 年月水位预报值

月份	1	2	3	4	5	6	7	8	9	10	11	12
水位	29.01	30.58	31.90	35.77	34.03	35.57	36.12	32.15	31.06	29.8	29.11	28.59

致谢　作者感谢侯振挺教授和谢衷洁教授的关怀指导！

参 考 文 献

[1] 杨位钦，顾岚．时间序列分析与动态数据建模．北京：北京工业学院出版社，1986
[2] 曾忻等．ARMA 序列的半参数预报，控制理论与应用，1990，7(2)：95—101
[3] 安鸿志等．时间序列的分析与应用．北京：科学出版社，1983
[4] He Sha yuan. On Estimating the Hidden Periodicities in Linear Time Series Model. Thinan Japan, Symposium On Statistics, Beijing University press, 1984

The Application of the Semi-parametric Forecasting Method in the Water Level Forecasting

XIA Xuewen
(Hunan Textile Institute • Xiangtan, 411104, PRC)

Abstract: In this paper, the semi-parametric forecasting method for ARMA series is used for the Xiangjiang's water level forecasting. The result shows that with this method, application value is good.

Key words: ARMA series; innovation theorem; semi-parametric forecasting

本文作者简介

夏学文　1964 年生．1985 年毕业于湖南师范大学数学系，获理学学士学位．1989 年毕业于长沙铁道学院科研所，获理学硕士学位．现为湖南纺织专科学校教师．目前主要从事教学及概率统计方面的研究工作，发表论文二十余篇，主要兴趣为时间序列分析的理论与应用