第13卷 第3期　武汉测绘科技大学学报　Vol.13 No.3
1988年 9月　Journal of Wuhan Technical University of Surveying and Mapping　Sep. 1988

应用时间序列方法作大坝变形预报

徐 培 亮

摘　要

本文首先介绍时序分析的三个基本模型——ARMA模型、AR模型和MA模型，以及各模型的统计性质。然后以某大坝1715廊道的激光视准线观测位移值（已利用倒垂观测把相对位移化为绝对位移）为例，着重叙述大坝变形分析的建模过程，得到了一个AR（2）模型并对大坝变形作了预报，结果具有相当好的预报精度。从而说明，时序分析法将是大坝变形分析的一个有力工具。

【关键词】 时间序列；AR模型；回归分析；大坝变形

引　言

大坝变形观测得到了大量的观测数据（如引张线，视准线，正、倒垂观测等），为了认识大坝变形的规律，并对大坝变形进行预报，需要对这些数据进行适当的处理。然而，如果仅对这些数据进行简单的处理是不能满足这一要求的。

时间序列分析是数理统计学的一个重要分支，经过数十年的发展，已逐渐广泛地应用于工农业生产中[1-3]。这是一种处理随时间变化而又相互关联的数据的数学方法，或者说，是一种处理动态数据的参数化时域分析方法。其主要手段是选择恰当的数学模型来近似描述动态数据，通过研究分析，更本质地了解数据的内在结构和复杂特性，以达到控制和预报的目的。大坝变形观测数据也是一种随时间变化的动态数据。因此，有必要探讨时序分析方法在大坝变形分析中的应用前景。

本文主要讨论应用时序分析方法建立大坝位移观测数据的数学模型的过程，并阐述根据所建立的模型进行变形预报的方法。为讨论方便，文中首先简要地介绍时序分析的模型和预报的基本概念。

1　平稳时间序列模型的概念

设 x_1，x_2，…，x_N 是一组均值为零的平稳时间序列观测值，用差分方程

$$x_t-\varphi_1x_{t-1}-\cdots-\varphi_px_{t-p}=a_t-\theta_1a_{t-1}-\cdots-\theta_qa_{t-q} \tag{1}$$

本文1987年9月12日收到.

对这些观测数据作近似描述。其中$\{a_t\}$是白噪声序列。如果$E(a_t^2)=\sigma_a^2$，且当$s>t$时，$E(x_t x_s)=0$，而由实参数φ_1，φ_2，…，φ_p和θ_1，θ_2，…，θ_q，构成方程（p，q为非负整数）：

$$1-\varphi_1 x-\varphi_2 x^2-\cdots-\varphi_p x^p=0, \tag{2}$$

$$1-\theta_1 x-\theta_2 x^2-\cdots-\theta_q x^q=0 \tag{3}$$

无公因子且其根全在单位圆外。

称（1）式为$\{x_t\}$的P阶自回归-q阶滑动平均混合模型，并简称为ARMA(p，q) 模型。

若$q=0$，（1）式成为：

$$x_t-\varphi_1 x_{t-1}-\cdots-\varphi_p x_{t-p}=a_t, \tag{4}$$

称为p阶纯自回归模型，简称为AR(p)模型。

若 $p=0$，（1）式成为：

$$x_t=a_t-\theta_1 a_{t-1}-\cdots-\theta_q a_{t-g}, \tag{5}$$

称为q阶纯滑动平均模型，简称为MA(q) 模型。

可以证明，差分方程（1）的平稳解和有理谱密度之间存在一一对应的关系，用它来拟合具有有理谱的平稳零均值随机序列是行之有效的。同时，ARMA(p，q)模型的形式简单，只要给定p、q和相应的参数φ_i、θ_i和σ_a^2，模型就完全确定。因此，用这种模型对动态数据进行拟合是比较方便的。

根据ARMA模型的结构和性质，在对平稳序列建立模型时，首先对模型的类别进行判别，并对模型参数进行初步估计，然后对模型作检验和改进，最后利用所建模型进行控制和预报。

模型参数的初步估计一般采用矩估计法，而精细估计可以按最小二乘和极大似然法等准则进行。对于自回归模型，当样本量充分大时，矩估计与精细估计的结果几乎没有区别，也可以作为最终的估计。

为了求得AR(p)的参数的矩估计，将式（4）两边同乘$x_{t-k}(k>0)$并取数学期望得：

$$E(x_t x_{t-k})-\varphi_1 E(x_{t-1}x_{t-k})-\cdots-\varphi_p E(x_{t-p}x_{t-k})=0, \tag{6}$$

由式（6）可得方程组

$$\begin{pmatrix} 1 & \rho_1 & \ddots & \rho_{p-1} \\ \rho_1 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \rho_1 \\ \rho_{p-1} & \cdots & \rho_1 & 1 \end{pmatrix}\begin{pmatrix} \varphi_1 \\ \varphi_2 \\ \vdots \\ \varphi_p \end{pmatrix}=\begin{pmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_p \end{pmatrix}. \tag{7}$$

其中

$$r_k=E(x_t x_{t-k})=r_{-k}, \tag{8}$$

$$\rho_k=r_k/r_0, \tag{9}$$

即r_k为x_t的自协方差函数，ρ_k为其自相关函数。

显然，在（7）式中，如果把x_t的自相关函数ρ_k的估计$\hat{\rho}_k$代替ρ_k，则可得 AR(p) 模型的参数估计为

$$\begin{pmatrix} \hat{\varphi}_1 \\ \hat{\varphi}_2 \\ \vdots \\ \hat{\varphi}_p \end{pmatrix} = \begin{pmatrix} 1 & \hat{\rho}_1 & \cdots & \hat{\rho}_{p-1} \\ \hat{\rho}_1 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \hat{\rho}_1 \\ \hat{\rho}_{p-1} & \cdots & \hat{\rho}_1 & 1 \end{pmatrix}^{-1} \begin{pmatrix} \hat{\rho}_1 \\ \hat{\rho}_2 \\ \vdots \\ \hat{\rho}_p \end{pmatrix}. \tag{10}$$

为了了解AR(p)模型的统计特性，改写（7）式为：

$$\begin{pmatrix} 1 & \rho_1 & \cdots & \rho_{k-1} \\ \rho_1 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \rho_1 \\ \rho_{k-1} & \cdots & \rho_1 & 1 \end{pmatrix} \begin{pmatrix} \varphi_{k1} \\ \varphi_{k2} \\ \vdots \\ \varphi_{kk} \end{pmatrix} = \begin{pmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_k \end{pmatrix}, \tag{11}$$

称 $\varphi_{kk}(k\geqslant 1)$为 x_t 的偏相关函数。显然，当 x_t 为AR(p) 序列时，偏相关函数具有截尾性，即当 $k>p$ 时，$\varphi_{kk}=0$；反之亦然。

如果在（11）式中，以$\hat{\rho}_i$代ρ_i，则有

$$\lim_{N\to\infty} E(\hat{\varphi}_{kk}) = 0\,,\quad k>p \tag{12}$$

$$E(\hat{\varphi}^2_{kk}) \approx \frac{1}{N}\,.\quad k>p \tag{13}$$

同理，为了求得MA(q)模型的参数的矩估计，将（5）式两边同时乘以 $x_{t-k}(k\geqslant 0)$，并取数学期望得：

$$r_k = \begin{cases} \sigma_a^2(1+\theta_1^2+\theta_2^2+\cdots+\theta_q^2) & k=0\,, \\ \sigma_a^2(-\theta_k+\theta_1\theta_{k+1}+\cdots+\theta_{q-k}\theta_q) & 1\leqslant k\leqslant q\,, \\ 0 & k>q\,. \end{cases} \tag{14}$$

$$\rho_k = \begin{cases} 1 & k=0\,, \\ \dfrac{-\theta_k+\theta_1\theta_{k+1}+\cdots+\theta_{q-k}\theta_q}{1+\theta_1^2+\theta_2^2+\cdots+\theta_q^2} & 1\leqslant k\leqslant q\,, \\ 0 & k>q\,. \end{cases} \tag{15}$$

由此可见，当 x_t 为MA(q)序列时，自相关函数 r_k 或 ρ_k 具有截尾性；反之，如果平稳序列的 ρ_k 具有截尾性，则该模型是MA(q)序列。

在（15）式中，以$\hat{\rho}_k$代替ρ_k，可解得MA(q)模型参数 θ_j 的矩估计$\hat{\theta}_i$。且有：

$$\lim_{N\to\infty} E(\hat{\rho}_k) = \begin{cases} \rho_k & 1\leqslant k\leqslant q\,, \\ 0 & k>q\,, \end{cases} \tag{16}$$

$$E(\hat{r}^2_k) \approx \frac{1}{N}\left(1+2\sum_{l=1}^{q}\rho_l^2\right)r_0^2 \quad k>q. \tag{17}$$

2 大坝位移数据的模型建立

下面以一个实例来说明对大坝位移观测建立模型的方法。数据采用 1978 年 3 月 30 日至1982年11月30日的观测结果，每半月观测一次，共 113 个数据。如果数据的观测日期不是每月的15日或30日，则应用内插技术得到。

2.1 数据的预处理

记原始数据为 $y(t)$。另计算 3 组平稳化、零值化数据，即对 $y(t)$ 取一阶差分 $\nabla y(t) = y(t) - y(t-1)$，年周期差分 $\nabla^{24}y(t) = y(t) - y(t-24)$ 和 $\nabla\nabla^{24}y(t) = \nabla^{24}y(t) - \nabla^{24}y(t-1)$，结果列于表 1（只列出部分数据）。

2.2 模型的初步识别

根据 AR、MA 模型的偏相关函数及自相关函数的截尾性质知：对于某一 p，如果在 $\hat{\varphi}_{(p+1)(p+1)}, \cdots, \hat{\varphi}_{(p+M)(p+M)}$ 中，满足 $|\hat{\varphi}_{ii}| \leqslant 1/\sqrt{N}$ 的 i 的总个数达到 68.3% M 或 $|\hat{\varphi}_{ii}| \leqslant 2/\sqrt{N}$ 的 i 的总个数达到95.5%M，则该序列可初判为 p 阶 AR 模型，其中 M 一般取 $N/10$ 或 $\sqrt{N}$；而对于某一 q，如果在 $\hat{\rho}_{(q+1)}, \cdots, \hat{\rho}_{(q+M)}$ 中，满足 $|\hat{\rho}_k| \leqslant 1/\sqrt{N}\left(1+2\sum_{l=1}^{q}\hat{\rho}_l^2\right)^{\frac{1}{2}}$ 的 k 的个数达到68.3%M 或 $|\hat{\rho}_k| \leqslant \frac{2}{\sqrt{N}}\left(1+2\sum_{l=1}^{q}\rho_l^2\right)^{\frac{1}{2}}$ 的 k 的个数达到95.5%M，则该序列可初判为MA(q)模型，否则定为ARMA模型。

因此，对 $y(t)$，$\nabla y(t)$，$\nabla^{24}y(t)$ 和 $\nabla\nabla^{24}y(t)$ 计算自协方差函数 $\hat{r}_k$ 和自相关函数 $\hat{\rho}_k$ 以及偏相关函数 $\hat{\varphi}_{kk}$。$\hat{\varphi}_{kk}$ 采用（34）式计算，而 $\hat{r}_k$ 和 $\hat{\rho}_k$ 的计算式分别为：

$$\hat{r}_k = \frac{1}{N}\sum_{i=1}^{N-k}x(i)x(i+k) \ , \tag{18}$$

$$\hat{\rho}_k = \hat{r}_k / \hat{r}_0 \ . \tag{19}$$

计算结果列于表 2。从该表可见，$y(t)$、$\nabla\nabla^{24}y(t)$ 或 $\nabla y(t)$ 的 $\hat{\rho}_k$ 或 $\hat{\varphi}_{kk}$ 没有截尾的现象，而 $x_t = \nabla^{24}y(t)$ 的偏相关函数呈截尾性，且满足上述百分比的要求。因此，把 $x(t)$ 判为AR(2)模型为：

$$x_t = \varphi_1 x_{t-1} + \varphi_2 x_{t-2} + a_t \ . \tag{20}$$

2.3 模型参数 φ_1，φ_2 的初步估计

利用（10）式对模型（20）的参数 φ_1，φ_2 作矩估计得：

$$x_t = 1.0813x_{t-1} - 0.3536x_{t-2} + a_t \ . \tag{21}$$

2.4 模型的检验

用时间序列方法进行建模是对实际观测序列的一种拟合，因此须对所建模型的拟合优度进行检验，实际上就是要检验 $\{a_t\}$ 是否是白噪声序列。

设 x_t 的样本长度为N_1（对于本例$N_1=113-24$），则由（20）式可得 a_t 的估计为：

$$\hat{a}_t=x_t-1.0813x_{t-1}+0.3536x_{t-2}, \tag{22}$$

其样本长度为 $N_a=N_1-2$。

计算

$$\hat{r}_k(a)=\frac{1}{N_a}\sum_{j=1}^{N_a-k}\hat{a}_j\hat{a}_{j+k}, \tag{23}$$

$$\hat{\rho}_k(a)=\hat{r}_k(a)/\hat{r}_0(a). \tag{24}$$

构造统计量

$$Q=\sum_{i=1}^{k}[\sqrt{N_a}\ \hat{\rho}_i(a)]^2=N_a\sum_{i=1}^{k}\hat{\rho}_i^2(a)=23.99, \tag{25}$$

当原假设成立时，

$$Q\sim\chi^2(k). \tag{26}$$

取$\alpha=0.05$，查表得$Q(17)=\chi^2_{0.05}(17)=27.587>Q$，因此接受原假设，就是说，用（21）式来拟合原观测序列是合适的。

且因二次方程：

$$1-1.0813x+0.3536x^2=0 \tag{27}$$

的根全在单位圆外，从而保证了解的平稳性。

2.5 精估计

对于AR模型而言，其精估计与矩估计差别不大。在本例中，$p=2$，样本长度达到113，因此容易证明两者几乎没有差别，故没有必要作精估计。

2.6 模型的改进

如果在第4步的检验被拒绝，或估计得到的$\hat{\varphi}_i$或$\hat{\theta}_i$构成的多项式方程的根不在单位圆外，则须返回第2步或重新进行数据预处理。

3 应用时序模型预报大塌变形

在（1）式中记

$$B^k x_t=x_{t-k}, \tag{28}$$

则（1）式成为

$$\varphi(B)x_t=\theta(B)a_t, \tag{29}$$

其中

$$\varphi(B)=1-\sum_{i=1}^{p}\varphi_i B^i, \tag{30}$$

$$\theta(B)=1-\sum_{i=1}^{q}\theta_i B^i. \tag{31}$$

将 $\theta^{-1}(B)$ 展开成级数为：

$$\theta^{-1}(B)=1+\sum_{j=1}^{\infty}d_jB^j \quad , \tag{32}$$

把（32）式代入（29）式得：

$$a_t=\varphi(B)x_t+\sum_{j=1}^{\infty}\varphi(B)d_jB^jx_t \quad ; \tag{33}$$

表 1

t	y(t)	$\nabla y(t)$	$\nabla^{24}y(t)$	$\nabla\nabla^{24}y(t)$
1	0.63	0.06	0.89	−0.32
2	0.69	−0.25	0.57	−0.19
3	0.44	−0.21	0.38	−0.44
4	0.23	−0.19	−0.06	−0.17
5	0.04	−0.32	−0.23	−0.43
6	−0.28	−0.49	−0.66	0.45
7	−0.77	−0.02	−0.21	0.29
8	−0.79	0.00	0.08	0.33
9	−0.79	0.74	0.41	0.35
10	−0.05	0.73	0.76	−0.44
11	0.68	0.74	0.32	−0.01
12	1.42	0.76	0.08	−0.23
13	2.18	0.57	−0.03	−0.11
14	2.75	0.36	−0.17	−0.14
15	3.11	0.29	−0.41	−0.24
16	3.40	−0.04	−0.55	−0.14
17	3.36	0.01	−0.68	−0.10
18	3.37	0.01	−0.15	−0.03
19	3.38	−0.58	−0.1	0.53
20	2.80	−0.19	0.01	0.05
21	2.61	−0.24	0.05	0.11
22	2.37	−0.19	−0.01	0.09
23	2.18	−0.19	0.29	−0.06
24	1.99	−0.47	0.12	0.30
25	1.52	−0.26	0.24	−0.17
26	1.26	−0.44	0.65	0.12
27	0.82	−0.65	0.81	0.41
28	0.17	−0.36	0.72	0.16
⋮	⋮	⋮	⋮	⋮

表 2　自相关函数和偏相关函数表

k	y(t)		∇y(t)		∇^{24}y(t)		$\nabla\nabla^{24}$y(t)	
	ρ_k	φ_{kk}	ρ_k	φ_{kk}	ρ_k	φ_{kk}	ρ_k	φ_{kk}
1	0.94	0.94	0.73	0.73	0.80	0.80	0.25	0.25
2	0.73	0.73	0.54	0.00	0.51	-0.35	-0.08	-0.16
3	0.63	-0.29	0.33	-0.14	0.26	-0.03	-0.19	-0.13
4	0.42	-0.08	0.16	-0.07	0.10	0.03	-0.09	-0.02
5	0.21	-0.07	0.01	-0.09	-0.01	-0.09	-0.16	-0.18
6	0.00	-0.14	-0.18	-0.25	-0.04	0.10	-0.40	-0.41
7	-0.19	-0.05	-0.28	-0.02	-0.07	0.30	-0.19	-0.08
8	-0.35	-0.07	-0.33	0.10	0.23	0.10	0.19	0.13
9	-0.48	-0.16	-0.35	-0.20	0.30	-0.12	0.23	-0.03
10	-0.57	-0.03	-0.40	-0.11	0.28	-0.02	0.21	0.12
11	-0.63	-0.08	-0.42	-0.15	0.19	-0.07	-0.03	-0.15
12	-0.65	-0.07	-0.37	-0.18	0.10	0.03	-0.10	-0.24
13	-0.63	0.00	-0.39	-0.02	-0.05	0.16	0.02	0.07
14	-0.57	-0.04	-0.36	-0.21	0.00	-0.07	-0.17	-0.16
15	-0.48	-0.02	-0.33	-0.14	0.02	0.14	-0.16	-0.04
16	-0.36	0.08	-0.21	-0.17	0.12	0.06	-0.19	-0.12
17	-0.21	0.14	-0.07	0.00	0.20	-0.03	0.14	0.08

表 3

日期(1983年)	观 测 值	AR(2)预报值*	时序预报的残差	逐步回归预报值	回归预报的残差
1.24	3.25	3.24	-0.01	2.97	-0.28
2.15	2.83	2.89	0.06	2.79	-0.04
2.23	2.78	2.88	0.10	2.71	-0.07
3.24	2.71	2.89	0.18	2.33	-0.38
4.11	2.92	2.68	-0.24	2.10	-0.82
4.25	2.94	2.23	-0.71	1.95	-0.99
5.9	2.41	1.58	-0.83	1.10	-1.31
5.23	1.55	1.18	-0.37	0.97	-0.58
6.13	1.50	0.80	-0.70	1.33	-0.17
6.27	1.25	0.51	-0.74	1.40	0.15

* 如果预报的日期不是每月的15日或30日，则采用内插技术.

$$x_t=\varphi_1 x_{t-1}+\cdots+\varphi_p x_{t-p}+a_t-\sum_{j=1}^{\infty}d_j\varphi(B)B^j x_t. \tag{34}$$

在上式中，令 $d_j=0$，并取 $E(a_t)=0$ 为 a_t 的预报值，则AR(p) 序列的预报公式为：

$$\hat{x}_t=\hat{\varphi}_1 x_{t-1}+\hat{\varphi}_2 x_{t-2}+\cdots+\hat{\varphi}_p\hat{x}_{t-p}. \tag{35}$$

在（34）式中，令 $\varphi_i=0$，取 $E(a_t)=0$ 为 a_t 的预报值，则MA(q) 序列预报公式为：

$$\hat{x}_t=-\sum_{j=1}^{t-1}d_j\hat{x}_{t-j}, \qquad N+1\leqslant t\leqslant N+q. \tag{36}$$

可以证明，当 $t>N+q$ 时，

$$\hat{x}_t=0. \tag{37}$$

利用（21）式和（35）式很容易地建立大坝变形的预报公式为：

$$\hat{x}_t=1.0813\hat{x}_{t-1}-0.3536\hat{x}_{t-2} \tag{38}$$

或

$$y(t)=y(t-24)+1.0813(y(t-1)-y(t-25))-0.3536(y(t-2)-y(t-26)). \tag{39}$$

其中 $y(t)$ 为大坝位移或变形值。

由（39）式得到的位移预报值列于表3的第3列。与第2列的原始观测数据比较可知，利用(39)式的AR（2）模型有效地对大坝变形作了预报，预报残差列于该表的第4列中，为了与逐步回归预报法作比较，把回归预报结果列于第5列中，其预报均方差为 $\hat{\sigma}_R^2=\frac{1}{10}\sum_{i=1}^{10}\nu_i^2=0.3985$。而用AR(2)预报得到的均方差为 $\hat{\sigma}_T^2=0.2471$，精度略高。因此，时间序列分析方法可望在大坝变形分析中得到广泛的应用。

4 結　語

本文把时间序列分析方法应用于大坝变形分析，建立了AR(2)模型并对大坝变形作了预报，与原始观测值比较可见，文中提出的方法具有相当高的预报精度。在本例中，AR(2)模型的预报效果优于逐步回归法。因此，时间序列分析法是大坝变形分析的一个有力工具。但是，由于预报是短期的（半年），而又缺少长期原始观测数据，所以时间序列分析预报的长期特性还有待进一步研究。

刘大杰教授和罗玉芳副教授对本文的写作给予了极大的帮助，特此致谢。

参 考 文 献

〔1〕 安鸿志等，时间序列的分析与应用，科学出版社，1983.

〔2〕 徐立子等，随机序列的新息方法在电力系统负荷预报中的应用，科学通报，Vol.25, No.22，1980.

〔3〕 黎令仪等，震级序列的门限建模，地球物理学报，Vol. 28，No. 3，1985.

Application of Time Series Analysis to the Prediction of Deformation for Large Dams

Xu Peiliang

Abstract

This paper first introduces three basic models—the ARMA model, the AR model and the MA model in the time series analysis and their statistical properties. Taking the absolute displacements of a large dam in the western of China as an example, the paper puts the emphasis on the procedure of the establishment of the model for deformation analysis of the large dam. An AR(2) model has been built and used in the prediction of deformations. The result shows good prediction accuracy. Compared with the stepwise regression analysis, it is of better accuracy. Therefore, if the regression factors have not been observed, the method mentioned in this paper will show its advantages in deformation analysis.

[Key words] time series; AR model; regression analysis; displacements of dams