

时间序列方法在观测资料分析中的应用*

刘金元 杨春林 吴正文

(同济大学地下建筑与工程系 上海 200092) (南京水利科学研究院材料结构所 南京 210024)

摘 要 利用时间序列方法 ,分析去除确定性部分后的残差序列 ,提取其中的有规律成分作为预报时的补偿项 ,形成叠合模型 ,进行建筑物观测资料的分析 .实例表明 ,该方法提高了预报精度 ,证明了叠合模型方法的适用性 .

关键词 观测资料 叠合模型 时间序列方法 残差序列

中图号 TU12

实际的工程建筑物 ,由于其本身材料的不均匀性和周围许多因素的影响而呈现出复杂的物理表现 .因此 ,对于工程结构物的分析 ,在从理论角度分析材料本身特性及影响因素作用机理的同时 ,也存在测试结构物的实际表现 ,获得反应记录而反推结构性能参数的方法 .实测资料的分析中存在物理分析和数学分析两种途径 .物理分析是从力学角度研究各影响因素同结构物之间的关系 ,数学分析则是利用数学方法进行分析 .最好也是最复杂的分析方法是物理分析方法 ,称为确定性模型方法 .但是多数情况下结构物影响因素不能完全确定 ,因而很难从力学角度分析 ,而是用被称为统计模型的数学方法进行处理 ,再就是将确定性模型同统计模型结合的混合模型方法 .

然而 ,上述三种方法存在着一个共同点 :都是对建筑物的实测资料分析后获得最佳的物理量描述方程和残差序列 .一般情况下 ,此物理量描述方程就用以代表该物理量而进行预报和安全评估 ,而不考虑残差的影响 .由于各种方法中假定的存在以及简化的不恰当等原因 ,残差中也存在有规律的成分需要加以提取 .本文则利用时间序列方法(以下简称为时序方法)分析统计模型建模后的残差序列 ,提取其中有规律的成分作为原模型的补充 ,从而提高预报精度 .

1 时序方法的数学原理

1.1 ARMA 模型的定义

时序方法包括参数模型和非参数模型 .本文用的是参数模型中的 ARMA 模型 .如果下列模型

$$x_t - \varphi_1 x_{t-1} - \varphi_2 x_{t-2} - \cdots - \varphi_p x_{t-p} = \alpha_t - \theta_1 \alpha_{t-1} - \cdots - \theta_q \alpha_{t-q} \tag{1}$$

满足一定条件^[1] ,则称为 p 阶自回归- q 阶滑动平均混合模型 ,记为 ARMA(p , q)模型 .其中 $\{\alpha_t\}$ 是白噪声序列 , $\{x_t\}$ 是平稳且零均值序列 .

1.2 建模原理

a. 设 $\{x_t\}$ 为平稳零均值序列 ,考虑由 $x_{t-1} \cdots x_{t-k}$ 对 x_t 的线性最小方差估计 ,整理得

$$\begin{bmatrix} 1 & \rho_1 & \cdots & \rho_{k-1} \\ \rho_1 & 1 & \cdots & \rho_{k-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \cdots & 1 \end{bmatrix} \begin{Bmatrix} \varphi_{k1} \\ \varphi_{k2} \\ \vdots \\ \varphi_{kk} \end{Bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_k \end{bmatrix} \tag{2}$$

其中

$$\rho_k = E_{x_t x_{t+k}} / \sqrt{E_{x_t}^2 E_{x_{t+k}}^2}$$

式(2)即称为 Yule-Walker 方程 ,系数 φ_{k1} 称为 $\{x_t\}$ 的偏相关系数 .

收稿日期 :1997-04-14
第一作者简介 :刘金元 ,男 ,博士研究生 ,岩土工程专业 ,主要从事软土地基与地下深基础工程设计理论研究 .
* 南京水利科学院科研基金资助项目(494001)的部分研究成果 .

b. 根据时序分析理论^[1],通过 φ_{k1} 和 ρ_k 的计算值来初步差别模型的类别,利用赤池(Akaike)的 BIC 准则来确定模型阶数,然后经过粗估计和精估计确定模型参数.

c. 对于模型序列 $\{x_t\}$ 在已知 $x_1, x_2, \dots, x_k$ 情况下,对 x_{k+1} 预报可由式(1)得到,第 L 步的预报值是从第 1 步一直递推得到.

1.3 ARIMA 模型

ARIMA 模型是利用差分将非平稳序列化为准平稳序列加以分析的. 设 $\{Z_t\}$ 是非平稳序列,存在 $\nabla^d Z_t \equiv x_t(t > d)$ 其中 ∇^d 是 d 阶差分算子,而其中的 $\{x_t\}$ 是 $\text{ARMA}(p, q)$ 序列,这样的 $\{Z_t\}$ 称为 p 阶自回归- d 阶差分- q 阶滑动平均模型,用 $\text{ARIMA}(p, d, q)$ 表示.

2 基于时序方法的叠合模型的应用

2.1 叠合模型的定义

叠合模型是将时间序列看作是确定性部分和随机部分的叠加. 实际测值大多是非平稳的,常常呈现明显的确定性趋势,在同时有影响因素的观测资料分析中,这些影响因素常常用来表示模型中的确定性部分,而在没有影响因素的情况下就从数学角度来提取. 剩下的部分则认为是随机的,要用随机序列的分析方法去分析. 本文利用的是时序分析方法. 确定性部分一般包含线性趋势、指数趋势和周期趋势. 为方便,记

$$y_t^T = \sum_{j=0}^s A_j e^{k_j t}, \text{ 因此叠合模型的一般形式为 } y_t = \sum_{j=0}^s A_j e^{k_j t} + Z_t \tag{3}$$

其中 Z_t 是随机部分.

2.2 实际应用的几个问题

前面时序分析方法的介绍中不止一次地强调时序是平稳序列,所以对于实际问题首先要进行平稳性检验,采用非参数检验^[2].

另一个问题是数据的等距问题. 实际的观测数据,由于各种因素的影响,大多情况下是非等距测值,因此在利用时序分析之前要对数据进行等距化处理.

2.3 叠合模型的应用

采用某大坝 4 号坝段的水平位移测值(1983 年 6 月至 1990 年 1 月)进行实例分析. 首先采用拉格朗日插值公式进行等距处理,然后再进行叠合模型的建模.

2.3.1 确定性部分的提取

由于水平位移时效的特点^[2],假定时效曲线

$$y_{t0} = a_1 + b_1 \log(1 + t) \tag{4}$$

对于周期部分,根据测值的近似年变化特点,分别进行了年周期、两年周期和半年期趋势的提取,用带阻尼的非线性最小二乘法求得参数值. 当上述几项提取后,再增加趋势项提取,残差平方和已不再明显降低,说明确定性部分分离已大体完毕.

$$y_t = a_1 + b_1 \log(1 + t) + e^{b_2 t} (b_3 \sin 2\pi t + b_4 \cos 2\pi t) + e^{b_5 t} (b_6 \sin \pi t + b_7 \cos \pi t) + e^{b_8 t} (b_9 \sin 4\pi t + b_{10} \cos 4\pi t) \tag{5}$$

式(5)中 t 以年计. 通过几次计算,最后得计算参数见表 1. 确定性部分和实测值的比较示于图 1. 由图可知,确定性模型是适宜的.

表 1 确定性部分的计算参数
Table 1 Parameters in the model

参数	a_1	b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	b_9	b_{10}
数值	1.87	-0.746	-0.017	-2.556	-1.136	-0.219	-0.377	0.299	-0.093	0.047	0.312 17

2.3.2 残差的 ARIMA 分析

去除确定性部分后的残差,经平稳性检验知此残差序列是非平稳序列,最后经两次差分后序列建模为 $\text{ARMA}(2, 0)$ 即 $\text{AR}(2)$ 模型.

$$x_t = -1.135x_{t-1} - 0.6587x_{t-2} + a_t \tag{6}$$

模型 $\text{ARMA}(2, 0)$ 拟合后,使 $\{x_t\}$ 残差平方和由 211.7 减到 58.6. 将残差的有规律部分形成的补偿项 Z_t

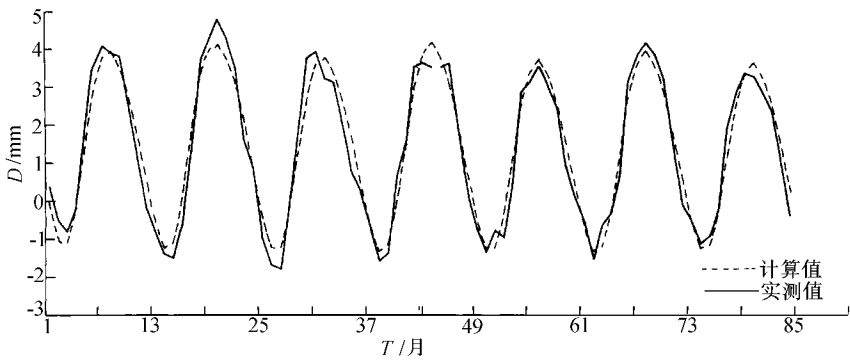


图1 确定性部分和实测值比较

Fig.1 Comparison of calculated and observed data

加到原模型 y_t 上形成叠合模型(见式(6))进行预报,即

$$Y = y_t + Z_t$$

(7)

式中 Z_t 是残差形成的补偿项.

$$\nabla^2 Z_t = x_t$$

(8)

$\{x_t\}$ 的预测值通过式(6)得到,然后依式(8)算得补偿项 $\{Z_t\}$ 的预测值,代入叠合模型得出总的预测值.预测值与实际值的比较示于表2.

利用后延预报方法知: $S_1^2 = 4.19$, $C = S_2/S_1 = 0.174$, $P\{|e - \bar{e}| < 0.6745S_1\} = 100\%$. 根据预报精度评估指标^[2]可知预报精度很好.

表2 预测值与实测值比较

Table 2 Comparison of predicted and observed data

步长	实测值	确定部分	$\alpha(1/N)$	$\alpha(1/N)$	计算值	误差
L	y	y_t			$y_t + Z$	e
1	-2.74	-3.09	1.03	0.34	-2.75	0.01
2	-1.89	-2.55	-0.26	0.53	-2.02	0.13
3	-0.67	-1.39	-0.38	0.44	-0.95	0.28
4	-0.09	-0.22	0.60	0.79	0.57	-0.66
5	2.56	1.12	-0.43	0.93	2.05	0.51

实测值为经等距处理后的值. $\bar{e} = 0.054$, $S_2^2 = 0.127$, $|\bar{e}|_{\max} = 0.66$, $|\bar{e}| = 0.32$.

3 结 语

- a. 从叠合模型的预报可看到效果是可以的,但随着步长的增大,误差显著增大,所以时序分析方法较适用于短期预报.为进一步提高预报精度,有必要引入新息预报.
- b. 实例分析中利用了 ARIMA 模型,虽然是向非平稳序列处理推进一步,但理论和方法都不是本质的,而且在差分阶数的确定上尚无较好的原则,所以有必要研究非平稳序列的非线性模型,以获得更好的效果.

参 考 文 献

1 黄俊钦,静.动态数学模型的实用建模方法.北京:机械工业出版社,1988.120~145

2 吴中如,沈长松,阮焕祥.水工建筑物安全监控理论及其应用.南京:河海大学出版社,1990.94~110

Application of Time Series Method to Structural Observed-data Analysis

Liu Jinyuan Yang Chunlin

(Dept. of Geotechnical Engineering, Tongji Univ., Shanghai 200092)

Wu Zhengwen

(Div. of Structure & Material, Nanjing Hydraulic Research Institute, Nanjing 210024)

Abstract The time series method is used to extract the regular part from the residual series received by modeling of observed data. A supplementary model is obtained by attaching the supplementary item to the original model. An example indicates the feasibility of the supplementary model and its ability of improving the prediction precision.

Key words observed data; supplementary model; time series method; residual series