

曲线拟合与自回归模型在地铁变形监测中的运用*

金彪, 吴北平, 李艳芳

(中国地质大学, 湖北 武汉 430074)

摘要: 曲线拟合与自回归是变形监测数据分析的常用方法, 为此, 介绍了如何建立曲线拟合与自回归模型, 以及这两种模型的适应性检验, 并结合武汉市某地铁竖井地表的沉降变形数据, 分别建立了曲线拟合与自回归模型进行沉降预测结果的分析与检验, 并用程序实现了沉降预测的自动化。通过实验比较, 结果表明: 这两种模型在相对平稳的变形中基本都能满足变形监测的要求, 但是自回归模型的预测残差要比曲线拟合模型小。

关键词: 曲线拟合; 自回归; 沉降观测; 最小二乘法

中图分类号: P 226⁺.3

文献标识码: B

文章编号: 1007-9394(2009)01-0035-03

Application of Curve Fitting and Auto Regression Model to Distortion Monitoring in Subway

JIN Biao, WU Bei-ping, LI Yan-fang

(China University of Geoscience, Wuhan Hubei 430074, China)

Abstract: The curve fitting and Auto-Regressive is the commonly used method for distortion monitor data analysis, this article introduced how to establish the curve fitting and the Auto-Regressive model, as well as the adaptability examination of these two kinds of model, and established curve fitting and Auto-Regressive model separately unifies the data of Settlement distortion of some subway shaft surface in Wuhan, for analysising and examining the result of the Settlement forecasting, and having realized the automation of settlement forecast with the procedure. Through the comparison of experiment, the result indicates that although these two kinds of models basically can satisfy the request for distortion monitoring in the relatively steady distortion, forecast residual of the Auto-Regressive model must be smaller than the curve fitting model.

Key words: curve fitting; auto-regression; subsidence observation; least squares method

0 引言

随着地下工程(如地铁等)建设的逐渐增多,对工程建筑物的时空特性、动态变化监测和分析显得尤为重要。通常是在所要研究的建筑物上布设一些特征点,然后进行周期性的重复观测,再通过对被观测点群的研究,进而找出地下建筑物随时间变化的规律及其未来的发展趋势,为工程管理机构以及政府部门的管理、决策提供依据。本文主要结合武汉某一地铁竖井的部分监测数据,采用曲线拟合与自回归对部分特征点进行分析预报,并通过比较,分析了不同模型的预测结果。

1 曲线拟合模型

曲线的数据拟合,通常也被称为离散数据的曲线拟合,指的是给定函数 $y = u(t)$ 的一组观察值 (t_i, y_i) ($i = 0, 1, \dots, m$) 选定一组简单函数 $\Phi(t)$ ($k = 0, 1, \dots, n$) 作为基函数,通过确定拟合

模型 $f(t) = x_1 \Phi_1(t) + x_2 \Phi_2(t) + \dots + x_n \Phi_n(t)$ 的待定参数 x_k , 使 $f(t)$ 与观察值 (t_i, y_i) 在总体上尽可能接近^[1]。

在确定拟合模型的过程中,存在多种选择函数 $f(t)$ 的方法,其中最常用的是最小二乘法,即选择函数 $f(t)$, 使

$$\sum_{i=1}^m (y_i - \bar{y}_i)^2 = \sum_{i=1}^m [y(i) - f(t_i)]^2 = \min \quad (1)$$

在沉降观测中,设某特征点在某一观测时间的高程为 y , y 的变化规律是随时间 x 变化的函数。假设 $y = a_0 + a_1 x_1 + a_2 x_2 + \dots + a_m x_m$, 利用最小二乘法求出该方程的系数,有:

$$A = \begin{bmatrix} x_1 & x_1^2 & \dots & x_1^m \\ x_2 & x_2^2 & \dots & x_2^m \\ \dots & \dots & \dots & \dots \\ x_n & x_n^2 & \dots & x_n^m \end{bmatrix} \quad (2)$$

$$X = [a_0, a_1, \dots, a_n]^T \quad (3)$$

$$B = [Y_1, Y_2, \dots, Y_n]^T \quad (4)$$

式中: n 为特征点重复观测次数; m 为多项式次数。利用最小二乘法得到曲线拟合的参数方程:

$$X = [A^T A]^{-1} A^T B \quad (5)$$

2 曲线拟合模型假设检验

求得的拟合预测方程是否与实际相符合,一般来说需要进行假设检验。若曲线拟合模型的假设符合实际情形,即 $y(x)$ 与 $\Phi_i(x)$ ($i=0,1,2,\dots,n$) 之间确实存在线性关系,这时方程系数 $c_i \neq 0$ ($i=0,1,2,\dots,n$), 否则, $y(x)$ 就不依赖于 $\Phi(x)$ 或者说不线性地依赖于 $\Phi(x)$ 。

因此,作假设检验的假设: $H_0: c=0, H_1: c \neq 0$ 。在实际的计算过程中,通常使用公式:

$$Q_T = Q_R + Q_E \quad (6)$$

式中: $Q_R = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$ 是线性关系引起的偏差平方和; $Q_E = \sum_{i=1}^n (y_i - \hat{y}_i)^2$ 是除去线性因素以外其它因素引起的偏差平方和^[2]。

由 F 分布的定义可知: $F = \frac{Q_R}{Q_E/(n-2)} \sim F(1, n-2)$ 可用此作为检验统计量,对给定的显著水平 α , 当 $F > F_\alpha(1, n-2)$ 时,拒绝假设 H_0 , 认为 $y(x)$ 与 $\Phi_i(x)$ ($i=0,1,2,\dots,n$) 之间确实存在线性关系,即所建立的拟合方程有效,否则认为无效。

本文利用武汉某一地铁竖井变形监测前 16 期的数据建立预报方程,预报后两期的沉降值,以此对比。表 1 为 2008 年 11 月 13 日至 11 月 30 日的沉降观测资料,表 2 和图 1 为观测值与曲线拟合预测值比较。

表 1 2008 年 11 月 13 日至 11 月 30 日的沉降观测资料

Tab. 1 Subsidence observation data from Nov 13, 2008 to Nov 30, 2008

观测时间	观测值/m	观测时间	观测值/m
2008.11.13	24.047 5	2008.11.22	24.048 1
2008.11.14	24.047 7	2008.11.23	24.048 0
2008.11.15	24.048 3	2008.11.24	24.048 0
2008.11.16	24.048 1	2008.11.25	24.048 1
2008.11.17	24.048 5	2008.11.26	24.048 3
2008.11.18	24.048 0	2008.11.27	24.047 9
2008.11.19	24.048 2	2008.11.28	24.048 0
2008.11.20	24.048 0	2008.11.29	24.048 1
2008.11.21	24.048 3	2008.11.30	24.047 8

利用 VB 程序分别取 $m=2$ 和 $m=3$ 进行计算,得到预报方程为:

$$\begin{cases} Y = 24.076 5 + 1.204 2E-04X - 6.512 6E-06X^2 \\ Y = 24.072 7 + 3.540 1E-04X - 3.984 6E-05X^2 + 1.307 2E-06X^3 \end{cases}$$

曲线拟合采用了表中前 16 期的数据,预测后两期的沉降值。当 $m=2$ 时,曲线预测的残差平方和为 $6.03E-07 \text{ m}^2$; 当 $m=3$ 时,曲线预测的残差平方和为 $4.48E-07 \text{ m}^2$ 。这说明三次曲线拟合的效果要比二次曲线拟合好,于是选择 $m=3$ 时的模型为预测方程,由 f 检验可知, $Q_R = 4.293E-07$, $Q_E = 4.482E-07$ 。

$F = 13.408$, 显著水平为 0.05 时, $F_\alpha(1, 14) = 4.6$, 因 $F > F_\alpha(1, 14)$, 故认为观测值与观测时间之间存在较好的相关关系,所得曲线能较好地反映实际情况,在预测后两期的沉降观测值时,第 17 期的预测误差为 0.000 1 m, 第 18 期的预测误差为 0.000 6 m。

表 2 观测值与曲线拟合预测值比较

Tab. 2 Comparison between observation values and curve fitting forecast values

期数	观测值 /m	二次曲线 预测值 /m	二次曲线 预测残差 /m	三次曲线 预测值 /m	三次曲线 预测残差 /m
1	24.047 5	24.047 8	0.000 3	24.047 6	0.000 1
2	24.047 7	24.047 9	0.000 2	24.047 8	0.000 1
3	24.048 3	24.048 0	-0.000 3	24.048 0	-0.000 3
4	24.048 1	24.048 0	-0.000 1	24.048 1	0.000 0
5	24.048 5	24.048 1	-0.000 4	24.048 2	-0.000 3
6	24.048 0	24.048 1	0.000 1	24.048 2	0.000 2
7	24.048 2	24.048 2	0.000 0	24.048 2	0.000 0
8	24.048 0	24.048 2	0.000 2	24.048 2	0.000 2
9	24.048 3	24.048 2	-0.000 1	24.048 2	-0.000 1
10	24.048 1	24.048 2	0.000 1	24.048 1	0.000 0
11	24.048 0	24.048 2	0.000 2	24.048 1	0.000 1
12	24.048 0	24.048 2	0.000 2	24.048 0	0.000 0
13	24.048 1	24.048 1	0.000 0	24.048 0	-0.000 1
14	24.048 3	24.048 1	-0.000 2	24.048 0	-0.000 3
15	24.047 9	24.048 0	0.000 1	24.048 0	0.000 1
16	24.048 0	24.047 9	-0.000 1	24.048 1	0.000 1
残差平方和/m ²		6.03E-07		4.48E-07	

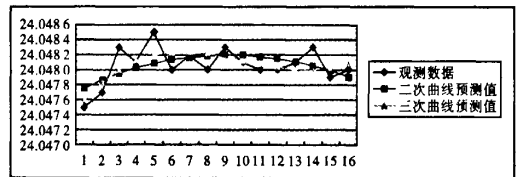


图 1 曲线拟合与观测值比较

Fig. 1 Comparison between curve fitting values and observation values

3 时间序列的自回归模型

3.1 时间序列的基本概念

随机序列就是指一串随机变量 X_1, X_2, \dots 所构成的序列, X_t 用 $(t=1, 2, \dots)$ 表示。当随机序列中的时间 t 是连续变化时就称为随机过程, 记为 $X(t)$ 。研究较多的一类随机序列是平稳随机序列或平稳随机过程它具有如下两个特点^[3]:

第一, 它的数学期望、方差不随时间变化, 亦即:

$$E(X_t) = \mu \quad (7)$$

$$E(X_t - \mu) = \sigma^2 \quad (8)$$

第二, 协方差函数只与时间间隔 $t-s$ 有关

$$\text{cov}(x_t, x_s) = E[(x_t - \mu)(x_s - \mu)] = R(t-s) \quad (9)$$

对某一随机序列, 若和它的一次现实求平均得到的统计特征, 当求平均的区间增大时, 能以任意接近 1 的概率逼近用各次现实的整个集合求平均得到的相应特征量, 就称这样的随机序列具有态历经性质, 即对具有这种性质的平稳随机过程 X_t , 数学期望为:

$$E(X_t) = \mu = \frac{1}{T} \int_0^T x_t dt \quad (10)$$

式中: T 为求平均的时间区间。方差为:

$$E(X_t - \mu) = \sigma^2 = \frac{1}{T} \int_0^T (x_t - \mu)^2 dt \quad (11)$$

时间间隔为 τ 的协方差为:

$$\begin{aligned} \text{cov}(x_i, x_{i+\tau}) &= E[(x_i - \mu)(x_{i+\tau} - \mu)] \\ &= \frac{1}{T - \tau} \int_0^{T-\tau} [x(t) - u][x(t + \tau) - u] dt \end{aligned} \quad (12)$$

相应的相关函数为:

$$R(\tau) = \frac{\text{cov}(x_i, x_{i+\tau})}{\delta} \quad (13)$$

如果平稳序列 z 的数学期望为 0, 自协方差函数有如下性质:

$$R_z(\tau) = \text{cov}(z_i, z_{i+\tau}) = \begin{cases} \delta^2, & \tau = 0 \\ 0, & \tau \neq 0 \end{cases} \quad (14)$$

则称 z_i 为白噪声序列或纯随机序列, 白噪声在时间序列分析中常常表示各种复杂过程的剩余分量, 其实质可理解为各种相互独立的随机干扰。

3.2 p 阶自回归模型 (AR(P))

时间序列 $\{x_i\} (i=1, 2, \dots, n)$ 的自回归模型为:

$$x_i = \Phi_1 x_{i-1} + \Phi_2 x_{i-2} + \dots + \Phi_p x_{i-p} + \alpha_i \quad (15)$$

自回归模型是一种线性模型, Φ_1, \dots, Φ_p 为模型参数, p 为模型阶数, $\{\alpha_i\}$ 假设为白噪声序列, 即 $E(\alpha_i) = 0$, 均方差为 δ^2 , 各 α_i 之间不相关。由式 (16) 得误差方程为:

$$V_i = \Phi_1 x_{i-1} + \Phi_2 x_{i-2} + \dots + \Phi_p x_{i-p} - x_i \quad (16)$$

式中: $i = p + 1, p + 2, \dots, n (n - p > p)$ 。写为矩阵形式为:

$$V = X\Phi - y \quad (17)$$

在 $V^T V = \min$ 条件下, 模型参数最小二乘解^[4]为:

$$\Phi = (X^T X)^{-1} X^T Y \quad (18)$$

4 自回归模型的适应性检验

所谓模型的适应性是指一个时间序列模型解释系统动态性的程度。一个时间序列的适合模型应该完全或基本上解释了系统的动态性, 从而模型中残差序列 $\{\alpha_i\}$ 应该是白噪声序列。模型的适应性检验实质上就是检验 $\{\alpha_i\}$ 序列是否为白噪声序列^[5]。由式 (16) 可知: $\alpha_i = x_i - \Phi_1 x_{i-1} - \Phi_2 x_{i-2} - \dots - \Phi_p x_{i-p}$, $\{\alpha_i\}$ 为残差序列。

4.1 相关函数法

相关函数法是通过计算并考察 $\{\hat{\alpha}_i\}$ 的自相关函数来判断残差序列的独立性。设 $\hat{\rho}_k$ 表示 $\{\alpha_i\}$ 序列的自相关函数, 即

$$\hat{\rho}_k = \frac{\sum_{i=1}^{n-k} \hat{\alpha}_i \hat{\alpha}_{i+k}}{\sum_{i=1}^n \hat{\alpha}_i^2} \quad (19)$$

如果残差序列 $\{\hat{\alpha}_i\}$ 为白噪声序列, 可以证明, 样本个数充分大时, $\hat{\rho}_k$ 是互不相关的, 且近似于正态分布, 即 $\hat{\rho}_k \sim N(0, 1/N)$ 。因此, 如果 $|\hat{\rho}_k| \leq 1.96/\sqrt{N}$, 就可以在 0.05 的显著水平下接受 $\rho_k = 0$ 的假设, 认为 $\{\alpha_i\}$ 是独立的^[6]。

4.2 χ^2 检验法

在有些情况下, 由于偶然因素, 可能有个别的 $\hat{\rho}_k$ 不满足 $|\hat{\rho}_k| \leq 1.96/\sqrt{N}$, 此时也可以将自相关函数放在一起整体进行检验, 为此, 构造如下的 Box - pierce 统计量:

$$Q = n \sum_{k=1}^m \hat{\rho}_k^2 \quad (20)$$

式中: m 为模型中的最大时滞; n 为时间序列的观测值的个数^[7]。对于给定的置信概率 $1 - \alpha$, 可查 χ^2 分布表中自由度为 m 的 $\chi^2_{\alpha}(m)$, 将 Q 与 $\chi^2_{\alpha}(m)$ 比较, 若 $Q \leq \chi^2_{\alpha}(m)$, 则判定所选 AR 模型是合适的, 可以用于预测; 否则是不合适的, 应进一步改进模型^[8]。

下面同样用上述程序中的数据建立 AR(3) 模型, 并且也是采用前 16 期的数据建立模型, 然后对后两期的沉降值进行预测, 由程序计算可知, 该预测模型为 AR(3):

$$\begin{aligned} X(t) &= 24.987\,8 + 1.727\,7E-03X(t-1) + 0.305\,5X(t-2) \\ &\quad - 0.346\,2X(t-3) \end{aligned} \quad (21)$$

沉降观测值和自回归预测值的比较见表 3, AR(3) 模型预测值与观测值比较见表 4。

表 3 沉降观测值与自回归预测值比较

Tab. 3 Comparison between subsidence observation values and auto-regression forecast values

期数	观测值/m	AR(3) 预测值/m	AR(3) 预测残差/m
1	24.047 5	24.047 5	0.000 0
2	24.047 7	24.047 7	0.000 0
3	24.048 3	24.048 3	0.000 0
4	24.048 1	24.048 1	0.000 0
5	24.048 5	24.048 3	-0.000 2
6	24.048 0	24.048 0	0.000 0
7	24.048 2	24.048 1	-0.000 1
8	24.048 0	24.048 0	0.000 0
9	24.048 3	24.048 1	-0.000 2
10	24.048 1	24.048 0	-0.000 1
11	24.048 0	24.048 1	0.000 1
12	24.048 0	24.048 0	0.000 0
13	24.048 1	24.048 1	0.000 0
14	24.048 3	24.048 0	-0.000 3
15	24.047 9	24.048 1	0.000 2
16	24.048 0	24.048 0	0.000 0
残差平方和/m ²			2.32E-07

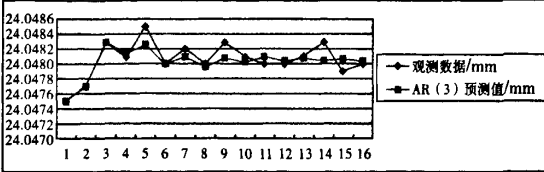


图 2 AR(3) 模型预测值与观测值比较

Fig. 2 Comparison between AR(3) model forecast values and observation values

回归模型也是采用表中前 16 期数据建模, 然后对后两期的沉降值进行预测, 当取 $p = 3$ 时建立 AR(3) 模型, 得到前 16 期的残差平方和为 $2.32E-07\text{ m}^2$, 然后 AR(3) 模型进行模型适应性检验, 首先进行第一步相关函数检验, 求出自相关系数, 进行自相关检验, 判断是否具有随机性。表 4 为前 4 个自相关系数。

表 4 自相关系数表

Tab. 4 Autocorrelation coefficient

自相关系数	值	自相关系数	值
ρ_1	-0.078 2	ρ_3	0.030 7
ρ_2	0.127 9	ρ_4	0.345 9

(下转第 44 页)

1) 控制测量: 对于测量监理而言, 把握好市政工程中控制测量是至关重要的, 对原有控制成果不管何种施工控制网, 在放样前, 必须进行复测, 以确定点位和测量成果的一致性和使用的可靠性。

2) 编制高程一览表: 要求施工部门编制高程一览表, 并进行逐项检核, 因为这是施工的直接依据。

3) 中线放样: 线路设计阶段所测设的中线桩至开始施工时, 一般均有丢失现象, 为保证施工中中线位置准确可靠, 施工前应根据定线条件复核中桩。

4) 横断面测量(原地面测量): 为了核实土方工程量, 故一般在开工前需进行横断面测量, 相邻横断面的间距越小则方量越精确, 但测量工作也将相应增大。一般可按设计图提供横断面间隔施测, 地形变化较大可加桩, 而对地势平坦或工程面较窄的地方可不施测横断面, 而以中线桩的地面高程代表两侧地形即可。

5) 隐蔽工程测量: 隐蔽工程一般均为标外工程, 监理应请业主共同参与, 便于确认工程量。

工程竣工验收就是对工程施工阶段的质量通过检查评定, 考核项目质量是否达到设计要求, 并通过验收确保工程项目的质量。所以工程竣工验收对质量的影响是保证最终产品的质量, 测量监理人员应督促施工单位完成收尾工程和竣工验收测量资料的准备。

4 结束语

市政工程测量时间性强, 涉及的数据多, 作业环境复杂, 是一项非常精细的工作, 若稍有不慎, 发生错误而又不能及时发

现, 就会直接影响工程的顺利进行, 甚至造成严重损失。所以测量监理人员要全面了解设计意图、工程情况、施工特点, 有的放矢地采取有效措施做好监理工作。监理人员不仅要以监理规程、技术标准为准绳, 秉公办事、一丝不苟、严格监理, 同时要热情服务, 及时帮助和指导施工人员解决技术业务问题, 以保证工程质量达到预定的质量标准和等级。

致谢: 本文在撰写过程中得到云南省地矿测绘院教授级高级工程师张述清的悉心指导, 在此表示感谢!

[参 考 文 献]

- [1] GB50026-2007, 工程测量规范[S].
- [2] CJJ8-99, 城市测量规范[S].
- [3] CJJ1-90, 市政道路工程质量检验评定标准[S].
- [4] CJJ1-2008, 城市道路工程施工与质量验收规范[S].
- [5] GB50268-97, 给水排水管道工程施工及验收规范[S].
- [6] CJJ2-90, 市政桥梁工程质量检验评定标准[S].
- [7] CJJ3-90, 市政排水管道工程质量检验评定标准[S].
- [8] 洪立波, 等. 城市测量手册[M]. 北京: 测绘出版社, 2003.
- [9] 江苏华宁工程咨询有限公司. 公路工程施工质量监理手册[M]. 北京: 知识产权出版社, 2003.
- [10] 谭明建. 测绘监理的现状与对策[J]. 四川测绘, 2007, 30(1): 45-47.

作者简介: 李明(1961-), 男, 上海人, 工程师, 现主要从事市政工程、交通工程监理方面的工作。

(上接第37页)

如表4所示, 前4个自相关系数均落在 $\pm(1.96/\sqrt{18}) = \pm 0.4198$ 内, 选择置信水平为0.05时, $\chi^2_{\alpha}(m) = 26.2962$, 当 $\rho = 3$ 时, $Q = n \sum_{k=1}^{\rho} \rho_k^2 = 17.9863 < \chi^2_{\alpha}(m)$, 则表明AR(3)模型适合, 可以用于预测, 用AR(3)模型预测后两期的沉降值时, 第17期的预测残差为 $-2.6E-05$ m, 第18的预测残差为0.0002 m。

5 结束语

通过曲线拟合法、自回归分析法对武汉市某地铁竖井地表沉降观测中的监测数据进行计算分析, 并对两种方法分别建立了两个模型进行比较, 然后通过适应性检验选择了更适合的模型用于预测, 最后选择了三次曲线拟合模型和AR(3)自回归模型, 通过这两个模型对后两期的沉降值进行预测, 并对预测残差进行比较, 可以看出自回归模型的预测残差要比曲线拟合模型小。但是, 当实测沉降值出现较大波动时, 曲线拟合与自回归模型预测结果都会变差, 这主要是由于本文的预测模型只是建立在观测数据的基础之上, 故对沉降突变的预测能力较差, 对于这种情况就需要加入相应的物理参数才能提高预测精度。总体上说, 这两种模型在相对平稳的变形中基本上都能满足变形监测的要求。

[参 考 文 献]

- [1] 黄声享, 尹晖, 蒋征. 变形观测数据处理[M]. 武汉: 武汉大学出版社, 2007: 98~108.
- [2] 王振龙, 胡永宏. 应用时间序列分析[M]. 北京: 科学出版社, 2007: 60~72.
- [3] 何书元. 应用时间序列分析[M]. 北京: 北京大学出版社, 2003: 23~35.
- [4] 王有良, 唐跃刚. 曲线拟合与GM(1, 1)模型沉降预测及相关性分析[J]. 测绘科学, 2008, (3): 33~35.
- [5] 朱海国, 尹晖, 陈雪丰. 地铁变形监测中曲线拟合与自回归模型的综合应用[J]. 测绘与空间地理信息, 2006, (6): 29~32.
- [6] 林正华, 冯仁忠. 自回归模型参数的最小二乘估计[J]. 吉林大学自然科学学报, 2001, (2): 2~6.
- [7] 邹自力, 李信. 曲线拟合在变形监测预报中的应用[J]. 江西测绘, 1997, (2): 15~18.
- [8] 万志远, 等. 概率论与数理统计[M]. 南昌: 江西高校出版社, 1995: 45~57.

作者简介: 金彪(1985-), 男, 湖北仙桃人, 硕士研究生, 主要研究方向为大地测量与测量工程。