

变形观测数据时间序列建模中的几个问题

王 卫¹, 杨志强², 杨建华², 王滕军²

(1 长安大学 地球科学与资源学院, 陕西 西安 710054; 2 长安大学 地质工程与测绘学院, 陕西 西安 710054)

摘要: 针对在变形观测数据时间序列建模中所遇到的问题, 对变形观测数据时间序列建模中的数据预处理、模型选择、模型定阶与系统稳定性检验等问题进行了研究, 提出了分析数据趋势项提取的 AR 模型方法, 编写了 C 语言计算程序, 实现了样条函数插值方法、周期项提取的差值法, 建立 ARMA 分析模型, 对变形预测数据与实际观测数据进行了比较。结果表明: 该方法的正确性以及时间序列分析方法在变形数据处理与分析中的适用性与可行性。

关键词: 时间序列; 建模; 模型; 预报

中图分类号: P207 **文献标志码:** A **文章编号:** 1672-6561(2008)02-0214-04

Several Important Problems in Time Series Modelling of Deformation Measurement Datum

WANG Wei¹, YANG Zhi-qiang², YANG Jian-hua², WANG Teng-jun²

(1 School of Earth Sciences and Resources, Chang'an University, Xi'an 710054, Shaanxi, China;

2 School of Geological Engineering and Surveying, Chang'an University, Xi'an 710054, Shaanxi, China)

Abstract According to the real problems that are encountered in setting the time series model, this paper studies such problems as the issues of deformation datum pre-proposing, model selecting, model setting and dynamic system stability judging. The methods to calculate the trend datum, periodical datum of deformation are put forward, and the C-programs of ARMA model to analyze the real problem is developed. Comparison between the surveying datum and calculated ones shows the correctness and availability of time series analysis in the data analysis of deformation survey.

Key words: time series; modelling; deformation; forecasting

0 引言

时间序列分析是 20 世纪初出现的一种现代数据分析与处理方法, 是系统辨识与系统分析的重要方法之一。时间序列分析的基本思想是, 对平稳、正态、零均值的时间序列 $\{x_t\}$, 若 x_t 的取值不但与其前 n 步的各取值 $x_{t-1}, x_{t-2}, \dots, x_{t-n}$ 有关, 而且还与其 m 前步的各个干扰 $a_{t-1}, a_{t-2}, \dots, a_{t-m}$ 有关, 按照多元线性回归的思想, 则可以得到最一般的 ARMA(n, m) 模型, 即 $x_t = \varphi_1 x_{t-1} + \varphi_2 x_{t-2} + \dots + \varphi_n x_{t-n} - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_m a_{t-m} + a_t$, $a_t \sim \text{NID}(0, \sigma_a^2)$, a_t 是白噪声序列。时间序列

分析方法研究系统输出(观测值)之间基于时间的变化关系, 是一种动态的数据分析与处理方法。同时, 由于该方法只以系统输出数据为基础对系统进行研究, 对于许多难以确定因果关系的系统而言, 是一种理想的分析方法和途径。根据时间序列分析方法的特点, 要求在数据分析与处理时有大量、准确可靠的系统输出数据, 从而尽可能逼真地反映系统的行为特征^[1]。

由于时间序列分析的理论和方法现在仍然处于发展之中, 实际应用中有许多问题仍有待进行研究。笔者根据在时间序列建模中遇到的具体问题, 对建模过程中的数据预处理、数据检验、模型定阶

与系统稳定性等问题进行相关讨论。

1 数据采集中的几个问题

数据采集与预处理是时间序列建模中的基础工作,其包括数据采集、检验与预处理,以得到平稳、正态、零均值的时间序列数据。

1.1 插值函数

时间序列分析的基础数据一般是离散、等间隔的数据序列,而现实中大多数的物理系统输出为连续信号,这就存在一个确定采样间隔的问题。另外,人们对某些系统的观测数据虽然是离散数据,但一般不是也不需要等间隔观测,而且在时序分析时也不一定需要用一个样本的所有数据,所以必须象对待连续信号一样,对离散型数据进行重新采样。在变形测量问题中,由于观测数据一般不是等间隔观测,所以应当进行等间隔重采样^[2]。

对离散型数据进行重采样,首先必须将离散型数据转化为连续型数据,显然数据的这种转化会导致一定程度的信号失真^[3]。如何把失真控制在一个很小的范围内,是个值得研究的问题。在建模过程中,笔者曾试图使用拉格朗日插值函数对原数据序列等间隔化,但经过系统分析发现这种方法在大多数情况下是不适用的。因为时序分析一般数据量很大,若建立一个统一的插值函数以期反映整个系统的性质,则正好掉进了拉格朗日插值的“陷阱”,当拉格朗日插值模型阶数较高时,这种插值方法在插值区域两端具有非常严重的信号失真现象。对于变形分析与预测,这种信号失真现象的影响是致命的。另外,拉格朗日插值方法计算稳定性差,个别数据的误差将影响所有插值结果的精度,有时甚至会引发灾难性后果。样条函数具有计算的稳定性、最佳逼近性和一致收敛性,并且在插值过程中具有很好的抗差能力,对于含有误差的数据序列进行插值处理效果很好^[4]。在此,选择样条函数作为数据插值函数模型。

1.2 采样间隔与采样样本长度的确定

插值函数确定以后,确定采样间隔与计算样本长度就成为一个重要问题。根据时间序列理论,当采样间隔 Δ 取得过大,在 $\{x_t\}$ 中将会损失观测信号中原有的相关性,以至使ARMA模型产生不应有的降阶;另外,如果 Δ 取得过小,又会使高频噪声与有用的信号难于区别,使ARMA模型产生不利的升阶^[1]。以上两种情况均会导致不同程度的系统

模拟失真。当原始数据为离散型,在进行合理插值之后,其高频噪声部分很不明显,当然, Δ 的取值也不应过小,但主要问题是 Δ 不能取的过大。在实际中,笔者认为, Δ 的取值不应大于原始数据序列的平均间隔,一般可以取的比较满意的效果。另外,为保证采样后序列与原序列没有显著差异,应当进行假设检验。在假设检验满足的前提下,适当减小 Δ 的取值,通常会取得令人满意的结果。

样本长度的确定主要关系到信号在频域中的能量泄漏效应与不同频率协波的分辨力问题^[2]。实际中,有些原始数据序列具有很长的样本长度,理论上应该使用所有原始数据,但是涉及到计算速度与计算机存储能力,一般应当对采样样本长度进行限制。另外,对于平稳序列而言,其数据为短记忆时间序列,没有必要使用过长的样本数据进行处理。但时间序列是建立在数理统计的基础上,太短的样本长度又不能满足这种分析方法的理论要求。通过实验,认为对变形观测数据,采样样本长度应当足够大,以满足时间序列分析的要求。对较小的离散型数据样本,在采样时一般要进行适当地放大。对过长的原始记录,可以改变采样长度进行检验,当两个不同的样本长度建立的模型没有显著差异时,选择较大的样本长度值作为采样样本长度。

1.3 趋势项提取

时间序列要求建模数据是短记忆的平稳数据序列。实际中的观测数据一般反映了动态系统的性质及运动过程,是长记忆的非平稳时间序列。所以在建模前,首先要将长记忆序列化为短记忆序列,也就是提取该时间序列的趋势项及周期项。趋势项提取有很多方法,例如差分法、最小二乘函数拟和法、指数平滑法以及移动平均平滑法等。建模中,根据Parzen的ARARMA时间序列分析思想,笔者使用AR模型对采样数据进行趋势项提取。周期项提取采用二次差分方法,进行实验数据计算。

2 模型识别与模型适用性检验

2.1 模型识别

模型识别是根据时间序列的统计特性来确定适合于该序列的时间序列模型,其形式分别为ARMA、AR和MA。一般根据时间序列的自相关函数和偏相关函数的性质来确定(表1)。

表 1 模型识别
Tab. 1 Model Identifying

类别	AR(<i>n</i>)	MA(<i>m</i>)	ARMA(<i>n</i> , <i>m</i>)
自相关函数	拖尾	截尾	拖尾
偏相关函数	截尾	拖尾	拖尾

实际建模中, 由于自相关函数和偏相关函数一般是其真值的估计值, 通常含有较显著的误差, 所以很难直接用上面的方法进行建模。对此, 实际建模时, 笔者先对时间序列建立 ARMA 模型, 然后, 根据模型 AR 和 MA 部分的参数值来确定最终的模型形式。这样就形成一个统一模型, 在具体分析中, 不去强调具体模型形式, 根据模型适应性确定模型的形式, 从而简化了模型检验的工作和过程。

2.2 模型适用性检验

模型适用性检验有很多准则, 例如, 白噪声准则、残差平方和检验准则、Akaike 信息准则等。实际应用可以借鉴差比检验准则, 它是上述各种准则的统一表达形式。在建模中, 应当根据具体问题决定采用何种准则。

笔者在建模中, 序列的趋势项提取实际上就用到一个 AR 模型, 这个 AR 模型由于其特殊的用途和目的, 即剔出时间序列中的趋势项, 所以该 AR 模型不能用上述任何准则进行检验。时间序列剔除趋势项后, 应该成为一个平稳时间序列。所以作者在这里应用数据的平稳性检验方法来作为该模型的适用性检验准则。在实际应用中, F-检验准则具有比较广泛的应用价值。在对整个系统模型的检验中, 使用白噪声准则进行模型检验可以保证具体问题分析的正确性。要进行白噪声检验, 要求有足够大的数据序列样本, 以保证统计量计算的精确性。这也是在具体问题分析中, 要求采样样本足够大的主要原因。

3 系统稳定性检验

系统稳定性分析, 是根据已建立时间序列动态模型的统计特性, 对动态系统的稳定性进行检验。整个建模过程中任何误差都会对稳定性分析产生影响, 所以在建模过程中应尽量减少产生误差的环节和过程。系统平稳性分析的理论已很成熟, 但在实际操作上, 这些方法似乎过于严格。实际分析中, 可根据以下原则对系统稳定性进行判断: 即在周期项与趋势项基本稳定的前提下, 如果预测步数不断增大, 预测值残差变化不明显或变化的过程缓

慢, 则可以认为系统是近似稳定的; 如果预测值的残差随着预测步数的增大迅速增加, 则这个系统是很不稳定的。

4 算例与分析

根据以上思路, 对西安市某引水工程变形观测中一个滑坡变形点的水平位移观测数据建立了时间序列 ARMA 模型, 参数估计使用最小二乘法解算, 使用 C 语言编写了数据分析程序。利用该模型对形变过程进行模拟计算、分析与预测, 并对预测变形值与实际观测变形值进行了比较, 结果如表 2。表中实测与预测值为该点 *X* 坐标的截断数据。

表 2 预报值与实测值比较

Tab. 2 Compare of Prediction Values and the Real Measurements			
时间	实测值/mm	预测值/mm	残差/mm
1993 年 1 月	2 225. 5		
2 月	2 224. 1		
3 月	2 223. 1		
4 月	2 224. 0		
5 月	2 221. 4		
6 月	2 224. 6	2 224. 957	−0. 357
7 月	2 222. 7	2 222. 939	−0. 239
8 月	2 223. 9	2 223. 422	0. 478
9 月	2 226. 0	2 225. 532	0. 468
10 月	2 224. 4	2 224. 374	0. 026
11 月	2 221. 6	2 222. 295	−0. 695
12 月	2 223. 5	2 224. 361	−0. 861
1994 年 1 月	2 223. 5	2 224. 017	−0. 517
2 月	2 223. 3	2 223. 266	0. 033
3 月	2 224. 2	2 224. 036	0. 164
4 月	2 224. 7	2 224. 332	0. 368
5 月	2 225. 8	2 224. 858	0. 942
6 月	2 227. 8	2 227. 150	0. 650
7 月	2 224. 7	2 224. 417	0. 283
8 月	2 226. 2	2 226. 408	−0. 208
9 月	2 224. 5	2 225. 832	−1. 332
10 月	2 222. 1	2 222. 721	−0. 621
11 月	2 228. 4	2 228. 064	0. 335
12 月	2 228. 7	2 227. 392	1. 308

在进行计算时, 由于只有该点的 24 期观测资料, 为满足时间序列模型对数据序列样本长度的要

求,在数据采样时,把样本长度放大了一倍,以满足白噪声模型检验的数据充分性要求。在数据处理中,使用样条函数插值进行采样数据的插值计算,利用AR模型进行趋势项提取,数据趋势项提取模型使用平稳序列检验原则。为建立模型简便与统一,选择ARMA模型作数据分析模型,参数计算采用最小二乘法进行。对预测数据与实测数据的残差进行检验,满足白噪声序列零均值、正态分布特征,说明了该方法的可行性与正确性,对实际变形曲线与预测曲线的比较说明,该模型准确地反映了变形体变形的系统特征。

变形模型的一步预报值(表2)高度逼近实际的变形值。多步预报,特别是三步以上预报不能准确模拟变形过程的系统特征,说明变形系统很不稳定,这与对该点的地质资料分析结果相吻合。

5 结语

通过以上问题的时间序列计算与分析可见,时间序列分析在变形观测数据分析与预测中具有重要的应用价值。通常,在变形物理分析较为困难的前提下,通过统计学的方法分析变形体的形变问题,使用时间序列分析可以研究变形体的系统形变规律与特征,为变形观测数据处理与分析提供了一种全新的研究方法和途径。当然,在变形观测数据处理中,也可以用于某些具有不确定性的系统误差

(有色噪声)问题的研究,对于更精确地研究变形体的变形过程与状态特征具有重要的意义。然而,尽管时间序列分析理论发展很快,在具体应用方面还有许多具体问题有待研究,特别是结合小波分析理论,对变形观测资料进行分析,将有利于对变形机理的深入研究,使变形测量数据分析技术得到迅速的发展。所以在实际工作中,应当根据具体问题对这种方法的应用进行深入研究,以期得到满足实际问题的最好模型和方法。

参考文献:

- [1] 杨叔子,吴雅.时间序列分析的工程应用[M].武汉:华中理工大学出版社,1991.
- [2] 田铮.时间序列的理论与方法[M].第2版.北京:高等教育出版社,2001.
- [3] 刘大杰,陶本藻.实用测量数据处理方法[M].北京:测绘出版社,2000.
- [4] 邓建中,葛仁杰,程正兴.计算方法[M].西安:西安交通大学出版社,1985.
- [5] 李日云,王利,张双成.灰色预测模型在高层建筑物沉降预测中的应用研究[J].地球科学与环境学报,2005,27(1):84-87.
- [6] 李珍照.大坝安全监测[M].北京:中国电力出版社,1997.
- [7] 郭英,高建国.统计学[M].北京:中国财政经济出版社,2001.
- [8] 易丹辉.统计预测——方法与应用[M].北京:中国统计出版社,2001.