

# 跨断层形变测量动态数据模型的研究\*

张 兴 飞

(广东省国土厅)

**【摘 要】** 本文主要讨论了在已知气温、地温、降雨等主要气象参数的前提下,如何运用时间序列分析方法处理跨断层观测数据,其中对混合回归方程的建立,模型方程与地震异常之间的联系以及对可能包含异常信息的观测数据进行计算和分析,结合两个实例做了较系统的探讨。此外还提出了用GM(1,1)模型对残差进行修正的灰色动态数据模型。实际应用表明上述做法是形变数据处理的一个有力工具。

**【关键词】** 形变测量; 时间序列; 混合回归; 疏系数; 最大迟后; 灰色模型

## 1 引 言

地震预报是当今自然科学面临的最具挑战性的问题之一,它与社会、经济密切相关,有着广泛的社会影响。我国是世界上首先采用跨断层形变测量的国家之一,并证实了它对地震预报的作用。关于跨断层形变测量数据排除干扰识别异常的处理,人们进行了广泛的研究,并取得了较好的成果。目前通常采用的是线性回归方法、动态灰箱分析法及动态处理的卡尔曼滤波法等。

本文将动态数据处理的时间序列分析法运用于跨断层形变测量数据处理,指出了适于处理跨断层形变测量数据的混合回归模型建立的途径和步骤,提出了用于形变测量数据的灰色动态模型。具体对四川鲜水河断裂带测点和北京八宝山大灰厂测点建立了动态数据的模型方程,用 FORTRA77 语言编制了相应的计算程序,并进行了计算分析,取得了较好的拟合效果。

## 2 跨断层形变测量及其数据处理

跨断层形变测量,就是在断层上盘和下盘布设控制点,用以监测断层两侧地表面的相对变形。其布点形式通常很简单,只布两点,上、下盘各一点,同时测两点之间的高差和基线长度。由观测所得的高差或基线长度值中,包含着断层活动信息、观测误差和各种外界因素

收稿日期:1991-10-28

\*本文是硕士论文的一部分,指导教师为刘大杰教授

的影响。观测值可由两种方式表达：

1、观测值 = 高差（或基线长度）理论值 + 观测误差；

2、观测值的变化 = 断层活动信息 + 各种外界因素的影响 + 观测误差；

现有的跨断层形变测量数据处理方法的途径为：选择正常期，建立正常模型，推估正常期外的各时刻状态，分析观测值与推估值的残差并找出异常。异常来自两方面，一方面是断层活动不正常了；另一方面是外界因素中有些因素不正常了。

### 2.1 用回归法处理跨断层形变测量数据

多元回归处理方法是在正常期内用多元回归法建立模型，用正常期内的观测数据来求各个回归系数，这样所得到的方程就认为是描述了正常期内状态过程的正常模型。推估的方法是利用回归方程直接推估，因为回归方程是时间的函数，给定一个具体时刻，就有一个回归值，也就是这个时刻的正常状态推估值。回归法具体有多元线性回归法、逐步回归法、褶积滤波法等。

### 2.2 卡尔曼滤波的动态处理

卡尔曼滤波是六十年代发展起来的一种现代滤波方法，它是通过建立相应的状态方程和量测方程，从而得到一系列递推式的卡尔曼滤波方程，由此做出最优估计。卡尔曼滤波法较回归法有更大的灵活性，它所处理的过程是随机过程，可以处理非线性过程，很容易解决各因子的影响为可变的情况，同时包含着一步预测，是考虑目标运动过程时的最优估计方法。在实际工作中，可运用卡尔曼滤波法建立水准、基线等的单独和联合处理模型。

## 3 基于混合回归理论的跨断层测量动态数据模型

### 3.1 混合回归理论的合理性

动态数据的时间序列模型实际上是一个差分方程，按其阶数不同可分为ARMA( $p, q$ )，AR( $p$ )，MA( $q$ ) 三种形式。通过分析上述时间序列模型和通常的回归模型的理论特性及应用情况，可看出它们都存在各自的局限性：普通回归模型没有体现出因变元自身前后及因变元与各自变元“过去”的依赖关系；而自回归模型则没有考虑其它变元对因变元的作用；多变量自回归模型在一定程度上弥补回归模型与单变量自回归模型的欠缺，但由于多变量自回归模型的具体形式及递推算法的限制，必须按一种固定的程式计算，可能使得估计的模型参数过多，计算结果不够理想。

设因变元序列为 $\{Z_t\}$ ，用 $X_{i,t}$  ( $1 \leq t \leq N, 1 \leq i \leq S$ ) 表示自变元，则线性混合回归模型为：

$$\begin{aligned} Z_t = & \alpha_0 + \alpha_1^{(0)} \cdot Z_{t-1} + \dots + \alpha_{p_0}^{(0)} Z_{t-p_0} + \alpha_1^{(1)} X_{1,t-1} + \dots + \alpha_{p_1}^{(1)} X_{1,t-p_1} \\ & + \dots + \alpha_1^{(s)} X_{s,t-1} + \dots + \alpha_{p_s}^{(s)} X_{s,t-p_s} + \varepsilon_t \end{aligned} \quad (1)$$

其中 $\{\varepsilon_t\}$ 为模型残差序列。

$P = \max(P_0, P_1, \dots, P_s)$  称最大迟后

混合回归模型能弥补单纯回归模型与单变量自回归模型的不足，而且由具体计算方法可

知, 它比多变量自回归模型构造灵活, 可以用较少的参数建立阶数较高的模型。当进行参估计时, 若量测变量个数较多且延迟步数较长, 采用 BIC (AIC) 准则挑选最佳线性混合回归模型的计算量很大, 且拟合模型的统计稳定性较差。因此本文采用改进了的 BIC 准则, 建立疏系数的线性混合回归模型, 其中心思想为: 先确定一个足标集合序列  $R_m, R_{m-1}, \dots, R_1$ , 令  $R_m = \{1, 2, \dots, m\}$ , 由高阶模型到低阶模型, 逐步剔除不重要变量。

### 3.2 基于混合回归理论建立数据模型

1) 数据的预处理。通常混合回归模型中的因变元, 应为三阶平稳序列。因此对原始数据  $y(t)$ , 如它具有明显的增长趋势或周期性变化的特点, 需通过一阶差分  $Y(t) = Y(t) - Y(t-1)$ , 周期差分  $\nabla Y(t) = Y(t) - Y(t-T)$ , 使之平稳化。对跨断层测量中的水准和基线测量数据, 由于它基本上保持在一常值附近波动, 因此只需使之零均值化后再代入运算, 然后将所得结果加上  $EY(t)$  即得到关于  $Y(t)$  的拟合结果。

2) 合理选取因子。在建立跨断层形变测量数据的模型前, 应根据实际情况合理选取对因变元影响较大的因子。由于跨断层形变测量是在地层表面进行野外测量, 无疑将受到各种外界因子的影响。无论是作为位移测量的标面, 还是断层两盘的相对运动, 在很大程度上受地温与地湿度的影响, 其中地下温、地表温、气温等综合反映出热胀冷缩的作用, 而地下水、降雨、蒸发等直接影响着地湿度, 因此它们都是应予以考虑的因素。同时因各流动测点未进行气象观测, 通常取离流动测点最近的气象观测站的有关资料作为干扰参数, 对在各个本身时间获得的观测数据, 有时应选取趋势活动背景项目, 以考虑断层正常趋势活动背景。

3) 建立混合回归模型。设  $\{Z_t\}$  表示跨断层的基线测量或水准测量数据序列,  $\{X_{i,t}\}, 1 \leq t \leq N, 1 \leq i \leq S$  代表各气象因子序列, 如气温、地表温、地下温、降雨等, 在考察某条水准或基线测量数据系列与其它测量数据系列的相关关系时,  $\{X_{i,t}\}$  中还可以包含其它测量数据系列。故可建立如下模型:

$$Z_t = \alpha_0 + \alpha_1^{(0)} Z_{t-1} + \dots + \alpha_{P_0}^{(0)} Z_{t-P_0} + \alpha_1^{(1)} X_{1,t-1} + \dots + \alpha_{P_1}^{(1)} X_{1,t-P_1} + \dots + \alpha_1^{(s)} X_{s,t-1} + \alpha_{P_s}^{(s)} X_{s,t-P_s} \quad (2)$$

令  $P = \max(P_0, P_1, \dots, P_s)$

在上式中取  $t = P+1, P+2, \dots, n$ , 则可得此  $n-P$  个等式的向量形式为:

$$Z = A \cdot \alpha + \varepsilon \quad (3)$$

式中

$$Z = \begin{pmatrix} Z_{P+1} \\ Z_{P+2} \\ \vdots \\ Z_n \end{pmatrix} \quad \alpha = \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \dots \\ \alpha_s \end{pmatrix}$$

$$\alpha_i = \begin{pmatrix} \alpha_1^{(i)} \\ \alpha_2^{(i)} \\ \dots \\ \alpha_{P_i}^{(i)} \end{pmatrix} \quad i = 1, 2, \dots, S$$

$$\alpha_0 = \begin{pmatrix} \alpha_1^{(0)} \\ \alpha_2^{(0)} \\ \dots \\ \alpha_{P_0}^{(0)} \end{pmatrix} \quad \varepsilon = \begin{pmatrix} \varepsilon_{P+1} \\ \varepsilon_{P+2} \\ \dots \\ \varepsilon_n \end{pmatrix}$$

$$A = (A_0, A_1, \dots, A_s)$$

$$A_0 = \begin{pmatrix} Z_P & Z_{P-1} & \dots & Z_{P-P_0+1} \\ Z_{P+1} & Z_P & \dots & Z_{P-P_0+2} \\ \vdots & \vdots & & \dots \\ Z_{n-1} & Z_{n-2} & & Z_{n-P_0} \end{pmatrix}$$

$$A_i = \begin{pmatrix} X_{i,P} & X_{i,P-1} & \dots & X_{i,P-P_i+1} \\ X_{i,P+1} & X_{i,P} & \dots & X_{i,P-P_i+2} \\ \vdots & \vdots & & \vdots \\ X_{i,n-1} & X_{i,n-2} & & X_{i,n-P_i} \end{pmatrix} \quad i = 1, 2, \dots, S$$

此即转化为一般的线性回归模型进行运算求解。

由式②可知，观测数据 $Z_t$ 不仅与其历史值有关，而且可能与各因子的当时值及历史值有关。

(4) 采用疏系数方法。根据跨断层测量观测数据较多，需估参数较多的特点，采用疏系数方法，用改进的 BIC 准则挑选变元。

其中最高阶 $n$ 阶模型的 BIC 值为：

$$BIC_n = \log(\hat{\sigma}_e^{(n)})^2 + 2n \log N / N$$

$(\hat{\sigma}_e^{(n)})^2$  表示 $n$ 阶模型的拟合残差方差。

假定 $q$ 阶模型的 BIC 值最小，便确定 $q$ 阶模型为最佳，其相应参数为 $\hat{\alpha}_0, \hat{\alpha}_{i_1}, \dots,$

$\hat{\alpha}_{i_p}$ ，即得模型：

$$Z_t = \alpha_0 + \hat{\alpha}_{i_1} X_{i_1,t} + \dots + \hat{\alpha}_{i_q} X_{i_q,t} \quad (4)$$

在模型(4)中，下标 $i_1, i_2, \dots, i_q$ 是 $\{1, 2, 3, \dots\}$ 的不连贯子集，表明此模型为式(2)的特殊形式，用近似最小信息量准则挑选变元，这样做虽然在模型优化程度上做了些让步，但计算量却大为减少。从而使得在有些情形下几乎无法解决的问题得以近似地解决。保证了估计的稳定性，提高了预报的可靠性。

5) 最大迟后量的选取。对于用混合回归方法建立的模型，需要根据因变元的周期特性，适当地选取最大迟后量 $P$ 参加计算。若因变元有明显的周期变化特点，则最大迟后量 $P$

选为比周期大几个单位的数即可。一般若 $P$ 选得过大, 不仅会增大计算量, 而且还可能破坏估计精度; 若 $P$ 选得过小, 不利于充分利用足够的的数据相关性。对跨断层测量中的基线和水准测量数据, 往往由于周期不明显或者说根本不存在所谓周期, 使得对 $P$ 的选取遇到困难。通过大量试算, 发现当 $P$ 取为12时, 即可获得较为满意的模型。

6) 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混合回归模型亦是如此, 因此需要建模型的优度进行检验, 实际上就要检验 $\{\varepsilon_t\}$ 是否白噪声序列。同时由所得模型出发, 进行预报、控制等工作, 由预报或控制效果的好坏, 可进行考核和验证模型的合理性。

### 3.3 拟合预测分析

若在某一时段内运用混合回归方法已建立了相应的模型, 则要看在该时段内各拟合值与实际值的差异如何, 而拟合值的计算由(2)式可知, 其 $l$ 步预测值为:

$$Z_{n+1} = \alpha_0 + \alpha_1^{(0)} \cdot Z_{n+1-1} + \dots + \alpha_p^{(0)} Z_{n+1-p} + \alpha_1^{(1)} \cdot X_{1, n+1-1} + \dots + \alpha_p^{(1)} X_{1, n+1-p} + \dots + \alpha_1^{(s)} X_{s, n+1-1} + \dots + \alpha_p^{(s)} \cdot X_{s, n+1-p} \quad (5)$$

要由此式得到 $l$ 步的预测值, 还需要对自变元向量 $X$ 建立多元自回归模型, 以得到自变元的预测值, 然后代入式(5)中得出因变元 $Z_t$ 的拟合预测值。因此在计算得出疏系数的混合回归模型后, 还需要对上述 $S$ 个自变元分别计算出 $S$ 个相应的疏系数自变元自回归模型, 然后由此 $S$ 个方程组成多元自回归模型。即:

$$X_{n+1} = \alpha + \Phi_1 \cdot X_{n+1-1} + \Phi_2 \cdot X_{n+1-2} + \dots + \Phi_p \cdot X_{n+1-p}$$

但在实际计算分析中, 为简便起见, 同时又能说明一定的问题, 对此处的拟合值中因变元与其本身的过去以及各自变元的关系, 是以其历史值与因子本身的实际值代入, 故此处的拟合有些特殊, 这完全是为了检验模型的拟合度。

类似于回归分析理论, 因变元 $Z_t$ 的方差可分解为:

$$\sum_{i=1}^n (Z_t - \bar{Z})^2 = \sum_{i=1}^n (Z_t - \hat{Z}_t)^2 + \sum_{i=1}^n (\hat{Z}_t - \bar{Z})^2$$

其中, 第一项是残差平方和, 它完全是随机误差的影响, 体现了除了自变元以外其它随机因素对 $Z_t$ 的影响; 第二项是回归平方和, 它虽包含有随机误差, 但主要是模型方程中一次项所引起的变动, 体现了估计量 $Z_t$ 涉及到的自变元对 $Z_t$ 变差的影响。在已建立的跨断层测量的水准和基线数据模型中如何识别异常, 这是一个关键。我们知道当断层处于正常活动期时, 其活动缓慢, 保持原有的活动趋势, 一旦附加应力场改变, 将破坏其趋势活动规律而出现异常活动。一般认为此时断层活动速率加快, 持续一段时间后, 速度减慢或停止或反向而导致地震的发生。因此欲从观测数据中提取断层的异常活动信息, 必须从中识别其正常趋势活动。通常认为在没有任何外界干扰的情况下, 当附加应力场不改变时, 断层基本上朝着一定方向运动, 只是当附加应力场发生变化, 其活动方式、运动方向、速率等才发生变化。因此, 一方面我们可以通过各时段模型本身的不同来划分其正常推和异常期, 另一方面可考察该时段内模型的拟合残差方差与观测数据的的样本方差的比的变化, 设在某时段里因变元观测数据的均方差为 $\sigma$ , 其模型的拟合残均方差为 $\hat{\sigma}$ , 定义 $R = \frac{\hat{\sigma}}{\sigma}$ 则可通过分析在不同时段内 $R$ 的变化情况, 进行判断分析。同时为了考察在同一测点的两条基线之间、两条水准之间

以及基线和水准之间的关系，在计算基线和水准的模型方程时，有意将另一基线与水准作为基线的因子，而将另一水准和基线作为水准的因子，分别得出模型方程，从而看出其相互关系。

具体步骤是按以下方式：

①全线数据模型分析。通过建立全线数据模型，计算残差，绘制相应的残差图，残差的大小反映出异常信息；

②分时段模型分析。首先取前面的若干个数据作为初始数据段，然后每次移动若干个数据点，分别建立相应的数据模型，计算各残差，绘制相应的残差图，求出各时段拟合残差均方差与样本均方差的比值  $R$ ，并通过模型方程和  $R$  的变化进行分析判断。

### 3.4 灰色动态预测模型

灰色系统理论自82年提出以来，已有了广泛的应用，它把一切随机变量看作是在一定范围内变化的灰色量，将随机过程看作是在一定范围内变化的、与时间有关的灰色过程，通过灰数生成杂乱无章的原始数据中去开拓、发现、寻找这种内在规律。对运用混合回归理论建立的动态数据模型，用灰色系统的 GM (1, 1) 模型对上述动态数据模型进行局部修正，使得修正后的模型在短期预报中能有较好的精度。

设用3.2所述方法建立了模型方程：

$$Z_t = F(Z_{t-1}, X_{R, t-P_s}) \quad (6)$$

其中：  $0 \leq l \leq P$ ,  $0 \leq P_s \leq P$ ,  $1 \leq t-l \leq N$ ,  $1 \leq t-P_s \leq N$

通过计算得到用 GM (1, 1) 模型修正后的方程：

$$Z_t = F(Z_{t-1}, X_{R, t-P_s}) + \delta(t-1) \cdot (-a) \cdot (X\{i\} - \frac{u}{a}) \cdot e^{-a(t-i+1)} \quad (7)$$

其中：  $0 \leq l \leq P$ ,  $0 \leq P_s \leq P$ ,  $1 \leq t-l \leq N$ ,  $1 \leq t-P_s \leq N$

$$\delta(t-i) = \begin{cases} 1 & t \geq i \\ 0 & t < i \end{cases} \quad \{X\{i\}\} \text{是残差序列}$$

(7) 式即为灰色动态预测模型。

实际计算表明在某些情况下，修正后的模型精度有所提高。

## 4 实例分析及结论

本文选取了北京大灰厂测点和四川鲜水河测点的水准测量和基线测量数据进行了大量的计算分析，其中1976年7月28日发生过唐山大地震，1981年在鲜水河地区的道孚也发生过地震，计算程序流程图见图1：

(1) 测点全线数据模型分析。计算得出拟合残差均方差

与样本均方差之比  $R = \frac{\hat{\sigma}}{\sigma}$  的值很小，表明基于混合回归理

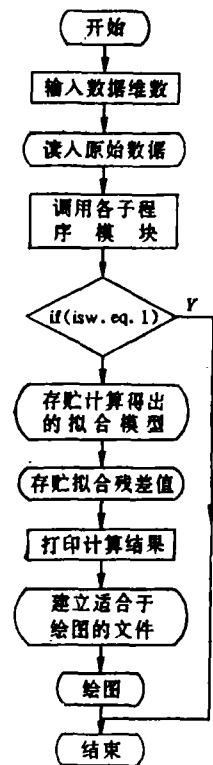


图 1

论的数据模型有较好的拟合效果,且在  $R$  的变化中,一定程度上能反映地震的异常。

(2) 测点各时段模型分析。取整个数据段  $N$  个数据的前面  $M$  个数作为初始数据段,进行计算,然后样本仍保持  $M$ ,每次向右移动  $P$  个点,对由此得到若干个子序列重复进行建模拟合。

通过比较各时段的计算结果,表明各时段的模型方程、残差及  $R$  的变化能较好地反映地震异常的出现,可以看到它与地球内部应力变化规律的联系,说明它对地震有较好的前兆反映。同时在不同时段基线测量之间、水准测量之间、以及基线测量和水准测量之间有不同的关系规律。用 GM(1,1) 模型对残差进行修正的灰色动态数据模型有较好的拟合精度,适应于在一定的条件下对基线和水准测量数据进行短期预报。当然这种方法的应用还处于初始阶段,还应从不同角度做更进一步的探索。

### 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] George E F BOX, Gwilym Jenkins. TIME SERIES ANALYSIS—forecasting and control. Holden—Day Inc. San Fransisco, 1976.

## Research on the Dynamic Data Model of Deformation Measurements Across Fault

Zhang Xingfei

**【Abstract】** In the paper, on the basis of known major atmospheric parameters, for example, air and underground temperature, amount of raintall, etc, how to use the method of time series analysis to process the data of observation across fault is emphatically disscussed. Through two practicality, detailed research about constructon of mixd linear regression equation, the relation between the model equation and the earthquake anomaly, processing and analysis of the observational data with possibly has anomaly is made. Besides, the grey dynamic data model that the residual is adjusted by GM(1,1) is given. Through practicality, it indicates that the metod mentioned will show its advatages in data proessing of deformation measurements across fault.

**【Key words】** deformation measurements; time series; mixed regression; sparse coefficient; maximum delay; grey model