

文章编号: 0253-3782(2000)06-0606-08

地震前兆异常识别技术的 数值模拟实验研究^{*}

严尊国¹⁾ 钱家栋²⁾ 陈俊华¹⁾ 李胜乐¹⁾

1) 中国武汉 430071 中国地震局地震研究所

2) 中国北京 100036 中国地震局分析预报中心

摘要 针对地震前兆观测数据可能含有异常信息的特点, 为客观描述数据序列的正常背景, 并且不失真地识别异常信号, 本文通过模拟实验序列, 论述了时间序列分析识别异常方法中建模样本序列遴选和经验预测序列估计的关键技术。实验序列分析结果表明, 该方法是可行的有效的, 所识别的异常信号清晰明确, 与模拟假设的异常信号基本一致。

关键词 地震前兆 时间序列 异常识别

中图分类号: P315.72

文献标识码: A

引言

一般认为, 地震孕震演化过程及其前兆的发展都是一种时间进程, 在数学上可以描述为时间变量函数。然而, 实际研究中还没有找到一种可用来作为描述地震前兆的物理量或数学量, 它们随时间的变化仅与孕震演化过程有关而与其它“非震过程”无关。也就是说, 用来描述前兆时间进程所观测的物理量或推演的数学量 $F(t)$, 其中不仅可能含有可能与地震孕育演化过程有关的前兆变量 $x(t)$, 而且还可能包含那些可能与地震孕育演化过程无关的部分, 我们称之为“非震变化”。这些“非震变化”可以分类归纳为: 观测物理量自身的确定性规律变化 $u_1(t)$, 其一般与地震无关; 机理清楚的非震因素影响产生的变化 $u_2(t)$; 未知或机理不清楚的非震因素影响产生的变化 $u_3(t)$; 观测量的本底噪音 $v_1(t)$ 和其它随机偶然性干扰引起的变化 $v_2(t)$ 。各量之间则相互独立。这样, 地震前兆观测量在时间上可以表达为

$$F(t) = u_1(t) + u_2(t) + u_3(t) + v_1(t) + v_2(t) + x(t) \quad (1)$$

地震前兆异常识别的基本任务是从观测量中寻求可能与地震孕育演化过程有关的前兆部分, 即从观测量中识别出“非震变化”和“前兆变化”两大部分。具体地说, 就是对式(1)右边前 3 项变化量建立“正常变化模型”, 并从观测量中消除掉, 以及从消除正常变化后的数据序列中识别“前兆变化”异常。显然, 对于某些观测量, 经过适当的预处理方法可以较好地消除已知的 $u_1(t)$ 及 $u_2(t)$ 的相关影响。但是, 一般说来, 大部分前兆观测量自身变化规

* 地震科学联合基金“九五”重点项目(95-07-431)资助。

2000-03-21 收到初稿, 2000-08-15 收到修改稿并决定采用。

律、非震影响量以及多种非震影响量之间的相关性尚不完全清楚。因而，经过预处理后的数据序列不完全是由“干净”的 $v_1(t)$, $v_2(t)$ 和 $x(t)$ 三项组成。认识观测数据正常变化背景始终是前兆识别技术中的基本问题。

已有的研究表明，在现有观测条件下，利用时间序列分析方法，通过观测数据建立正常背景模型，离析出可直观识别异常信号的数据序列是前兆识别技术发展的有效途径之一，常用的 AR, ARMA, ARBUG 等时序分析方法均能较好地适应前兆数据的分析，并且应用效果差别不大(Zhen et al., 1997; 杨叔子等, 1991a)。但是，这些研究与前兆数据分析的其它方法一样，较少注意到混杂在数据中的未知异常信号的影响，也没有论证所识别信号的真伪。而在前兆异常识别研究中，不仅要注重选用适当的参数模型以描述数据本身的结构与规律，而且更需关注数据中所存在的违背数据自身规律的异常信号对探求规律特征的影响。因此，本研究不再讨论时序分析方法运用的适应性，而是针对前兆数据中可能含有异常信号的特点，通过人工构建适当的模拟序列的数值模拟方法，重点论证在存在(或“混入”)异常信号的模拟序列中前兆数据分析方法的可行性，讨论运用其处理前兆数据的关键技术，以及论证用其发现(或检测)异常信号的可靠性。

1 前兆数据时序分析识别异常技术的关键问题

如果前兆数据中可能包含未知的异常及其它干扰等信号，将严重影响时序分析中正常变化背景模型的建立和经验预测估计的客观性。因此，从一个可能含有未知异常及其它干扰信号的前兆时间序列中，遴选出不含异常信号的相对“干净”的建模样本序列，进而对可能的异常信号作出“相对准确”的经验预测估计，是前兆数据处理的基本特点和要求，也是前兆识别中的关键技术。

1.1 建模样本数据序列遴选方法

假设在正常情况下，所分析的数据序列满足平稳正态过程，那么，根据建模样本序列的要求，它应该是从数据序列中遴选出的充分长的满足平稳性假设的子序列。

根据平稳随机过程的定义，平稳时间序列 $\{X_i\}$ 的均值 μ_x 和方差 σ_x^2 为常数；序列的自协方差函数 R_k 只与时间间隔 k 有关，而不依赖于序列的时间总长。依据平稳随机过程的这些性质，常用的检验数据序列平稳性的方法，主要有分段检验法、逆序检验法和自协方差函数检验法(杨叔子等, 1991a)等。这里主要应用分段检验法和自协方差函数检验法。

(1) 分段检验法。将数据序列均匀地分为 l 个子序列。如果数据序列是平稳时序，则各子序列的均值、方差和自协方差函数的估计值不应有显著差异。反之，如果某些子序列估算值与其它子序列存在显著差异，则表明数据序列整体上拒绝平稳，最多是分段平稳。这样，可以利用两个相邻子序列间的统计特性(均值 μ_j 、方差 σ_j 、自协方差 $R_{j,k}$)，采用数理统计中的假设检验来检验这种差异，遴选出其中差异较小、并且相邻的若干个子序列，合并为建模样本序列。

分段检验法概念简单，但计算繁琐。对于不同观测资料，统计检验的显著水平假设可能不同。本研究统计检验的显著性水平取为 0.05。

(2) 自协方差函数检验法。前兆数据序列中选取长为 n 的子序列，如果其自协方差函数 R_k 随时间间隔 k 的增大而向 0 衰减，则认为所检验的序列是平稳的，反之，则拒绝平稳。

1.2 经验预测数据序列的“无偏”估算

经验预测数据序列，是根据时序分析所得到的模型，利用原始数据序列值递推而求得的数据序列。原则上，经验预测数据序列是模型的一步预测结果，它描述前兆数据变化的正常背景。在前兆数据中，对于完全正常态变化的数据，模型的一步预测值能够较好地描述其正常背景；但是在前兆数据中含有异常信号时，一旦异常值代入模型递推，预测估计将失真，估算值将偏离正常背景的真值。因此，要消除可能的异常信号对预测估计的影响，需要对可能含有的异常作出初步判定，并且采用适当的数据处理方法，尽可能地进行合理的无偏预测估计。

(1) 预测估计中可能异常信号的初步判定。时间序列分析的模型定阶时，一般采用最终预测误差准则(FPE)，或信息准则(AIC , BIC)(杨叔子等, 1991a)。不同准则在前兆数据序列的处理中对建模结果影响差别不大，以 FPE 准则为例，最终预测误差平方表示为

$$(FPE)_n = \frac{n + (k + 1)}{n - (k + 1)} S_n^2 \quad (2)$$

式中， n 是建模样本序列长度， k 是模型阶数， S_n^2 是 $[n - (k + 1)]$ 个模型拟合值与观测数据的残差平方和。

这样我们可以用最终预测误差作为可能异常信号的判据。如果模型预测值与原始数据之差超过上述判定标准，则认为这时前兆数据中出现可能异常值。

(2) 出现可能的异常信号时的经验预测估计。依据最终预测误差判据，当数据序列出现可能的异常信号时，采用以一步预测值替代异常值继续进行预测估计，是一种简洁的“替代法”。

这里，“可能的异常信号”是指 t 时刻的观测数据 X_t 与模型预测值 Φ_t 之间出现 $|X_t - \Phi_t| \geq \sqrt{(FPE)_n}$ (或取 $2\sqrt{(FPE)_n}$) 情况时的数据信号。这时以 Φ_t 替代 X_t 值预测估计($t+1$)时刻的 Φ_{t+1} 值。当继续判定 X_{t+1}, X_{t+2}, \dots 仍是可能异常信号时，则以此类推。

当“可能的异常信号”时间尺度远小于模型定阶数时，利用替代法估计的模型预测序列有较好的结果。但是，当可能异常信号的时间尺度较大时，替代步数增大(相当于多步预测)，这时最佳长期预测值将渐进等于时间序列的均值(杨叔子等, 1991b)，预测准确性下降。也就是说，当可能异常信号时间尺度较大时，对其后期的正常背景预测估计欠佳。

另一种办法是最佳相关子序列替代方法：利用与出现可能异常信号前的一定长度子序列的最佳相关子序列的续后值替代可能异常信号的方法。

若 t 时刻观测值 X_t 被初步判定为“可能的异常信号”，截取 t 之前长度为 m (m 大于模型的阶) 的子序列 $X_{t-m}, X_{t-m+1}, \dots, X_{t-1}$ ，并选取若干位于时刻 $t-m$ 之前、长为 m 的过去子序列进行相关分析，选择其中最佳相关的过去子序列，以其后续的第一个数据替代 X_t 进行预测递推估计。当连续出现可能异常信号时，则将最佳相关过去子序列续后的相应值依次作替代。

2 数值模拟实验数据序列的构建

基于上述分析，本文的研究思路如下：第一，构建一个实验数据序列 $F(t)$ ，分别由满足式(1)右边各分项特征的人为的时间序列构成；第二，按上节提出的方法，检验从中提取

异常信号 $x(t)$ 的能力和条件.

首先讨论实验数据序列 $F(t)$ 的构建. 式(1)右边各项特征分述如下:

(1) 异常信号 $x(t)$. 在实际观测中, 前兆异常形态十分复杂, 不同观测量间的差别很大, 多样性是其主要特点. 但绝大多数异常是以其幅度随时间的变化来描述的. 总的说来可以分为两大类: 一类观测量在异常变化结束后回复到过去正常的背景状态; 另一类观测量在异常结束后观测值不再回复, 而明显进入一个新的背景状态为标志(类似阶变性质). 对于后一类异常变化, 原则上应利用新的观测数据重新进行时序分析. 这里暂考虑第一类异常情况. 在异常形态上简化为单一性上升、下降(或下降、上升)和起伏性上升、下降(或下降、上升)两种典型变化(图 1). 图 1 中横坐标是时间, 纵坐标是观测量的幅值. 在这里以长度量 $A(\text{cm})$ 示意描述, 对于不同前兆观测量, 将根据观测的物理量及其度量确定.

(2) 周期性信号 $u_1(t), u_2(t), u_3(t)$.

式(1)中这 3 项是数据正常变化背景, 可以合并为一个函数构建. 尽管其函数形式是未知的, 但是总能表述为若干不同周期性的正弦级数之和. 为充分反映前兆数据的复杂性, 本研究以 20 个不同周期的正弦级数叠加信号来描述. 每个正弦函数的周期、相位和幅度系数均分别随机确定.

具体确定时, 模拟实验时间序列样本长度以 200 为例, 首先由计算机随机函数输出 3 组随机数, 令各组中最大值分别对应于最大周期值(100 单位时间)、最大相位值(π)和最大幅值(单位长度 1 cm), 相应组中其它各随机数对应的周期、相位和幅度系数值, 由各随机数与最大随机数的比值大小, 按比例确定.

(3) 噪声背景 $v_1(t)$. 直接由随机序列构建.

(4) 偶然性干扰信号 $v_2(t)$. 在实验序列中以脉冲信号表现, 出现的时间与强度均随机确定, 其出现次数控制在全序列样本数的 2% 以下. 当然, 脉冲性偶然“干扰”信号在某些场合也许是某种异常信号.

模拟实验时间序列样本长度以 200 为例, 异常信号时间尺度为 10~40, 一个序列中允许出现 1~2 个时段异常信号, 异常与非异常信号(含周期性信号和噪声背景)的幅度比限制在 3~5 倍的范围内^①(梅世蓉等, 1993).

本研究按上述方法构建了近百个模拟数据序列, 图 2 仅例举了两个. 图 2 例 1 中的第一条曲线, 除周期信号外, 叠加有 2 个脉冲信号、1 段上升一下降异常信号、1 段下降一上升异常信号和噪声背景信号; 图 2 例 2 中第一条曲线含有的 1 段异常信号是起伏性下降一上升形态.

3 模拟实验数据序列分析与结果

模拟实验数据序列构建时未涉及趋势性变化等信号, 因此, 不再进行通常前兆数据处理中的零均值化和提取趋势项等预处理.

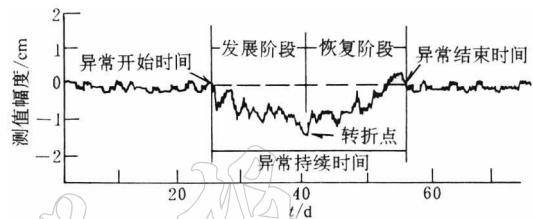


图 1 典型异常信号示意图(据高旭等, 1984)

^① 孙加林. 1997. 关于地震短临跟踪工作的几个问题. 震情研究, (1): 90~93.

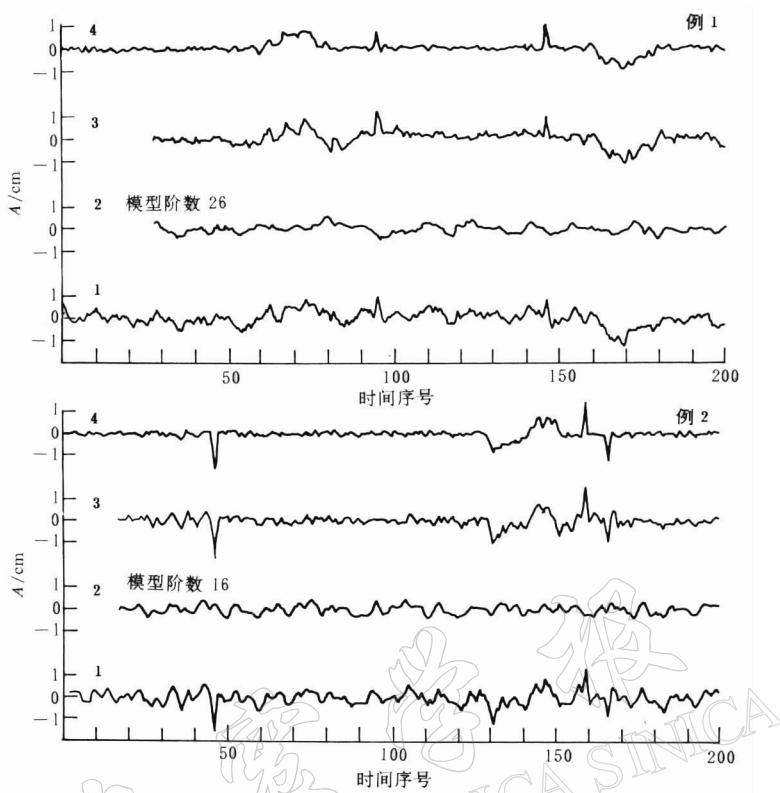


图 2 模拟实验时间序列分析例图

1. 前兆模拟数据序列；2. 经验预测序列；3. 异常识别序列；
4. 假设的已知异常信号、脉冲性干扰信号和噪声背景信号叠加序列

(1) 遴选建模样本序列. 模型阶次越高越能准确逼近真实系统, 但是参数增多会导致计算误差. 对于构建实验序列的建模阶数大多在 20 ± 10 阶范围, 因此建模样本长度至少在 60 以上.

采用分段检验法时, 分段子序列长度分别选取为 10 和 20 进行比较, 大多实验序列能够遴选出足够长的建模样本序列. 部分实验序列中通过检验的连续子序列较短, 主要原因是偶然性脉冲干扰的影响. 对于这些序列, 如果有脉冲信号, 利用其前后的数据进行二项式内插处理, 仍然可以得到足够长的有效建模样本序列.

采用自协方差检验法时, 为了充分利用数据资源, 保证所建模型的客观性, 总是希望建模样本序列达到最长, 因此不得不在数据序列中不断延长子序列的长度反复检验. 实际上, 只要能够满足建模所需的足够长的平稳数据序列, 大部分建模结果对数据序列的长度和起止时间要求不很严格. 为避免过多的盲目估计, 可以直观图形判断数据序列较长的可能平稳段进行检验. 另外, 为保证其估计的有效性, k 值不能取得过大, 一般取 $k_{\max} \leq n/4$, 通常只取 $k_{\max} = n/10$. 两种检验方法均适用于实验数据分析, 虽然遴选出的建模样本序列不一定相同, 并且建模参数也略有差异, 但对数据正常背景的描述及对异常的识别影响不大.

(2) 时间序列分析方法. 主要采用自回归模型(AR), 对少数序列利用自回归滑动平均模型(ARMA). 实验表明, AR 模型已经足以描述实验序列的正常背景, 比较简单. 其中, 模型定阶采用最终预测误差准则(FPE).

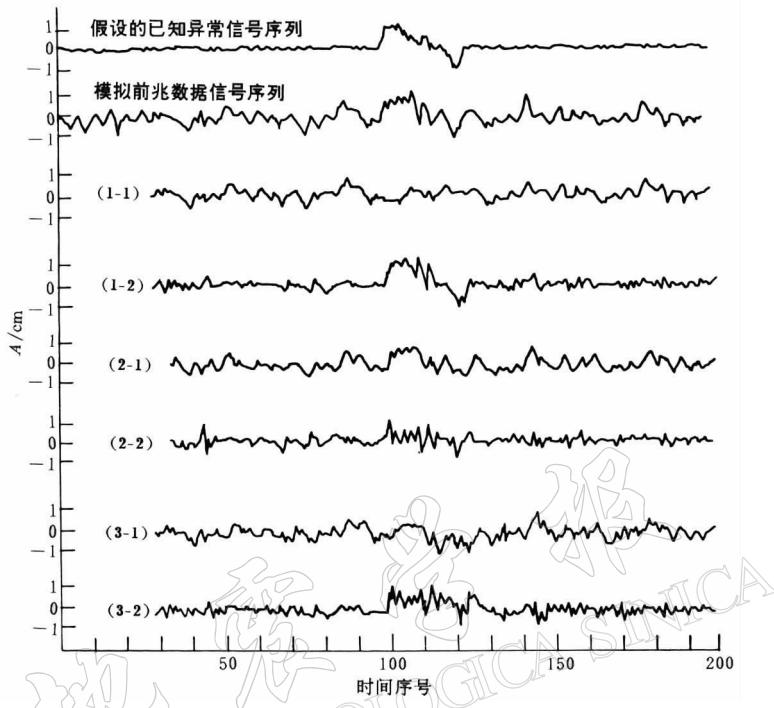


图 3 异常信号对时间序列分析结果影响的比较

(1-1) 本研究模型预测序列; 正常背景; (1-2) 模拟序列与序列(1-1)的差值: 低噪声背景下的异常识别;

(2-1) 不考虑模拟序列中异常信号的建模预测序列; (2-2) 模拟序列与序列(2-1)的差值; (3-1) 利用模拟序列中平稳样本序列建模, 但预测估计时忽略了异常信号的影响; (3-2) 模拟序列与序列(3-1)的差值

(3) 经验预测序列估算. 当递推预测估算值涉及初步判断的可能异常值时, 采用递推预测值替代或最佳相关子序列替代的方法继续递推估算. 两种方法对于判断异常信号的起始和基本形态效果差别不大. 而当异常信号较长时, 最佳相关子序列替代方法得到的异常后期形态的失真程度低于递推预测值替代方法, 但是最佳相关子序列替代方法计算量明显增大. 图 2 仅示例了两个典型的模拟实验序列分析结果. 图中曲线 1 是构建的实验数据序列; 曲线 2 是建模估算的经验预测序列; 曲线 3 是曲线 1, 2 两序列的差值序列, 即异常识别序列; 曲线 4 是假设的已知异常信号、脉冲性干扰信号和噪声背景信号叠加序列.

对构建的数十个实验序列分析结果表明: ① 实验模拟序列中无论是可直观的, 或不可直观识别的不同形态异常(图 2 中曲线 1), 利用时序分析方法均能有效提取. 最终输出异常识别数据序列上异常信号直观清晰. 一般讲, 不同分析人员都能够一致地提取信号的有关特征量; ② 在异常识别序列上所识别的异常信号与假设的已知异常信号, 在起始与截止时间、幅度和形态上基本一致(图 2 中曲线 3, 4), 大部分实验序列这两组信号的相关系数在 0.90 以上. 异常初期形态一致性较好, 只是当假设异常信号持续时间较长时, 其后期形态略有畸变; ③ 对于脉冲性偶然干扰(异常)信息, 在输出的异常识别序列上所识别的信号

与输入实验序列中已知信号的大小、起始及截止时间和形态等特征几乎完全一致。

为了进一步说明前兆数据中异常信号对数据处理的影响，实验中同时比较了忽视建模样本平稳性检验和模型预测估计中忽略存在异常信号影响的情况(图3)。图中(1-1)、(1-2)是依照本研究方法得到的结果；(2-1)、(2-2)是同时忽视建模样本平稳性检验和在模型预测估计中忽略存在异常信号影响的结果；(3-1)、(3-2)是遴选出平稳样本序列建模，但模型预测时忽略了可能存在异常信号影响的情况。很清楚，(1-2)中可直观识别的异常信号与假设的已知异常信号基本一致，而(2-2)中识别异常信号比较困难，(3-2)中所识别的异常与假设的已知异常信号相比明显失真。这一结果说明，对可能含有异常信号的数据序列适当处理后，时间序列分析方法得到的低噪声背景下直观识别的异常信号是有效可靠的，同时也说明忽视了前兆数据中含有异常信号的特点，简单地引入较成熟的数据处理方法，可能不会达到预期的处理效果。

4 结论

含有异常信号是实际前兆观测数据序列的特点。在数据处理技术中忽视了这些可能的异常量的影响，必将使异常识别产生不同程度的畸变，有时这种畸变甚至还会十分严重(图3)。而忽视数据中可能的异常量的影响，恰恰是前兆数据分析中比较普遍的现象。显然，寻求一种较完善的前兆数据处理技术，达到既能描述数据序列的正常变化背景，又能准确提取异常信号的要求，已经成为前兆研究中一个十分重要而且十分迫切的课题。在这一方面，本文关于地震前兆异常识别技术的数值模拟实验研究的思路和方法，可能有一定的借鉴和参考作用。

本模拟实验研究的结果给出了一个例证，利用时间序列分析方法，在遴选建模样本序列和估算经验预测序列时采用适当的技术处理，有可能有效地解决这类问题。从实际的前兆数据序列中识别出异常信号，从而为前兆定量研究提供了条件。当然，本文论述只是示范性的，采用人工模拟的实验序列进行数值模拟，只是实际观测数据的一般性的简化近似，目的在于阐述所研究问题的必要性、研究思路和方法以及意义。不言而喻，实际数据序列要复杂得多。因此，在处理实际前兆数据时，还需对数据特征做具体分析，还将涉及诸如分段检验法中显著水平、估算经验预测序列中异常信号初步判断的阈值[本文取 $2(FPE)$]等参数的选择等具体问题的讨论。特别是对于实际观测数据中可能含有多种时间尺度的异常变化(如较高频异常或较低频的趋势异常)的情形，还需要考虑将观测数据序列分解为两个或若干个不同频带的序列，分频段进行时序分析和异常识别。有关这些问题将在另文中专门阐述。

参 考 文 献

- 高旭, 邱竞男, 姜秀娥, 等. 1984. 中国地震前兆资料图集[M]. 北京: 地震出版社, 383~384
- 梅世蓉, 冯德益, 张国民, 等. 1993. 中国地震预报概论[M]. 北京: 地震出版社, 249~385
- 杨叔子, 吴雅稚, 王治国, 等. 1991a. 时间序列分析的工程应用(上册)[M]. 武汉: 华中理工大学出版社, 243~298, 383~384
- 杨叔子, 吴雅稚, 王治国, 等. 1991b. 时间序列分析的工程应用(下册)[M]. 武汉: 华中理工大学出版社, 7~10
- Zhen Wenheng, Chen Junhua, Yan Zunguo, et al. 1997. Using an artificial intelligence method to recognize seismic precursor anomaly[J]. *Earthquake Research in China*, 11(1): 99~107

STUDY OF TECHNIQUE OF IDENTIFYING THE EARTH- QUAKE PRECURSOR ANOMALOUS IN TERMS OF MATHEMATIC MODELING

Yan Zunguo¹⁾ Qian Jiadong²⁾ Chen Junhua¹⁾ Li Shengle¹⁾

1) Institute of Seismology, China Seismological Bureau, Wuhan 430071, China

2) Center for Analysis and Prediction, China Seismological Bureau, Beijing 100036, China

Abstract: This paper deals mainly with the key technique of identifying the anomalous signals without distortion, which might be the precursors associated with earthquakes from the real time series of observations, which would be usually the mixture of the anomalous signals and the normal background variations, some interference and noises.

The key technique of “no-biased estimation” is to construct an empirical time series and set up the creation for identifying the anomalous variation on the bases of time series analysis.

To the end of testing the method, a man-made time series including the normal variations and random interference as well as specific anomaly, has been constructed. And the test of picking up the anomaly has been conducted with the feasible and effective way of identifying the anomalous signal from a complicated time series.

Test results confirm that the technique under discussion are effective and applicable, and the signals extracted from the analysis, could be clear and precise, and is almost similar to the known simulated anomalous signals in the experiments.

Key words: earthquake precursor; time sequence; anomalous identification