

# 主成分分析法在辽宁地区地震预测中的应用

李永振

(辽宁省地震局, 辽宁 沈阳 110034)

**摘要:**采用多因子分析的主成分分析法,选择与地震活动强度有关的地震频次  $N(M_L \geq 3.0)$ 、 $b$  值、 $\eta$  值、 $M_f$  值、 $A(b)$  值和  $A_c$  值等 6 个参量进行分析,实现对上述参量的有效约简,得到反映地震强度特征的综合指标  $W$ 。发现指标  $W$  在辽宁海城 7.3 级、岫岩 5.4 级地震前出现明显的异常变化,震后恢复。表明综合指标  $W$  可以较好的反映地震活动的异常特征。

**关键词:** 辽宁; 地震活动强度参数; 主成分分析法; 相关性; 贡献率

中图分类号: P315.51

文献标识码: A

文章编号: 1000-0844(2011)01-0076-04

## Application of Primary Components Analysis Method to Earthquake Prediction in Liaoning Province

LI Yong-zhen

(Earthquake Administration of Liaoning Province, Liaoning Shenyang 110034, China)

**Abstract:** Using the primary components analysis method, six main parameters for indicating seismicity strength such as frequency  $N(M_L \geq 3.0)$ ,  $b$  value,  $\eta$  value,  $A(b)$  value,  $M_f$  value,  $A_c$ -value are selected to analyse. The parameters are simplified effectively and a new synthetic index  $W$  for indicating seismicity strength is obtained. It is found that the index  $W$  appeared obvious anomalies before Haicheng  $M_s 7.3$  earthquake in 1975 and Xiuyan  $M_s 5.4$  earthquake in Liaoning province, and returned after events. It is shown that the synthetic index  $W$  can reflect anomalous characteristics of seismicity well.

**Key words:** Liaoning province; Seismicity strength parameter; Primary components analysis method; Relativity; Contribution rate

## 0 引言

当今世界各类数据量正在以指数增长。数据挖掘<sup>[1]</sup>就是在数据库中对数据进行一定的处理,从大量不完全的、有噪声的、模糊的、随机的数据中提取隐含的、事先未知的、但又是潜在有用的信息和知识的过程。在数据挖掘中数据清洗是一个重要环节,包括去噪声,填补丢失的域,删除无效数据,对时序数据的整理和归并,以及数据属性的约简等。王伟等<sup>[2]</sup>就用主成分分析法对地震活动性参数进行了约简,同时用该方法在江苏溧阳 6.0 级地震中进行了应用<sup>[3]</sup>。

主成分分析方法是描述样本特征的多个可能

有一定相关性的指标化为少数几个综合指标的一种统计分析方法,能够在最大限度地保留原有信息的基础上对高维变量系统进行最佳地综合与简化,并能够客观地确定各个指标的权数,避免了主观随意性。应用主成分分析法可以找出由若干个指标线性组合而成的综合指标,即若干个主成分,同时彼此之间相互独立。

目前在地震活动分析中存在许多预报指标,这类参数数量较多,同时它们之间还可能存在一定的相关性<sup>[4-5]</sup>。另外在实际预报中常常出现有些参数在一些中强以上地震前出现较明显的异常,而另一些参数并不出现异常。这些都给实际预报带来困

难。为解决预报参数过多而造成预报意见的不一致。本文选用 1970 年以来辽宁地区 2 次 5 级以上地震前后震中附近地区的地震资料,选择与地震活动强度有关的 6 个参量,即 3 级以上地震的频次  $N$ 、 $b$  值、 $\eta$  值、 $A(b)$  值、 $M_f$  值和  $A_c$  值进行主成分分析,以期得到一个反映地震强度异常的综合指标。

## 1 主成分分析法

主成分分析法旨在力保原始数据信息丢失最小的情况下,对高维变量空间进行降维处理,经过线性变换和舍弃部分信息,以少数的综合变量取代原有的多维变量。

设原始变量为  $x_1, x_2, \dots, x_p$ , 进行主成分分析后得到的主成分(综合变量)为  $z_1, z_2, \dots, z_m$  ( $m < p$ ), 它们是  $x_1, x_2, \dots, x_p$  的线性组合 ( $m < p$ )。新变量  $z_1, z_2, \dots, z_m$  构成的坐标系是在原坐标系经平移和正交旋转后得到的,称其空间为  $m$  维主超平面。在主超平面上第一主成分  $z_1$  对应于数据变异(贡献率  $e_1$ ) 最大的方向,对于  $z_1, z_2, \dots, z_m$ , 依次有  $e_2 \geq \dots, \geq e_m$ 。因此,  $z_1$  是携带原始数据信息最多的一维变量,而  $m$  维主超平面是保留原始数据信息最大的  $m$  维子空间。

主成分分析法的步骤如下:

(1) 为了排除数量级和量纲不同带来的影响,首先对原始数据进行标准化处理:

$$x'_{ij} = \frac{x_{ij} - \bar{x}_i}{\sigma_i} \quad i = 1, 2, \dots, n \quad (1)$$

式中,  $x_{ij}$  为第  $i$  个指标第  $j$  个样本的原始数据;  $\bar{x}_i$  和  $\sigma_i$  分别为第  $i$  个指标的样本均值和标准差。

(2) 根据标准化数据表  $(x'_{ij})_{p \times n}$ , 计算相关系数矩阵  $R = (r_{ij})_{p \times n}$ 。其中

$$r_{ij} = \frac{1}{n} \sum \frac{(x_{ij} - \bar{x}_i)(x_{ij} - \bar{x}_j)}{(\sigma_i \sigma_j)} \quad (2)$$

(3) 根据特征方程  $|R - \lambda I| = 0$ , 计算  $R$  的特征根  $\lambda_i$ , 并使其从大到小排列:  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$ , 同时可得对应的特征向量  $u_1, u_2, \dots, u_p$ , 它们标准正交。  $u_1, u_2, \dots, u_p$  称为主轴。这里  $I$  为单位矩阵。

(4) 计算贡献率

$$e_i = \lambda_i / \sum_{k=1}^p \lambda_k \quad (3)$$

和累计贡献率

$$E_m = \sum_{k=1}^m \lambda_k / \sum_{k=1}^p \lambda_k \quad (4)$$

(5) 计算主成分

$$Z_m = \sum_{j=1}^p u_{mj} x_j \quad (5)$$

(6) 一个  $m$  维主超平面究竟以多大的精度来近似代替原始变量系统才能确保尽可能多的原始数据信息? 这可以通过求累计贡献率  $E_m$  来判断。一般  $E_m > 85\%$  的最小  $m$  ( $m < p$ ), 则可得主超平面的维数  $m$ , 从而可对  $m$  个主成分进行综合分析。

(7) 根据主成分分析得到的主成分  $Z_i$  和相应的权值(贡献率)  $e_i$ , 计算本文定义的反映地震活动时、空、强异常特征的综合指标

$$W = \sum_{i=1}^m e_i Z_i \quad (6)$$

由于  $m$  个主成分已基本保留了这些预报参数的信息, 所以综合指标  $W$  包含了这些参数从不同侧面反映地震活动时、空、强异常的基本特征。本文选择了从不同侧面反映地震活动时、空、强特征的一些参量: 地震频次  $N$  ( $M_L \geq 3.0$ )、 $b$  值、 $\eta$  值、 $M_f$  值、 $A(b)$  值、 $A_c$  值 6 个参量进行主成分分析。

## 2 地震实例

### 2.1 1975 年海城 7.3 级地震

1975 年 2 月 4 日辽宁海城发生  $M_s 7.3$  地震。图 1 为地震发生前后震中周围 300 km 范围内的一些地震活动性参数随时间变化曲线。除地震频次  $N$  ( $M_L \geq 3.0$ ) 外, 其它参数计算所选取的起始震级为  $M_L 2.0$ 。参数计算累计时间为 1 年, 滑动步长为 1 个月。从图 1 可以看到, 7.3 级地震前这些参数各有所异, 大多参数在这次地震前的异常变化并不显著。

表 1 为通过主成分分析方法得到的上述参量在各主成分中的系数(特征向量)、特征值与贡献率。当取 3 个主成分时累计贡献率已达到 93%; 当取 4 个主成分时累计贡献率已达到 98%。这表明取前 3 个主成分已包含了样本中的绝大部分信息量。由表 1 可知, 主成分 1 中的  $A(b)$ 、 $M_f$  值系数较大, 是构成主成分 1 的主要参数; 主成分 2 中的地震频次  $b$  值、 $\eta$  值系数较大, 是构成主成分 2 的主要参数; 主成分 3 中的  $A_c$  值、 $N$  值 ( $M_L \geq 3.0$ ) 系数较大, 是构成主成分 3 的主要参数。

本文取前 3 个主成分根据式(5)计算地震综合指标  $W$ 。图 2 反映了海城 7.3 级地震前后震中附近地区地震活动时、空、强异常特征的综合指标  $W$  随时间的变化。可以看到, 在 1975 年海城 7.3 级地震前 0.5 年左右时间海城震中附近地区  $W$  值出现

明显抬升的形态,异常幅度近 1.0,震后异常逐渐恢

复到正常范围内波动。

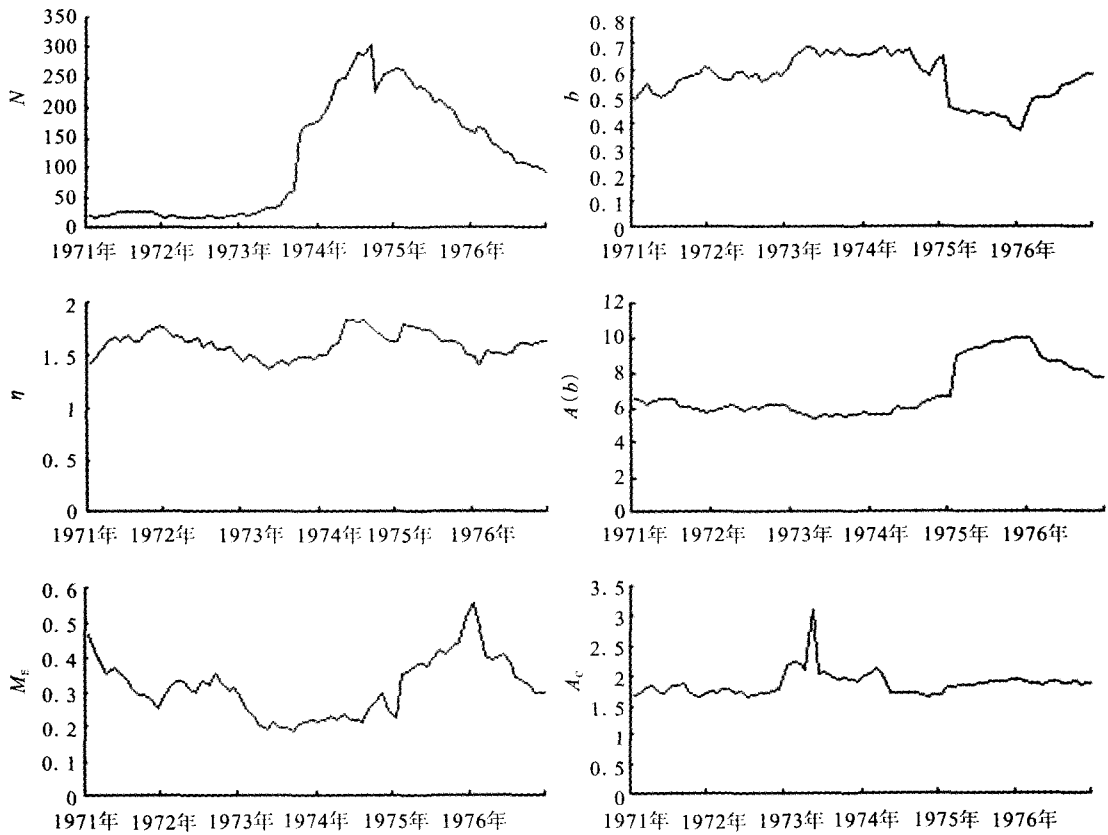


图 1 海震地震前各地震参数曲线

Fig. 1 Curves of different seismic activity parameters before Haicheng  $M_s7.3$  earthquake in 1975.

表 1 各参量在各主成分中的系数(特征向量)、特征与贡献率

| 参量      | 主成分 1    | 主成分 2    | 主成分 3    | 主成分 4    | 主成分 5    | 主成分 6    |
|---------|----------|----------|----------|----------|----------|----------|
| $N$     | 0.459 9  | 0.338 7  | 0.486    | -0.097 2 | 0.042 9  | -0.652 9 |
| $b$     | -0.308 1 | 0.519 5  | 0.188 8  | -0.551 4 | 0.450 3  | 0.304 6  |
| $\eta$  | -0.019 4 | -0.632 7 | 0.464 9  | 0.090 4  | 0.611 5  | 0.031    |
| $A(b)$  | 0.550 2  | -0.023 5 | 0.388 8  | -0.095 6 | -0.321 7 | 0.657 9  |
| $M_f$   | 0.459 4  | 0.319 7  | -0.369 3 | 0.465 9  | 0.548 1  | 0.181 3  |
| $A_c$   | -0.423 6 | 0.335 1  | 0.473 7  | 0.672 4  | -0.132 3 | 0.119 3  |
| 特征值     | 3.460 7  | 1.399 9  | 0.747 1  | 0.322 7  | 0.045 8  | 0.024    |
| 贡献率/%   | 57.677 9 | 23.331 2 | 12.450 9 | 5.377 6  | 0.763 2  | 0.399 2  |
| 累计贡献率/% | 57.677 9 | 81.009 1 | 93.46    | 98.837 6 | 99.600 8 | 100      |

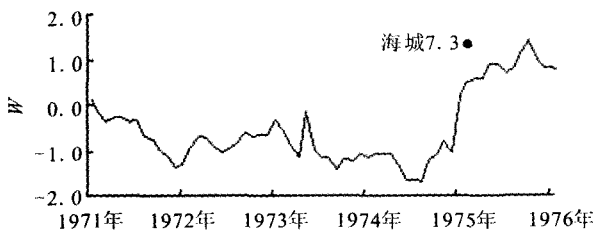


图 2 海城地震前后地震强度综合指标  $W$  值的变化

Fig. 2 Curve of earthquake intensity synthetic index  $W$  before and after Haicheng earthquake.

## 2.2 岫岩 5.4 级地震

本文选取 1970 年以来东北地区( $38^{\circ}\sim 55^{\circ}N$ ,  $117^{\circ}\sim 135^{\circ}E$ )小震目录,考虑到地震孕育过程中的可能影响范围及东北地区地震活动实际情况,对岫岩 5.4 级地震选取 200 km 左右圆域。除地震频次  $N(M_L \geq 3.0)$  外,其它参数的计算选取  $M_L 2.0$  地震,并剔除余震,资料计算时间窗为 12 月,滑动步长为 1 个月。

采用与海城 7.3 级地震震例相同的主成分分析方法,取累计贡献率  $E_m > 85\%$  的最小  $m(m < p)$  确定主成分个数  $m$ 。在此基础上对  $m$  个主成分使用

式(5)进行综合分析,得到反映地震活动时、空、强异常特征的综合指标  $W$ ,并计算其中强以上地震前后随时间变化曲线(图3)。由图3可知, $W$ 值大体在0以下波动,在中强震发生前1年左右的中期阶段出现明显的增高异常,异常高值通常超过1.0,这与文献[6]研究一致。

以上两个震例初步表明,反映地震时、空、强异

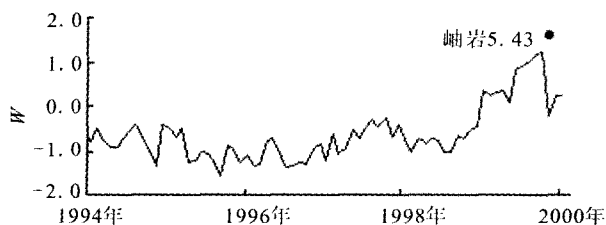


图3 岫岩地震前  $W$  值异常

Fig. 3 The  $W$  value anomaly before Xuyan earthquake.

常特征的综合指标  $W$  在震中附近地区震前的异常变化明显,可以较好地反映地震活动异常。

### 3 主成分对综合指标 $W$ 值的影响

图4为对与地震活动强度有关的6个参量进行主成分分析时得到的3个公共因子得分随时间变化曲线。可以看到图4中综合指标  $W$  值的基本形态主要由主成分1确定,这是由于主成分1的方差贡献率最大,达到58%。主成分1主要反映了异常相对较好的  $M_f$  值、 $A(b)$  值2个参数的信息。其它2个成分得分各有所异,且异常变化特征不甚显著,但因为它们的贡献率较小,因此综合指标  $W$  值可以较好地反映出震前的异常变化。由此可以看出,因子分析可以通过研究各参量的相关矩阵内部结构,得到各参量的因子载荷和因子贡献率,将一些包含异常信息较大的参量和因子赋予较大的权值,从而使综合指标  $W$  值的异常变化明显。

### 4 结论

目前在地震预测中由于地震活动性参数较多,指标之间有一定的相关性,因而所得的统计数据反映的信息在一定程度上有重叠;而且无论在正常情况还是异常情况下各个参数的形态变化各不相同<sup>[7-8]</sup>。这不但给分析预测工作带来不必要的麻烦,还可能存在主观片面性,不能很好地抓住事物的主要矛盾。而主成分分析有可能改善这一问题。主成分分析法可在力保原始数据信息丢失最少情况下,对高维变量空间进行降维处理,经过线性变换和舍弃部分信息,以少数的综合变量取代原有的多维变

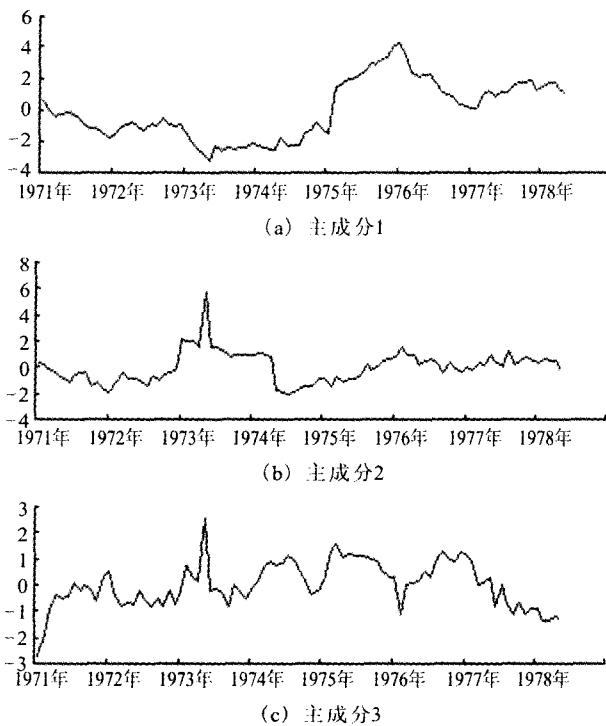


图4 主成分1、2、3分值曲线

Fig. 4 The value curves of the primary component 1, 2 and 3.

量。本文用主成分分析得到综合反映地震活动综合特征的  $W$  值,包含了以前描述地震活动强度各参数的绝大部分信息,在辽宁2次中强地震前出现明显的异常变化,表明综合指标  $W$  可以较好地反映地震活动异常特征,在地震预报中具有良好的应用前景。

### [参考文献]

- [1] 邵峰晶,于忠清.数据挖掘原理与算法[M].北京:中国水利水电出版社,2003:68-88.
- [2] 王伟.主成分分析及地震活动参数的约简[J].地震学报,2005,27(5):524-531.
- [3] 王伟.主成分分析法在地震预测中的应用研究[J].中国地震,2005,21(3):410-416.
- [4] 陆远忠,阎利军,郭若眉.用于中短期地震预报的一些地震活动性参量相关性讨论[J].地震,1999,19(1):11-18.
- [5] 韩渭滨.地震活动性参数分类及其相关性初步研究[J].四川地震,2003,(3):1-5.
- [6] 李永振.应用RTL算法对岫岩  $M_s 5.4$  地震的回顾性预测[J].东北地震研究,2008,24(1):39-43.
- [7] 王筱荣,李志海.新疆中强震预报的探讨[J].西北地震学报,2008,30(2):132-138.
- [8] 李亚荣,何新社,陈军.甘肃及邻近地区中强地震前弱震空区的时空参数研究[J].西北地震学报,2008,30(3):293-297.
- [9] 李永振,曹凤娟.主成分分析法在林甸  $M_s 5.1$  地震回顾性预测中的应用[J].东北地震研究,2008,24(2):29-34.