第 27 卷 第 5 期
2005 年 9 月 (524～531)

地 震 学 报
ACTA SEISMOLOGICA SINICA

Vol. 27, No. 5
Sep., 2005

**文章编号**: 0253-3782(2005)05-0524-08

# 主成分分析及地震活动参数的约简*

王 炜[1) ] 马钦忠[1)] 林命週[1)] 吴耿锋[2)] 吴绍春[2)]

1) 中国上海 200062 上海市地震局

2) 中国上海 200072 上海大学计算机工程与科学学院

**摘要** 选择从不同侧面反映地震活动时、空、强特征的地震频次 $N(M_L \geqslant 3.0)$、$b$ 值、$\eta$ 值、$A(b)$值、$M_f$ 值、$A_c$ 值、$C$ 值和 $D$ 值 8 个参量进行主成分分析. 通过主成分分析方法可以得到反映地震活动时、空、强异常特征的综合指标 $W$. 通常这 8 个参量之间具有一定的相关性，各参量在不同时段的变化各有所异，预报效果并不理想. 但是综合指标 $W$ 在华北 13 次 5.8 级以上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时、空、强异常特征. 本文还对主成分分析所得结论的有关问题进行了讨论.

**关键词** 主成分分析 数据挖掘 特征向量 贡献率

**中图分类号**: P315.5 **文献标识码**:

## 引言

数据挖掘(data mining)(邵峰晶，于忠清，2003)是近几年随着数据库和人工智能发展起来的一门新兴技术. 数据挖掘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，数据清洗是数据挖掘前的一个重要环节. 它包括去噪声，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. 本文将使用主成分分析方法对地震活动性参数进行约简.

在地震活动性分析中有许多预报指标，如 $b$ 值、$\eta$ 值、$C$ 值、$D$ 值、$M_f$ 值等. 这些指标都从不同侧面反映了地震活动时间、空间和强度特征. 目前这类参数较多，同时它们之间还可能存在着一定的相关性(陆远忠等，1999；韩渭宾，2003). 另外在实际预报中，常常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. 在正常情况下，也常常有些参数出现较明显的异常，而另一些参数并不出现异常. 这些都给实际预报带来困难.

如何解决预报参数过多而造成预报意见的不一致，本文将主成分分析方法(方开泰，1989；唐启义，冯光明，2002)用于多项预报指标的简化. 主成分分析方法是将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. 主成分分析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. 在原始数据的基础上应用主成分分析法，可以找出由若干个指标的线性组合而成的综合指标，即若干个主成分. 这些主成

* 地震科学联合基金(104090)资助项目.
2004-10-25 收到初稿，2005-04-14 收到修改稿，2005-05-16 决定采用.

分可以尽可能地反映原来的指标的信息，同时彼此之间相互独立.

本文使用 1975—2000 年大华北地区共发生的 13 次 5.8 级地震前后震中附近地区的地震资料，选择从不同侧面反映地震活动时、空、强特征的 8 个参量，即地震频次 $N(M_L \geqslant 3.0)$、$b$ 值、$\eta$ 值、$A(b)$值(吴佳翼，曹学峰，1983)、$M_f$ 值(王炜等，1997)、$A_c$ 值(吕悦军等，1997)、$C$ 值和 $D$ 值(王炜等，1997)进行主成分分析. 通过主成分分析方法可以得到反映地震活动时、空、强异常特征的综合指标 $W$. 在各次地震前，震中周围地区的这 8 个参量之间具有一定的相关性，各参量在不同时段的变化各有所异，效果并不一定都理想，但是综合指标 $W$ 在这 13 次 5.8 级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时、空、强异常的综合特征. 主成分分析方法是简化预报参量的有效工具.

## 1　主成分分析方法

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，即在保证原始数据信息损失最小的前提下，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 $x_1$，$x_2$，…，$x_p$，主成分分析后得到的主成分(综合变量)为 $Z_1$，$Z_2$，…，$Z_m$，它们是 $x_1$，$x_2$，…，$x_p$ 的线性组合($m < p$). 新变量 $Z_1$，$Z_2$，…，$Z_m$ 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 $Z_1$，$Z_2$，…，$Z_m$ 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 $Z_2$，$Z_3$，…，$Z_m$，依次有 $e_2 \geqslant$，…，$\geqslant e_m$. 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量最大的 $m$ 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：

$$x'_{ij} = \frac{x_{ij} - \bar{x}_i}{\sigma_i} \qquad i = 1, 2, \cdots, p;\ j = 1, 2, \cdots, n \tag{1}$$

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 $\boldsymbol{R} = (r_{ij})_{p\times n}$.

其中

$$r_{ij} = \frac{1}{n}\sum_{k=1}^{n}\frac{(x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j)}{\sigma_i \sigma_j} \tag{2}$$

3) 计算 $\boldsymbol{R}$ 的特征值和特征向量. 根据特征方程 $|\boldsymbol{R} - \lambda \boldsymbol{I}| = 0$，计算特征根 $\lambda_i$，并使其从大到小排列：$\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_p$，同时可得对应的特征向量 $u_1$，$u_2$，…，$u_p$，它们标准正交. $u_1$，$u_2$，…，$u_p$ 称为主轴. 这里，$\boldsymbol{I}$ 为单位矩阵.

4) 计算贡献率

$$e_i = \lambda_i / \sum_{k=1}^{p}\lambda_k \tag{3}$$

和累计贡献率

$$E_m = \sum_{k=1}^{m}\lambda_k / \sum_{k=1}^{p}\lambda_k \tag{4}$$

5) 计算主成分

$$Z_m = \sum_{j=1}^{p} u_{mj} x_j \tag{5}$$

这时，各主成分相互独立.

6）综合分析. 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. 一般取 $E_m > 85\%$ 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

$$W = \sum_{i=1}^{m} e_i Z_i \tag{6}$$

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 $N$（$M_L \geqslant 3.0$）、$b$ 值、$\eta$ 值、$A(b)$ 值、$M_f$ 值、$A_c$ 值、$C$ 值和 $D$ 值这 8 个参量进行主成分分析.

## 2 1979 年江苏溧阳 6.0 级地震前后地震活动性参数的主成分分析

为说明主成分分析方法如何用于实际地震预报中，下面以1979年7月9日江苏溧阳

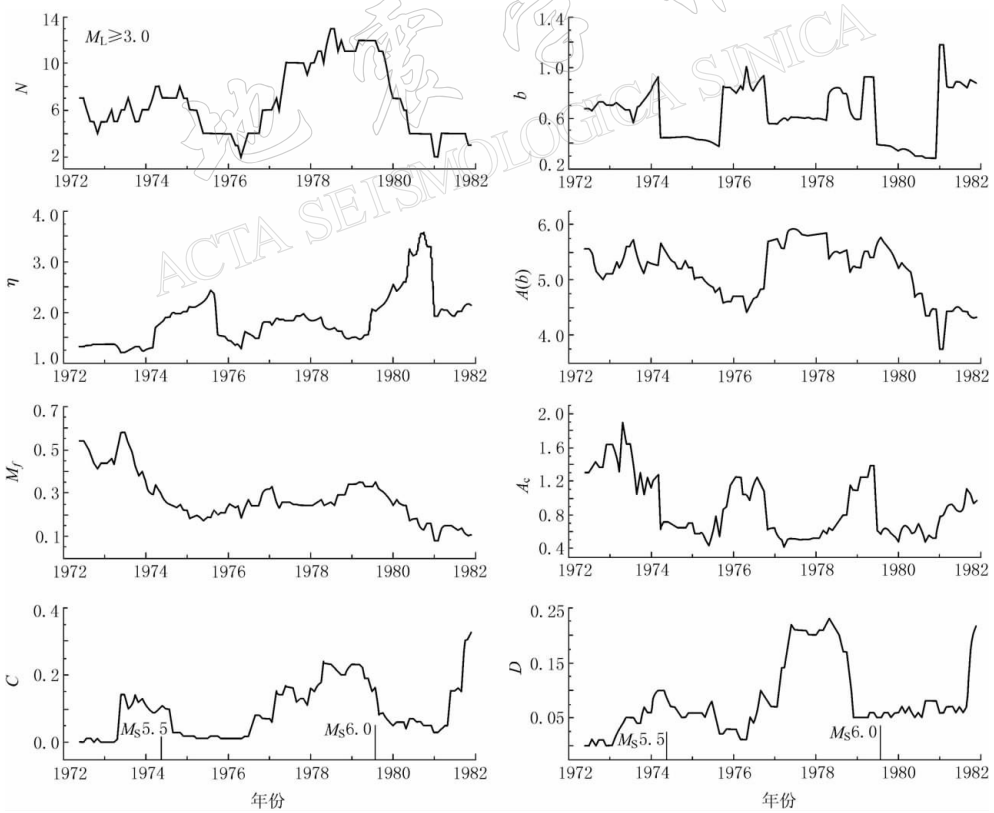


图 1 溧阳 6.0 级地震前后震中周围 200 km 范围内地震频次 $N$、$b$ 值、$\eta$ 值、$A(b)$ 值、$M_f$ 值、$A_c$ 值、$C$ 值和 $D$ 值随时间变化

6.0 级地震为例. 1979 年江苏溧阳发生 6.0 级地震，其前曾于 1974 年 4 月 22 日在同一地区发生 5.5 级地震. 由于该区大体在 1972 年开始建立较为密集的地震台网，所以本研究时间为 1972—1982 年. 图 1 为 6.0 级地震发生前后震中周围 200 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 $N$ ($M_L \geqslant 3.0$)外，其它参数计算所选取的起始震级为 $M_L 2.0$. 参数计算的累计时间为 18 个月，滑动步长为 1 个月. 从图 1 和一些文献(王炜等，1994；王炜等，1999)可以看到，6.0 级地震前这些参数各有所异，大多参数在这两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. 可以看到，当取 3 个主成分时，累计贡献率已达到 88 %；当取 4 个主成分时，累计贡献率已达到 93.6 %. 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 $M_f$ 值、$A_c$ 值、$\eta$ 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 $N$ ($M_L \geqslant 3.0$)、$A(b)$值系数较大，是构成主成分 2 的主要参数；主成分 3 中的 $b$ 值、$C$ 值、$D$ 值系数较大，是构成主成分 3 的主要参数.

表 1　各参量在各主成分中的系数(特征向量)、特征值与贡献率

| 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 | 主成分 7 | 主成分 8 |
|---|---|---|---|---|---|---|---|---|
| $N$ | 0.20458 | 0.55447 | 0.16753 | −0.28168 | 0.680877 | 0.06721 | 0.002849 | −0.27417 |
| $b$ | 0.03690 | −0.39502 | 0.57764 | −0.37439 | 0.088170 | 0.10789 | 0.549613 | 0.21739 |
| $\eta$ | −0.43546 | 0.10427 | −0.38672 | 0.29396 | 0.329825 | 0.39776 | 0.473405 | 0.26918 |
| $A(b)$ | 0.30802 | 0.55318 | 0.04883 | −0.08818 | −0.318129 | 0.06354 | 0.029231 | 0.69487 |
| $M_f$ | 0.51090 | 0.12412 | 0.01349 | 0.46319 | −0.227531 | 0.00006 | 0.548869 | −0.39471 |
| $A_c$ | 0.43237 | −0.32523 | 0.15936 | 0.39826 | 0.301545 | 0.48446 | −0.388171 | 0.21667 |
| $C$ | −0.30009 | 0.15672 | 0.52761 | 0.55811 | 0.179894 | −0.48204 | −0.043106 | 0.17022 |
| $D$ | −0.36653 | 0.27179 | 0.42704 | 0.04041 | −0.378921 | 0.59540 | −0.138417 | −0.30088 |
| 特征值 | 3.2205 | 2.0919 | 1.7629 | 0.4124 | 0.2949 | 0.1243 | 0.0588 | 0.0341 |
| 贡献率 | 40.256% | 26.148% | 22.036% | 5.1551% | 3.6868% | 1.5538% | 0.7357% | 0.4263% |
| 累计贡献率 | 40.257% | 66.405% | 88.442% | 93.597% | 97.284% | 98.837% | 99.573% | 100.00% |

本文取前 3 个主成分根据式(5)计算地震综合指标 $W$. 图 2 为溧阳 6.0 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 1974 年溧阳 5.5 级和 1979 年 6.0 级地震前 2～3 a 时间，溧阳地震震中附近地区 $W$ 出现明显增高，异常幅度一般大于 1.0(由于资料限制，1972 年前变化不清，但 1974 年溧阳 5.5 级地震前综合指标 $W$ 处于高值的过程是明显的)，震后异常恢复在 0 以下波动. 这表明地震活动时、空、强异常特征的综合指标 $W$ 可以很好地反映地震活动的增强和在时、空上的丛集异常综合特征.

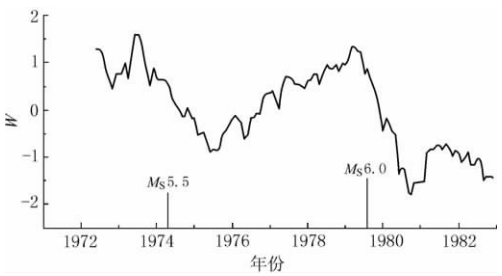


图 2　溧阳 6.0 级地震前后震中附近地区综合指标 $W$ 随时间的变化

比较图 1 中的 8 个反映地震活动时、空、强方面的参数随时间变化，由于这些参数分别反映了地震活动时、空、强不同侧面的特征，可以看到图 1 中的一些参数在两次溧阳地震前的异常变化并不明显，而一些参数在其它时段变化反而变化较大，故总体预报效果不

好. 而通过主成分分析得到，地震时、空、强异常特征的综合指标 $W$ 则可以较好地反映地震活动异常的综合特征，它在震中附近地区震前的异常变化明显.

## 3 大华北地区中强以上地震前地震活动性参数的主成分分析

1972 年以来大华北地区(30°～43°N，105°～125°E)开始有较为完整的地震目录. 为了保证在中强震前有足够时间的小震记录资料，本文取 1975—2000 年该地区除强余震以外共发生的 13 次 $M_S \geqslant 5.8$ 地震(表 2)，分别取震中周围一定范围内的地震资料计算上述 8 个参数；并在此基础上进行主成分分析，研究这 13 次中强以上地震前反映地震活动时、空、强异常特征的综合指标 $W$ 值的异常变化. 在研究区域的选取上，为消除人为因素的影响，文中一般选取以大震震中为圆心的圆域. 考虑到地震孕育过程中的可能影响范围，对 5～6 级左右地震一般选取 200～250 km 左右的圆域；对 7 级以上的地震选取半径为 300～350 km 左右的圆域. 根据华北地区的地震控制能力，除地震频次 $N(M_L \geqslant 3.0)$外，其它参数的计算一般选取 $M_L$2.0～2.3 地震，并剔除其余震. 资料计算的时间窗大体为 18 个月，步长为 1 个月.

采用与上述 1979 年溧阳 6.0 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 $m$. 根据本文不同震例的主成分分析，通常 $m=3$ 时就可达到要求. 在此基础上，可对 $m$ 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 $W$，并计算其在中强以上地震前后的随时间变化曲线. 图 3 为表 2 中各次地震前后的变化曲线. 由图 3 可知，在正常情况下综合指标 $W$ 大体在 0 以下. 但在上述中强以上地震发生前 2～3 a 的中期阶段，综合指标 $W$ 都出现明显的增高异常，异常高值通常超过 1.0，多数震例的异常在震后恢复. 这同样表明，地震时、空、强异常特征的综合指标 $W$，可以较好地反映地震活动异常的综合特征，它在震中附近地区震前的异常变化明显.

表 2 1972—2000 年华北地区 $M_S \geqslant 5.8$ 地震

| 编号 | 发震时间 | 震中位置 | | | $M_S$ |
|---|---|---|---|---|---|
| | | $\varphi_N$ | $\lambda_E$ | 地点 | |
| 1 | 1975-02-04 | 40°42′ | 122°42′ | 海城 | 7.3 |
| 2 | 1976-04-06 | 40°14′ | 112°12′ | 和林格尔 | 6.3 |
| 3 | 1976-07-28 | 39°38′ | 118°11′ | 唐山 | 7.8 |
| 4 | 1976-09-23 | 40°05′ | 106°21′ | 巴音木仁 | 6.3 |
| 5 | 1979-07-09 | 31°27′ | 119°15′ | 溧阳 | 6.0 |
| 6 | 1979-08-25 | 41°14′ | 108°07′ | 五原 | 6.0 |
| 7 | 1981-08-13 | 40°30′ | 113°25′ | 丰镇 | 5.8 |
| 8 | 1983-11-07 | 35°18′ | 115°36′ | 菏泽 | 5.9 |
| 9 | 1984-05-21 | 32°38′ | 121°36′ | 南黄海 | 6.2 |
| 10 | 1989-10-19 | 39°57′ | 113°49′ | 大同 | 6.1 |
| 11 | 1996-05-31 | 40°42′ | 109°36′ | 包头 | 6.2 |
| 12 | 1996-11-09 | 31°42′ | 123°06′ | 南黄海 | 6.1 |
| 13 | 1998-01-10 | 41°06′ | 114°18′ | 张北 | 6.2 |

## 4 讨论和结论

### 4.1 主成分与综合指标 $W$

图 4 为溧阳 6.0 级地震前后进行地震活动性参数的主成分分析时，根据式(4)得到的主成分 1、主成分 2、主成分 3 随时间变化曲线. 可以看到，图 2 中综合指标 $W$ 的基本形态主要由主成分 1 确定. 这是由于主成分 1 的贡献率最大，已达到 40%. 笔者也曾选择 3 个以上的主成分根据式(5)计算地震强度综合指标 $W$，所得结果与图 2 基本一致. 这是由于其它主成分贡献率较小的缘故.

另外，我们还可以看到，这 3 个主成分在两次溧阳地震前都出现了不同程度的上升变

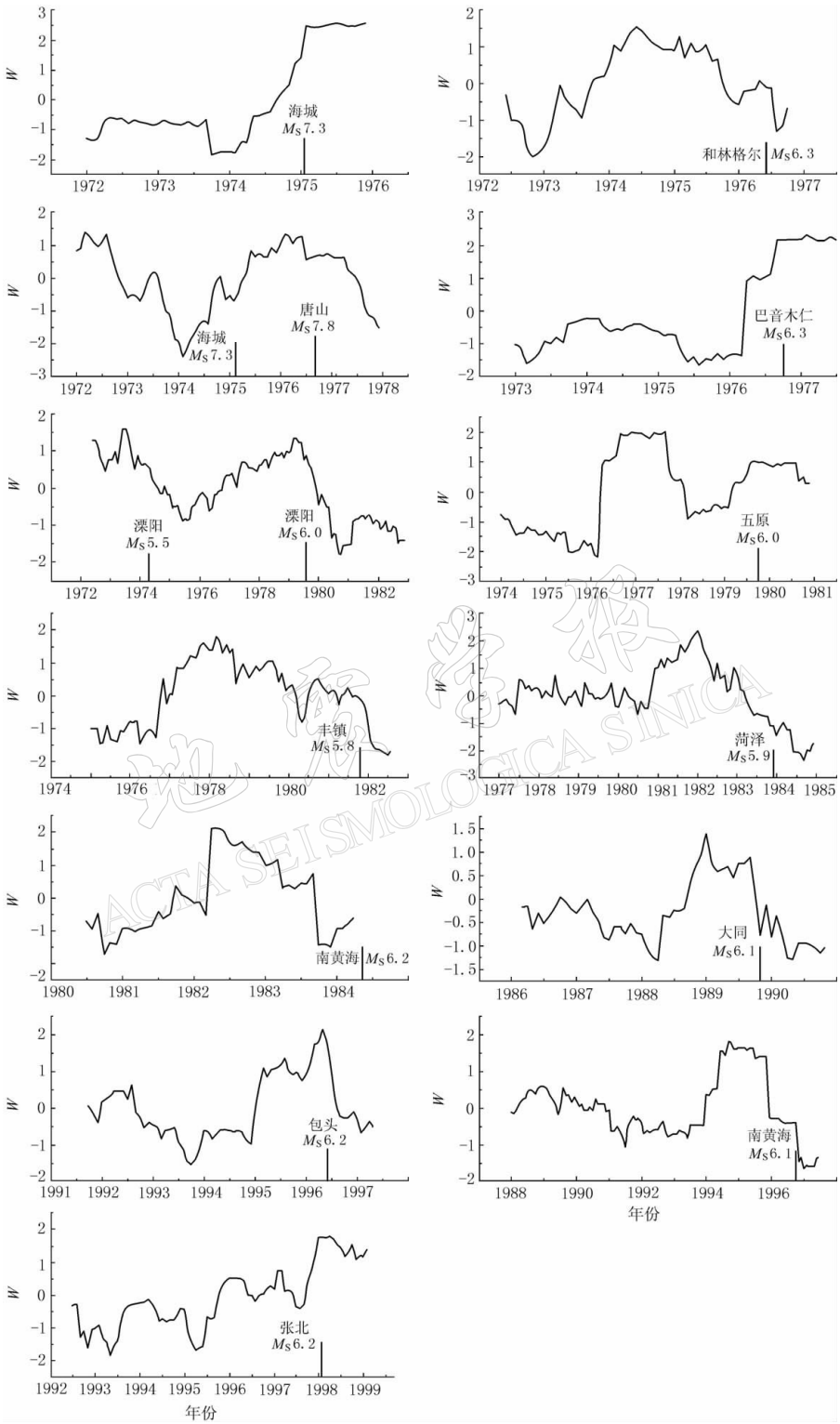


图 3　华北地区 13 次中强以上地震前反映地震活动时、空、强异常特征的综合指标 $W$ 值的异常变化

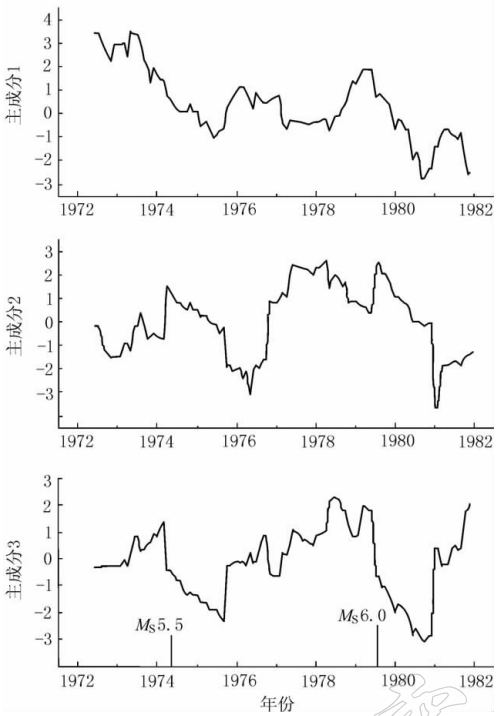


图 4 主成分分析得到的主成分 1，2，3 随时间的变化

化，但变化特征不甚显著，而综合指标 $W$ 可较好地反映出震前的异常变化. 这是由于主成分分析得到的前 3 个主成分是分别包含了上述 8 个参数中的部分信息，各个主成分对上述参数中的贡献率大小是不一样的. 而综合指标 $W$ 则包含了上述 8 个参数中的绝大部分信息，从而可以较好地反映出地震活动时、空、强异常变化的综合特征.

### 4.2 地震活动性参数间的相关性分析

不少研究者都注意到了一些地震活动性参数之间存在着一定的相关性(陆远忠等，1999；韩渭宾，2003). 主成分分析在对原始数据进行标准化处理后的第一步，就是计算各统计参数之间的相关系数. 表 3 为 1979 年江苏溧阳 6.0 级地震主成分分析所得到的上述 8 个参数之间的相关系数. 相关系数检验表明，地震频次 $N(M_L \geqslant 3.0)$ 与 $A(b)$ 值、$M_f$ 值存在着明显的相关性；$b$ 值与 $\eta$ 值、$A(b)$ 值、$A_c$ 值有一定的相关性；$\eta$ 值与 $A(b)$ 值、$M_f$、$A_c$ 值之间有一定的相关性；$A(b)$ 值与 $M_f$ 之间也有一定的相关性.

表 3 各参数之间的相关系数

| 参数 | 相关系数 | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | $N$ | $b$ | $\eta$ | $A(b)$ | $M_f$ | $A_c$ | $C$ | $D$ |
| $N$ | 1.000000 | −0.203113 | −0.247168 | 0.799411 | 0.388857 | −0.029053 | 0.105620 | 0.126864 |
| $b$ | −0.203113 | 1.000000 | −0.545903 | −0.358497 | −0.090714 | 0.424330 | 0.284029 | 0.151893 |
| $\eta$ | −0.247168 | −0.545903 | 1.000000 | −0.375847 | −0.652845 | −0.693142 | 0.156853 | 0.272972 |
| $A(b)$ | 0.799411 | −0.358497 | −0.375847 | 1.000000 | 0.647713 | 0.031787 | −0.107845 | 0.019104 |
| $M_f$ | 0.388857 | −0.090714 | −0.652845 | 0.647713 | 1.000000 | 0.671162 | −0.349537 | −0.48963 |
| $A_c$ | −0.029053 | 0.424330 | −0.693142 | 0.031787 | 0.671162 | 1.000000 | −0.295273 | −0.56560 |
| $C$ | 0.105620 | 0.284029 | 0.156853 | −0.107845 | −0.349537 | −0.295273 | 1.000000 | 0.792585 |
| $D$ | 0.126864 | 0.151893 | 0.272972 | 0.019104 | −0.489631 | −0.565600 | 0.792585 | 1.000000 |

本文的结果表明，目前地震预报中的地震活动性参数较多，这些参数虽然从不同侧面反映了地震活动在时间、空间和强度方面的特征，但是这些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下还是异常情况下各个参数的变化也各不相同. 这就给分析预报工作带来不必要的麻烦. 在实际预报中我们深有所感. 而主成分分析法可在力保原始数据信息丢失最小情况下，对高维变量空间进行降维处理，即在保证原始数据信息损失最小前提下，经过线性变换和舍弃部分信息，以少数的主成分取代原有的多维变量. 尤其本文将多个主成分根据其贡献率的大小综合成反映地震活动时、空、强异常特征的综合指标 $W$. 该指标包含了上述 8 个参数中的绝大部分

信息，从而可以较好地反映出地震活动时、空、强异常变化的综合特征.

总之，主成分分析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的综合与约简，并能够客观地确定各个指标的权数，避免了主观随意性. 通过主成分分析方法得到的反映地震活动时、空、强异常特征的综合指标 $W$ 在华北地区 13 次中强以上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时、空、强异常特征. 主成分分析方法是约简预报参量的有效工具，在地震预报中具有良好的应用前景.

## 参　考　文　献


方开泰. 1989. 实用多元统计分析[M]. 上海：华东师范大学出版社，286～295

韩渭宾. 2003. 地震活动性参数分类及其相关性初步研究[J]. 四川地震，(3)：1～5

陆远忠，阎利军，郭若眉. 1999. 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，**19**(1)：11～18

吕悦军，陆远忠，郑悦君. 1997. 用算法复杂性分析地震活动演化特征[J]. 地震，**17**(1)：25～33

邵峰晶，于忠清. 2003. 数据挖掘原理与算法[M]. 北京：中国水利水电出版社，68～88

唐启义，冯光明. 2002. 实用统计分析及其 DPS 数据处理系统[M]. 北京：科学出版社，367～373

王炜，庄昆元，宋俊高，等. 1997. 测震学分析预报方法[M]. 北京：地震出版社，86～98

王炜，戴维乐，黄冰树，等. 1994. 地震震级的统计分布及其地震强度因子 $M_f$ 值在华北中强以上地震前的异常变化[J]. 中国地震，**10**(增刊)：95～100

王炜，刘峥，宋先月，等. 1999. 地震活动性的定量化及其在地震中期预报中的应用[J]. 中国地震，**15**(2)：116～127

吴佳翼，曹学锋. 1983. 地震活动性的定量化问题[J]. 地震，**3**(6)：13～17


# PRIMARY COMPONENT ANALYSIS METHOD AND REDUCTION OF SEISMICITY PARAMETERS


Wang Wei[1)]　Ma Qinzhong[1)]　Lin Mingzhou[1)]　Wu Gengfeng[2)]　Wu Shaochun[2)]

1) *Earthquake Administration of Shanghai Municipality, Shanghai* 200062, *China*

2) *School of Computer Engineering and Science, Shanghai University, Shanghai* 200072, *China*



**Abstract**: In the paper, the primary component analysis is made using 8 seismicity parameters of earthquake frequency $N(M_L \geqslant 3.0)$, $b$-value, $\eta$-value, $A(b)$-value, $M_f$-value, $A_c$-value, $C$-value and $D$-value that reflect the characteristics of magnitude, time and space distribution of seismicity from different respects. Through the primary component analysis method, the synthesis parameter $W$ reflecting the anomalous features of earthquake magnitude, time and space distribution can be gained. Generally, there is some relativity among the 8 parameters, but their variations are different in different periods. The anomalous variation of some parameters before strong earthquakes is not very well. However, the synthesis parameter $W$ showed obvious anomalies before 13 earthquakes ($M_S \geqslant 5.8$) occurred in the North China area, which indicates that the synthesis parameter $W$ can better reflect the anomalous characteristics of magnitude, time and space distribution of seismicity. Other problems relating to the conclusions drawn by the primary component analysis method are also discussed.



**Key words**: primary component analysis method; data mining; eigenvector; contribution rate