

利用主成分分析法研究东北地震活动异常*

李永振

(辽宁省地震局, 沈阳 110034)

摘要: 选择与地震强度有关的地震频次 N ($M_L \geq 3.0$)、 b 值、 η 值、 M_f 值、 $A(b)$ 值和 A_c 值共 6 个参量进行主成分分析, 实现对上述参量的有效约简。这 6 个参量之间有一定的相关性, 在不同时段各参量的变化各有所异。根据主成分分析可以得到反应地震强度特征的综合指标 W 。1970 年以来东北 14 次 5 级以上浅源地震前 W 都出现明显的异常变化, 震后恢复。这表明综合指标 W 可以较好地反映地震活动的异常特征。

关键词: 主成分分析; 相关性; 特征向量; 贡献率

中图分类号: P315.7

文献标识码: A

文章编号: 1000-0666(2010)02-0164-06

0 引言

目前, 在地震活动分析中存在许多预报指标, 如 b 值、 C 值、 D 值、 Y 值、 η 值等, 这些指标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一定的相关性。在一些中强以上地震前有些参数会出现较明显的异常, 而另一些参数并不出现异常, 这些都给实际预报带来困难。王伟等 (2005)、李滔 (2007) 用主成分分析法对地震活动性参数进行了约简, 同时将该方法应用到江苏溧阳 6.0 级地震中。本文利用主成分分析法, 对 1970 年以来东北 14 次 5 级以上浅源地震进行了研究, 实现了多项预报指标的简化。

主成分分析方法是将被描述样本特征的多个可能有一定相关性的指标 (陆远忠等, 1999; 韩渭宾, 2003) 化为少数几个综合指标的一种统计分析方法。主成分分析法在最大程度保留原始数据信息的基础上, 还能对高维变量空间进行综合与简化, 并能够客观地确定各个指标的权数, 避免了主观随意性 (方开泰, 1989; 唐启义, 冯光明, 2002; 吕悦军等, 1997)。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线性组合而得到的综合指标, 即若干个主成分。这些主成分可以最大化地反映原来指标的信息, 同时彼

此之间相互独立。

本文选用 1970 年以来, 东北地区 14 次 5 级以上浅源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 级以上地震频次 N 、 b 值、 η 值、 M_f 值、 $A(b)$ 值和 A_c 值进行主成分分析。这 6 个参量之间具有一定的相关性, 在不同时段各参量的变化各有所异, 预报效果并不理想。但是根据主成分分析方法可以得到反应地震强度异常特征的综合指标 W , 发现在地震前 W 均出现明显的异常变化, 多数震例震后异常变化恢复, 表明综合指标 W 可以较好地反映地震活动的异常特征。主成分分析方法是简化预报参量的有效工具, 在地震预报中具有一定的应用前景。

1 主成分分析法

主成分分析法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, 经过线性组合, 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。综合变量 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, 故称 z_1, z_2, \dots, 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, \dots, z_m ，依次有 $e_2 \geq e_3 \geq \dots \geq e_m$ 。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：

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

式中， x_{ij} 为第 j 个指标第 i 个样本的原始数据， \bar{x}_i 和 σ_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：

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

(3) 计算 R 的特征值和特征向量。根据特征方程 $|R - \lambda I| = 0$ ，计算特征根 λ_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 和累计贡献率

$$e_i = \lambda_i / \sum \lambda_k, \tag{3}$$

$$E_m = \sum \lambda_k / \sum \lambda_k. \tag{4}$$

(5) 计算主成分

$$z_m = \sum u_{mj} x_j. \tag{5}$$

(6) 综合分析。一个 m 维主超平面究竟以多大的精度来代替原始变量系统，才能尽可能多的保留原始数据信息？这可以通过求累计贡献率 E_m 来判断。一般取 $E_m > 85\%$ 的最小变量 m ($m < p$)，就可确定主超平面的维数 m ，从而可对 m 个主成分进行综合分析。

(7) 根据主成分 z_i 分析得到的主成分和相应的权值（贡献率） e_i ，计算本文定义的反映地震活

动时间、空间和强度异常特征的综合指标：

$$W = \sum e_i z_i. \tag{6}$$

因为 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间、空间和强度异常的基本特征。本文选择了从不同侧面反映地震活动时间、空间和强度特征的 6 个参量：地震频次 N ($M_L \geq 3.0$)、 b 值、 η 值、 M_f 值、 $A(b)$ 值、 A_c 值进行主成分分析。

2 1975 年海城 7.3 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 1975 年 2 月 4 日海城 $M_s 7.3$ 地震为例。图 1 为 7.3 级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的变化曲线。除地震频次 N 选取的起始震级为 $M_L 3.0$ 外，其它参数计算所选取的起始震级均为 $M_L 2.0$ 。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，虽然地震前这些参数变化各有所异，但是大多参数变化并不显著。

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，累计贡献率达到 93%；当取 4 个主成分时，累计贡献率达到 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、 η 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 (6) 计算地震综合指标 W 。图 2 为海城 7.3 级地震前后震中附近地区地震活动的综合指标 W 随时间的变化。可以看到，在 1975 年海城 7.3 级地震前 6 个月时间，震中附近地区 W 值出现明显抬升的形态，异常幅度大于 1.0。这表明地震活动的综合指标 W 可以很好地反映地震活动增强特征（王炜等，1994，1999）。

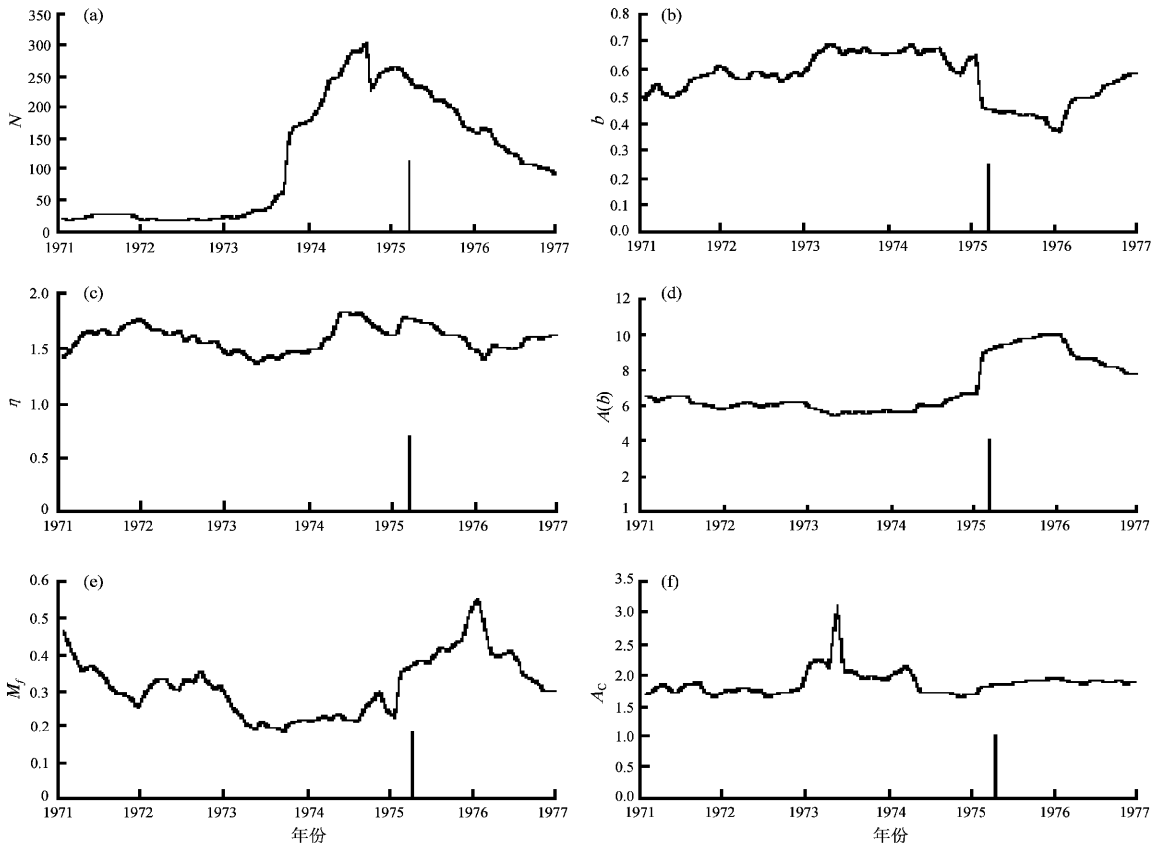


图 1 海城地震参数随时间变化曲线

Fig. 1 Curve of the seismic activity parameter with time

表 1 各参量在各主成分中的系数（特征向量）、特征与贡献率

Tab. 1 Parameters in primary component of the coefficient (eigenvector)、features and contribution rate

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
η	-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

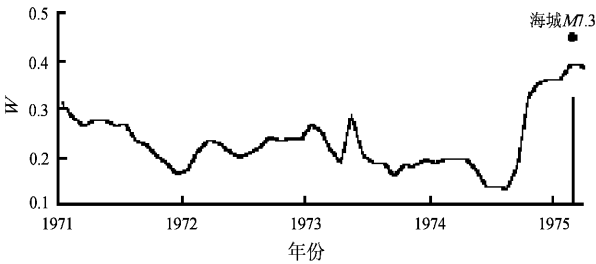


图 2 地震强度综合指标 W 值随时间的变化

Fig. 2 Curve of earthquake intensity synthetic index W with time

3 东北地区中强震前地震活动性参数主成分分析

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, 1970 年 ~ 2008 年东北地区共发生 $M_s \geq 5.0$ 浅源地震 14 次 (剔除了海城、岫岩地震强余震, 不含唐山老震区地震), 分别取上述 6 个参数, 进行主成分分析, 进而研究这些中强震前 W 值的异常变化。考虑到地震孕育过程中可

能的影响范围及东北地震活动实际情况，对 5 ~ 6 级地震一般选取 200 ~ 250 km 左右的圆域；对 7 级以上地震选取 300 km 左右的圆域。除地震频次 N 选取的起始震级为 $M_L3.0$ 外，其它参数的计算选取起始震级 $M_L2.0$ ，并剔除余震，资料计算时间窗为 12 月，滑动步长为 1 个月。

采用主成分分析方法，取累计贡献率 > 85% 的最小的 m ($m < p$) 确定主成分个数 m 。在此基础上，对 m 个主成分使用 (b) 式进行综合分析，得到反映地震活动时间、空间和强度异常特征的综合

指标 W ，并计算其中强以上地震前后随时间的变化曲线。图 3 为表 2 中各次地震前后的变化曲线，由图 3 可知，综合指标表现为 W 值大体在 0 以下波动，在中强震发生前 0.5 ~ 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。有的震例是在高值发震，如海城、岫岩地震；有的震例是高值恢复后发震，如前郭、鄂伦春地震。这表明，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 1970 年以来东北 $M_s \geq 5.0$ 地震

Tab. 2 $M_s \geq 5.0$ earthquakes in the Northeast China since 1970

序号	地震时间	震中位置		地点	震级
	年 - 月 - 日	北纬	东经		
1	1975 - 02 - 04	40°44′	122°43′	辽宁海城	$M_s7.3$
2	1978 - 05 - 18	40°43′	122°37′	辽海海城	$M_s5.9$
3	1980 - 02 - 10	48°53′	122°13′	内蒙古博克图	$M_s5.3$
4	1981 - 04 - 26	48°53′	122°01′	内蒙古博克图	$M_s5.3$
5	1986 - 02 - 01	48°38′	126°40′	黑龙江德都	$M_s5.0$
6	1986 - 03 - 01	48°38′	126°40′	黑龙江德都	$M_s5.4$
7	1986 - 08 - 16	48°35′	126°31′	黑龙江德都	$M_s5.5$
8	1999 - 11 - 29	40°24′	123°11′	辽宁岫岩	$M_s5.4$
9	2000 - 01 - 12	40°30′	123°05′	辽宁岫岩	$M_s5.1$
10	2003 - 08 - 16	43°50′	120°09′	内蒙古巴林左旗	$M_s5.9$
11	2004 - 03 - 24	45°24′	118°12′	内蒙古东乌珠	$M_s5.9$
12	2005 - 07 - 25	46°54′	125°00′	黑龙江林甸	$M_s5.1$
13	2006 - 03 - 31	44°42′	124°00′	吉林前郭	$M_s5.0$
14	2008 - 06 - 10	49°00′	122°30′	黑龙江鄂伦春	$M_s5.2$

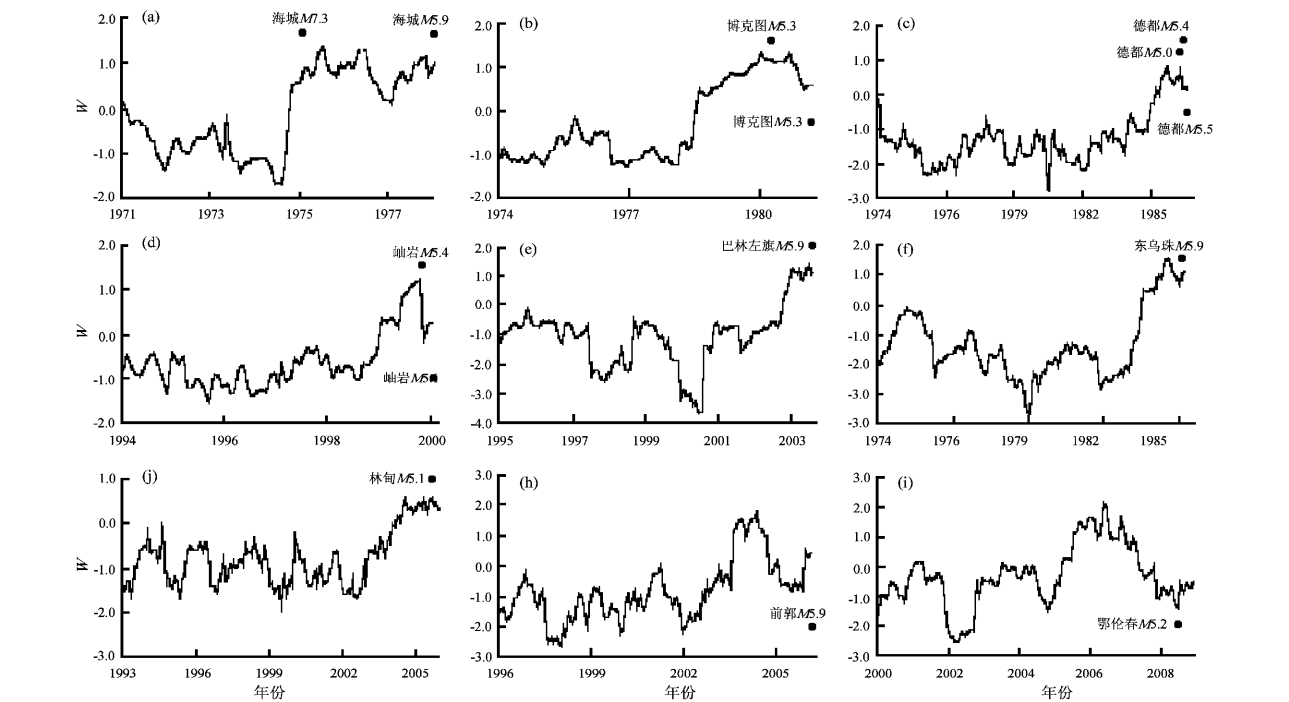


图 3 东北 14 次 $M_s \geq 5.0$ 浅源地震前 W 值异常

Fig. 3 W value anomaly before 14 shallow earthquakes ($M_s \geq 5.0$) in the Northeast China

4 主因子得分与综合指标 W 值

图4为对与地震活动强度有关的6个参量进行因子分析时,得到的4个公共因子得分随时间变化曲线,可以看到图2中综合指标 W 值的基本形态主要由主因子1确定,这是由于主因子1的贡献率最大,达到58%。而因子分析的结果显示,因子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个因子得分各有所异,且异常变化特征不甚显著,但由于它们的贡献率较小,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,在因子分析中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,从而使得综合指标 W 值的异常变化明显。

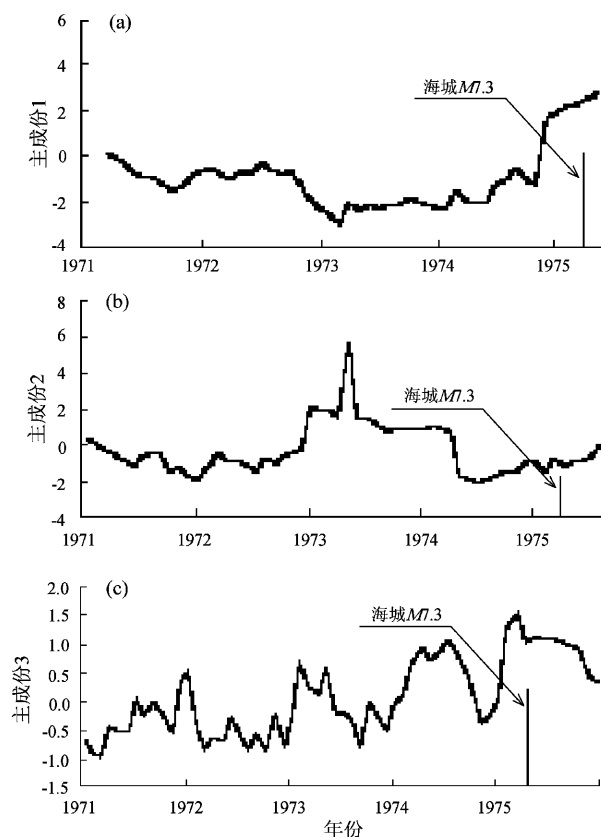


图4 主成分1 (a)、2 (b)、3 (c)
随时间的变化

Fig. 4 Curve of the primary component
1, 2 and 3 with time

5 结论

目前在地震预测中,反映地震活动性的参数较多,这些参数从不同侧面反映了地震在时间、空间和强度方面的特征(李献智,侯建盛,2000)。由于各个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重叠,而且各个参数的形态变化各不相同,这不但给分析预测工作带来不必要的麻烦,还可能使预测工作存在主观随意性。主成分分析法可在保证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,经过线性组合,以少数的综合变量取代原有的多维变量。本文所使用综合反映地震活动时间、空间和强度特征的 W 值,包含了上述6个参数中的绝大部分信息,可以较好地反映地震活动的时间、空间和强度异常变化特征。通过主成分分析方法得到的反映地震异常特征的综合指标 W 在东北14次5级以上中强浅源地震前出现一定的异常变化,表明综合指标 W 可以较好地反映地震活动异常特征,在地震预报中具有一定的应用前景。

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.
- 李滔. 2007. 滇西地区前兆震群参数的主成分分析[J]. 地震研究, 30(1): 24-28.
- 李献智, 侯建盛. 2000. 20世纪中国大陆6级以上地震活动图像及其意义[J]. 地震研究, 23(3): 263-268.
- 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地震[J]. 地震, 7(1): 25-33.
- 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, 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). 95-110.
- 王炜, 林命遇, 马钦忠, 等. 2005. 主成分分析法在地震预测中的应用研究[J]. 中国地震, 21(3): 410-416.

- 王炜, 刘峥, 宋先月, 等. 1999. 地震活动性的量化及其在地震中期预报中的应用[J]. 中国地震, 15(2): 116 – 127.
- 王炜, 马钦忠, 林命邇, 等. 2005. 主成分分析及地震活动参数的约简[J]. 地震学报, 27(5): 524 – 531.

Using the Primary Component Analysis Earthquake Parameters Reduction

LI Yong-zhen

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

Abstract

Abstract: the primary component analysis is made using 6 seismicity parameters of earthquake frequency N ($M_L \geq 3.0$), b , η , $A_{(b)}$, M_f , A_c , which had relation to the earthquake strenth, it was achieved to reduce the above-metioned parameters effectively. There were some relativities among the 6 parameters and each one was different in distinct stages, but the earthquake intensity synthetic index W can be gained by the primary component analysis. W value showed obvious anomalies before 14 shadow earthquakes occurred in the Northeast China area and returned to normal after that. It showed that the synthetic index W can reflect the anomalous characteristics of the seismicity.

Key words: primary component analysis method, relativity, eigenvector, contribution rate