

地震活动参数约简的因子分析方法

王 炜, 林命週, 赵利飞, 马钦忠

(上海市地震局, 上海 200062)

摘要:目前地震活动性分析有许多指标参量,各参量在不同时段变化各异,预报效果不理想。由于各参量之间通常具有一定的相关性,本文选择地震频次 $N(M_L \geq 3.0)$ 、 b 值、 η 值、 $A(b)$ 值、 M_f 值、 A_c 值、 C 值和 D 值 8 个参量进行因子分析,得到的反映地震活动时、空、强异常特征的综合指标 W_{fa} 。结果表明 W_{fa} 在 2005 年江西九江 5.7 级地震和华北 14 次 5.7 级以上地震前出现明显的异常变化,可以较好地反映地震活动时、空、强异常特征。还对因子分析的有关问题进行了讨论。

关键词: 地震活动参数; 因子分析; 特征向量; 贡献率; 九江地震

中图分类号: P315.75 文献标识码: A 文章编号: 1000-0844(2006)04-0303-06

Factor Analysis Method for Reducing Seismicity Parameters

WANG Wei, LIN Ming-zhou, ZHAO Li-fei, MA Qin-zhong

(Earthquake Administration of Shanghai, Shanghai 200062, China)

Abstract: There are many parameters in analysis of seismicity. The variations of those parameters are different in different period of seismic process, and the earthquake forecast results of the parameters are not well. For existing some correlations among those parameters, 8 parameter, such as $N(M_L \geq 3.0)$, b -value, η -value, $A(b)$ -value, M_f -value, A_c -value, C -value and D -value are selected for factor analysis method, and an synthetic parameter W_{fa} , that reflects the temporal, spatial and intensity characteristics of seismicity, is made. The result shows that W_{fa} had obvious anomalous change before Jiujiang $M_s 5.7$ earthquake in 2005, Jiangxi Province, and 14 $M_s \geq 5.7$ earthquake events in North China, can reflect the anomalous variation in temporal, spatial and intensity characteristics of seismicity well. Some problems in the factor analysis method are discussed also.

Key words: Seismicity parameter; Factor analysis method; Eigenvector; Contribution rate; Jiujiang earthquake

0 引言

在地震活动性分析中有许多预报指标,如 b 值、 η 值、 C 值、 D 值、 M_f 值、 A_c 值等,从不同侧面反应了地震活动时间、空间和强度特征。目前这类参数较多,同时它们之间可能存在着一定的相关性^[1-2]。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;或在正常情况下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。

为解决预报参数过多而造成预报意见的不一致,本文将因子分析(Factor Analysis)方法^[3-4]用于多项预报指标的约简。因子分析法是主成分分析方法^[5-6]的推广和深化,是把一些可能有一定相关性的变量归结为少数几个综合因子的一种多元统计分析方法。因子分析从一组观测数据出发,通过研究其相关矩阵(或协方差矩阵)的内部结构,找出对变量起支配作用的几个互不相关的因子,在尽量减

少原始数据信息损失的前提下,用少数几个公共因子去代替数量较多且有一定联系的原始变量,从而达到揭示原始变量之间的内在联系、合理解释原始变量与主要因子关系的目的。同时,因子分析方法能够在最大限度地保留原有信息的基础上,客观地确定各个因子的权数,避免了主观随意性,从而可由若干主要因子得分的线性组合构成综合指标 W_{16} ,对地震活动性参数进行有效的约简。

本文首先以2005年11月26日江西九江5.7级地震为例,检验了综合指标 W_{16} 的预报效果;进而使用1975—2005年期间大华北地区共发生的14次5.7级地震前后震中附近地区的地震资料,选择地震频次 $N(M_i \geq 3.0)$ 、 b 值、 η 值、 $A(b)$ 值、 M_f 值、 Ac 值、 C 值和 D 值8个参量进行因子分析,得到综合指标 W_{16} 随时间的变化。结果显示 W_{16} 在这14次5.7级地震前出现明显的异常变化,表明可以较好地反映地震活动时、空、强异常的综合特征。因子分析方法是简化预报参量的有效工具,在地震预报中具有良好的应用前景。

1 因子分析方法

1.1 因子分析的数学模型

因子分析法^[3-4]是在为数众多的因素中,在不损失或很少损失原有信息的基础上,通过对样本相关矩阵的内部依赖关系的研究,将多个指标归结为少数几个公共因子以简化观测系统,并最终对各样本点进行定量评价的多元统计分析方法。其关键在于正交因子模型的建立和分析应用。

设 $x_i (i=1, 2, \dots, p)$ 为地震预报指标变量,其中某些指标变量之间具有不同程度的相关性。假定每个变量已经过标准化。本文所说的地震预报指标的正交因子模型指的是:存在 $m (m \leq p)$ 个公共因子 $F_i (i=1, 2, \dots, m)$, 使指标变量 x_i 可用它们的线性组合表示为

$$X = AF + \epsilon \quad (1)$$

式中, $X = (x_1, x_2, \dots, x_p)$; F 为公共因子, $F = (F_1, F_2, \dots, F_m)$; ϵ 为误差或特殊因子, $\epsilon = (\epsilon_1, \epsilon_2, \dots, \epsilon_p)$; A 为因子载荷矩阵, $A = (a_{ij})_{p \times m}$ 。 a_{ij} 称为第 i 个指标在第 j 个公共因子上的载荷,这是第 i 个变量与第 j 个公共因子的相关系数;特殊因子 ϵ_i 相互独立,且服从正态分布 $N(0, \sigma^2)$ 。

1.2 因子分析方法

设原始数据矩阵为

$$X = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \dots & \dots & \dots & \dots \\ x_{p1} & x_{p2} & \dots & x_{pn} \end{pmatrix} \quad (2)$$

这里 n 为样本数, p 为变量数。本文使用因子分析方法计算综合指标 W_{16} 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:

$$x'_{ij} = (x_{ij} - \bar{x}_i) / \sigma_i \quad (3)$$

($i = 1, 2, \dots, p; j = 1, 2, \dots, n$)

式中 x_{ij} 为第 i 个指标第 j 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 $R = (r_{ij})_{p \times p}$, 其中

$$r_{ij} = \frac{1}{n} \sum_{k=1}^n (x'_{ik} - \bar{x}_i)(x'_{jk} - \bar{x}_j) / \sigma_i \sigma_j \quad (4)$$

(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。这里 I 为单位矩阵。

(4) 计算公共因子的贡献率 $e_i = \lambda_i / \sum_{k=1}^p \lambda_k$ 和累计贡献率 $E_m = \sum_{k=1}^m \lambda_k / \sum_{k=1}^p \lambda_k$ 。公共因子 F_i 的方差贡献率反映了该因子包含信息量的多少, 贡献率越大该因子越相对重要。究竟选取多少公共因子来近似代替原始变量系统, 通常根据领域的特点来确定。本文选取累计方差贡献率大于85%的前 m 个公共因子。

(5) 建立初始因子载荷矩阵 A 。以上过程可以确定前 m 个公共因子, 这与主成分分析方法相同。在此基础上由式(1)可以得到初始因子载荷矩阵 A 。

(6) 因子载荷矩阵 A 的变换。建立因子分析数学模型的目的不仅是为了找出公共因子, 更重要的是要知道每个公共因子的意义, 以便对实际问题进行分析。但在初始因子载荷矩阵中, 往往一些变量在几个公共因子上都有较大的载荷, 使得初始公共因子的实际意义难以得到合理解释。这时可对因子载荷矩阵实行旋转, 即用一个正交阵右乘使旋转后的因子载荷矩阵结构简化, 便于对公共因子进行解释。所谓结构简化就是使每个变量仅在一个公共因子上有较大的载荷, 而在其余公共因子上的载荷比较小。这种变换因子载荷的方法称为因子旋转。因子旋转有方差最大正交旋转和斜交旋转, 本文使用方差最大正交旋转。

(7) 计算综合指标 W_{fa} 值。根据各公共因子得分 F_i 和相应的权值(贡献率) e_i , 本文定义反映地震活动时、空、强特征的综合指标为

$$W_{fa} = \sum_{i=1}^m e_i F_i \quad (5)$$

本文选择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、 η 值、 $A(b)$ 值、 M_f 值、 A_c 值、 C 值、 D 值共 8 个参量进行因子分析。由于 m 个公共因子已基本保留了这些预报参数的信息, 所以综合指标 W_{fa} 值包含了这些参数从不同侧面反映地震活动时、空、强异常的主要特征。

2 江西九江 5.7 级地震前地震活动性参数的因子分析

以下以江西九江 5.7 级地震为例说明因子分析方法如何用于实际地震预报中。2005 年 11 月 26 日江西九江发生 5.7 级地震, 该区历史上地震活动水平较低, 震前震中附近地震不多, 多次地震会商分析也没发现该区存在明显的地震活动性异常。图 1 为九江 5.7 级地震发生前震中周围 200 km 范围内

的地震频次 $N(M_L \geq 3.0)$ 、 b 值、 η 值、 $A(b)$ 值、 M_f 值、 A_c 值这 6 个地震活动性参数随时间变化曲线, 各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窗为 18 个月, 滑动步长为 1 个月。由图 1 可以看到, 九江 5.7 级地震前这些参数变化各有所异, 大多参数在这次地震前的异常变化并不显著, 仅 M_f 值自 2002 年 10 月开始出现明显的增高异常, 地震频次 $N(M_L \geq 3.0)$ 和 $A(b)$ 值也出现一定的增高异常, 但在 20 世纪 90 年代也曾出现过一定的异常, 而其后该区并无地震发生。

选择这 6 个参量进行因子分析。表 1 为通过因子分析方法得到的各因子的特征值与贡献率, 可以看到当取前 4 个因子时, 累计贡献率已达到 94.7%, 因此本文选取公共因子数 $m=4$ 。表 2 为各参量的初始因子载荷矩阵。

表 3 为对初始因子载荷矩阵进行方差最大正交旋转后得到的因子载荷, 使得每个变量仅在一个公共因子上有较大的载荷。由表 3 可知: 因子 1(F_1) 中的地震频次 $N(M_L \geq 3.0)$ 、 $A(b)$ 值、 M_f 值的因子载荷较大, 反映了地震活动增强的特征, 因此可称为

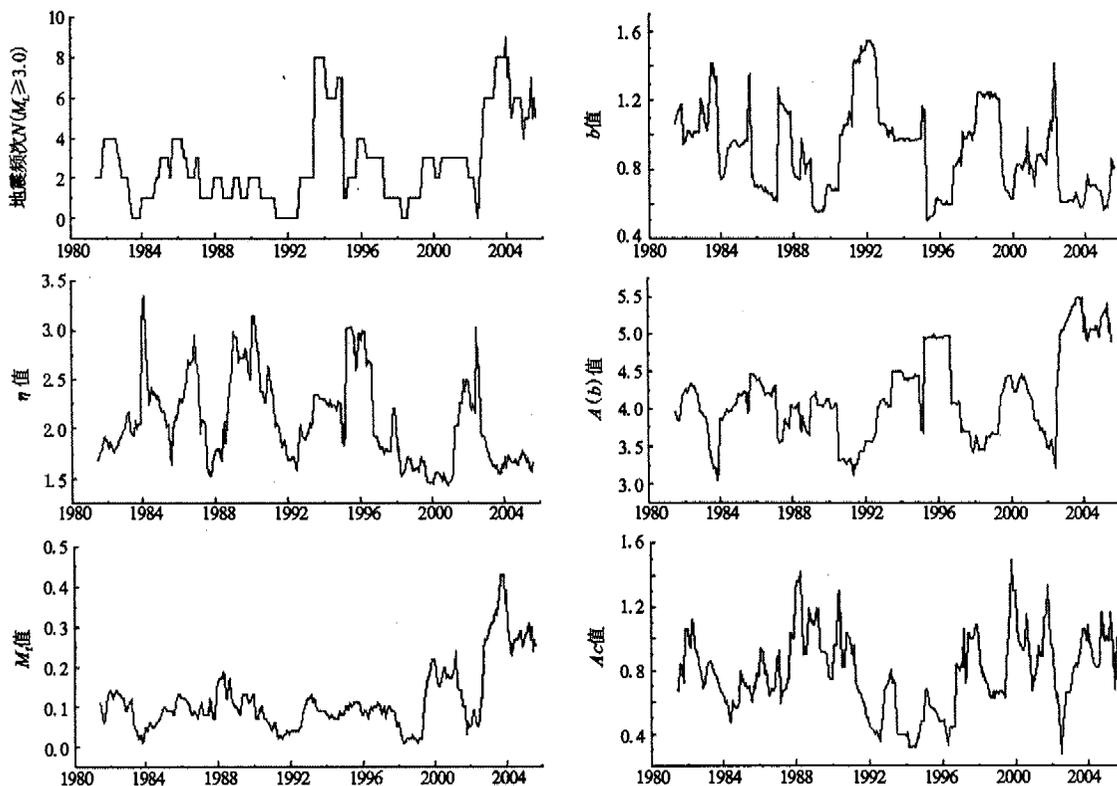


图1 九江 5.7 级地震前后震中周围 200 km 范围内地震频次 $N(M_L \geq 3.0)$ 、 b 值、 η 值、 $A(b)$ 值、 M_f 值、 A_c 值随时间变化

Fig. 1 Variation curves of $N(M_L \geq 3.0)$, b -value, η -value, $A(b)$ -value, M_f -value and A_c -value before and after Jiujiang $M_s 5.7$ earthquake in the area within 200 km from the epicenter.

地震活动增强因子;因子2(F_2)中 η 值的因子载荷较大,反映了地震活动偏离震级-频次(G-R)关系的特征,因此可称为地震活动偏离因子;因子3(F_3)中算法复杂性Ac值的因子载荷较大,反映了地震活动复杂现象的动力学特征,因此可称为地震活动复杂性因子;因子4(F_4)中b值的因子载荷较大,反映了地震活动的震级-频次关系的特征,因此可称为地震活动G-R关系因子。

表1 各因子的特征值与贡献率

因子	特征值	贡献率/%	累计贡献率/%
1	2.955 30	49.255 1	49.255 1
2	1.288 23	21.470 7	70.725 8
3	0.855 78	14.263 1	84.988 8
4	0.58529	9.7549	94.7438
5	0.212 44	3.540 7	98.284 5
6	0.102 93	1.715 5	100.000 0

表2 各参量的初始因子载荷

参量	因子1	因子2	因子3	因子4
N	0.847 431	-0.016 93	-0.311 14	0.295 28
b	-0.542 05	0.599 59	-0.189 02	0.54293
η	-0.188 43	-0.885 66	0.152 77	0.37268
A(b)	0.913 33	-0.170 18	-0.218 11	0.11767
M_L	0.915 60	0.222 41	-0.002 85	-0.099 45
Ac	0.485 12	0.256 15	0.807 67	0.201 75

表3 正交旋转后的因子载荷

参量	因子1	因子2	因子3	因子4
N	0.946 71	0.020 38	0.070 04	-0.028 99
b	-0.233 35	0.142 61	-0.068 46	0.950 97
η	-0.065 85	-0.976 81	-0.057 94	0.142 44
A(b)	0.913 65	-0.014 62	0.093 88	-0.284 22
M_L	0.743 87	0.389 25	0.329 79	-0.289 97
Ac	0.151 26	0.063 17	0.981 46	-0.062 06

图2为九江5.7级地震前,上述6个参量的综合指标 W_{16} 值随时间变化。可以看到该区在震前的20多年里 W_{16} 值一般在0附近波动,基本都小于0.5。

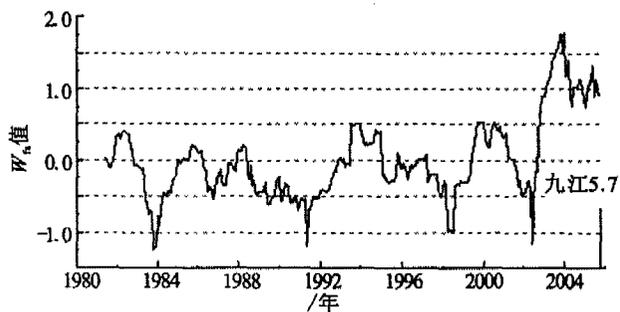


图2 九江5.7级地震前后6参数因子分析综合指标 W_{16} 随时间的变化

Fig. 2 Variation of the synthetic parameter W_{16} from 6 parameters before Jiujiang earthquake.

自2002年10月后出现明显的增高异常,其值超过1.0,这说明这一阶段地震活动出现明显增强。在高值略有恢复的过程中发生5.7级地震。说明综合指标 W_{16} 可以较好地反映该区地震活动异常的综合特征,它在震前出现明显的异常变化。

进一步分析可知,由于 F_1 的方差贡献率已达到49.25%,所包含信息量最大,是4个因子中最重要因子。对该因子起主要作用的是地震频次N($M_L \geq 3.0$)、A(b)值和 M_L 值3个参数,这3个参数在震前的异常反映相对较好(图1),因此使得 W_{16} 值在震前的异常反应较为突出。

3 大华北地区中强以上地震前地震活动性参数的因子分析

1972年以来大华北地区($30^\circ \sim 43^\circ N, 105^\circ \sim 125^\circ E$)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—2005年期间该地区除强余震以外共发生的14次 $M_s \geq 5.7$ 地震(表4),分别取震中周围一定范围内的地震资料计算地震频次N($M_L \geq 3.0$)、b值、 η 值、A(b)值、 M_L 值、Ac值、C值和D值8个参数,并在此基础上进行因子分析,研究这14次中强以上地震前 W_{16} 的异常变化。在研究区域的选取上,为消除人为因素的影响,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($M_L \geq 3.0$)外,其他参数的计算一般选取 $M_L 2.0 \sim 2.3$ 地震,对地震控制能力较差地区起始震级还将适当提高,并剔除余震。资料计算的时间窗大体为18月,步长为1个月。

表4 1972—2000年期间华北地区发生的 $M_s \geq 5.8$ 地震

编号	发震时间	震中位置		震级 M_s	
		纬度	经度		
1	1975-02-04	40°42'	122°42'	辽宁海城	7.3
2	1976-04-06	40°14'	112°12'	内蒙和林格尔	6.3
3	1976-07-28	39°38'	118°11'	河北唐山	7.8
4	1976-09-23	40°05'	106°21'	内蒙巴音木仁	6.3
5	1979-07-09	31°27'	119°15'	江苏溧阳	6.0
6	1979-08-25	41°14'	108°07'	内蒙五原	6.0
7	1981-08-13	40°30'	113°25'	内蒙丰镇	5.8
8	1983-01-07	35°18'	115°36'	山东菏泽	5.9
9	1984-05-21	32°38'	121°36'	南黄海	6.2
10	1989-10-19	39°57'	113°49'	山西大同	6.1
11	1996-05-31	40°42'	109°36'	内蒙包头	6.2
12	1996-11-09	31°42'	123°06'	南黄海	6.1
13	1998-01-10	41°06'	114°18'	河北张北	6.2
14	2005-11-26	29°42'	115°43'	江西九江	5.7

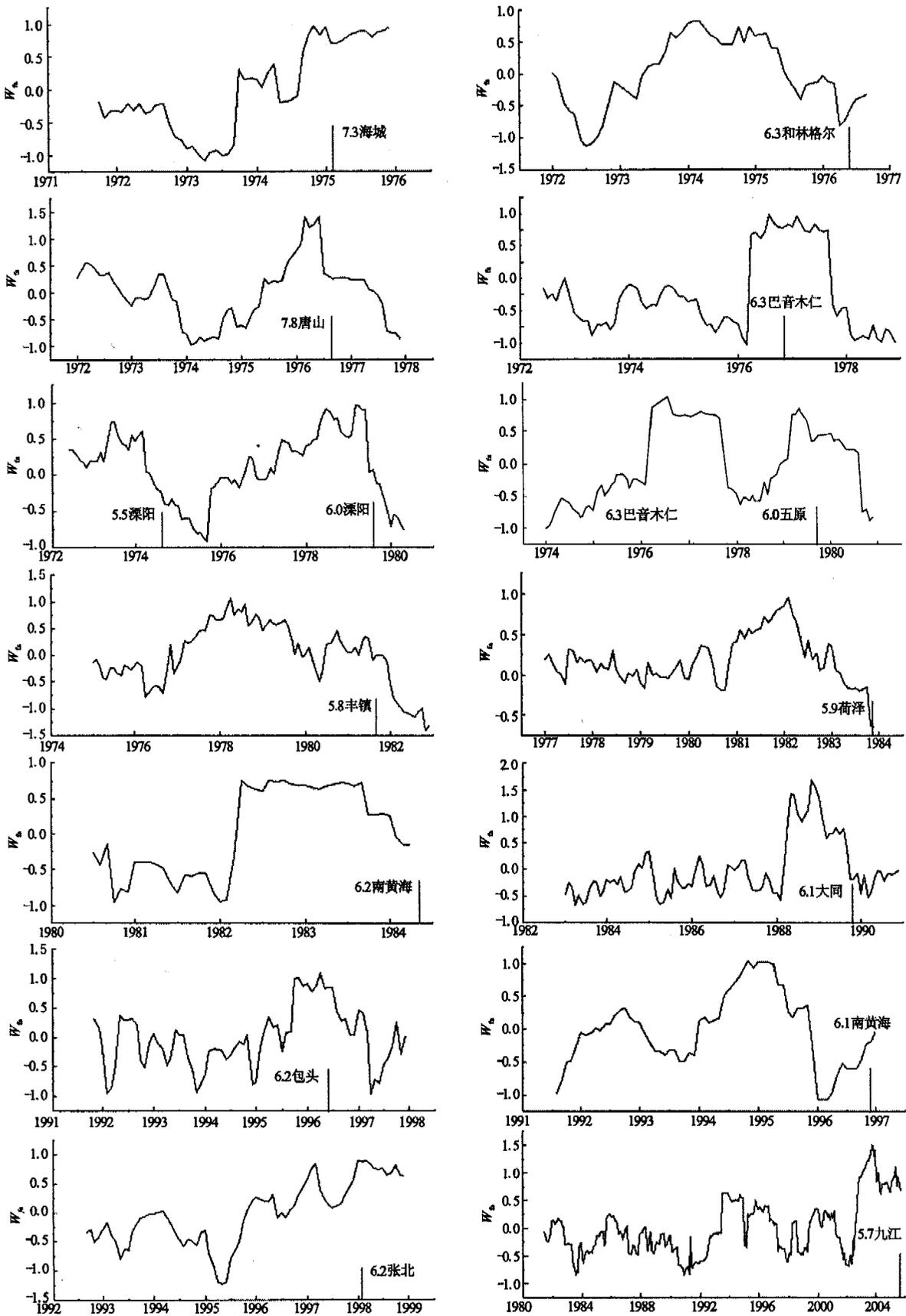


图 3 华北 14 次中强以上地震前综合指标 W_n 的异常变化

Fig. 3 Anomalous variation of the synthetic parameter W_n before the 14 moderate and strong earthquake events in North China.

采用与上述 2005 年江西九江 5.7 级地震震例相同的因子分析方法。取累计贡献率 $E_m > 85\%$ 的最小 $m (m < p)$ 确定公共因子个数 m 。根据本文不同震例的因子分析,通常 $m = 3 \sim 4$ 时就可达到要求。图 3 为各次地震前后 W_{ia} 的变化曲线,可知在正常情况下 W_{ia} 大体在 0 以下,一般不会超过 0.5;但在上述中强以上地震发生前 2~3 年的中期阶段都出现明显的增高异常,异常高值通常在 0.9~1.0 以上,多数震例的异常在震后恢复。

4 讨论与结论

文献[5]、[6]对 1979 年江苏溧阳 6.0 级地震和华北地区其他 12 次中强以上地震前震中周围地区的地震资料进行主成分分析,取得较好的预测效果,表明主成分分析方法也是约简预报参量的有效工具。那么因子分析与主成分分析有哪些区别和联系?

主成分分析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;因子分析则是主成分分析方法的推广和深化,通过研究一组观测数据的相关矩阵内部结构,找出对变量起支配作用的几个互不相关因子,在尽量减少原始数据信息损失的前提下,用少数几个主要公共因子去代替数量较多且有一定联系的原始变量,从而达到揭示原始变量之间的内在联系、合理解释变量与主要因子关系的目的。

因子分析与主成分分析也有很大差别:主成分分析是将主成分表示为原观测变量的线性组合;而因子分析是将原观测变量表示为公共因子的线性组合。主成分分析的主成分数 m 和原变量数 p 相等,是将一组具有相关性的变量变换为一组独立的综合变量;而因子分析的目的是要使公共因子数 m 比原变量数 p 小,而且要尽可能地选取小的 m ,以便尽可能地构造一个结构简单的模型。在主成分分析中,原观测变量对某一主成分的影响大小,由该主成

分相应的特征向量确定;而在因子分析中,原观测变量在某一公共因子上的载荷,由该公共因子相应的特征向量确定;此外,与主成分分析相比因子分析可以使用旋转技术帮助解释因子,在解释方面更具有优势。大致说来,当需要寻找潜在的因子并对这些因子进行解释的时候,更加倾向于使用因子分析,并且借助旋转技术帮助更好解释。

研究结果表明,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,但是它们之间有一定的相关性。无论在正常情况下还是异常情况下各个参数的变化是各不相同,这给分析预报工作带来不必要的麻烦。而因子分析法可在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,通过少数的公共因子表示原有的多维变量。本文将多个公共因子根据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_{ia} 值。该指标包含了上述参数中的绝大部分信息,可以较好地反映出中强以上地震前地震活动时、空、强特征的异常变化。

[参考文献]

- [1] 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.
- [2] 韩渭宾. 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.
- [3] 方开泰. 实用多元统计分析[M]. 上海: 华东师范大学出版社, 1989. 286-295.
- [4] 唐启义, 冯光明. 实用统计分析及其 DPS 数据处理系统[M]. 北京: 科学出版社, 2002. 367-373.
- [5] 王炜, 林命遇, 马钦忠, 等. 主成分分析法在地震预报中的应用研究[J]. 中国地震, 2005, 21(3): 409-416.
- [6] 王炜, 马钦忠, 林命遇, 等. 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.
- [7] 吕悦军, 陆远忠, 郑悦君. 用算法复杂性分析地震活动演化特征[J]. 地震, 1997, 17(1): 25-33.