

# 应用混合极值理论及最大似然法估计 中国大陆地震危险性<sup>\*</sup>

陈 虹 黄忠贤

(中国北京 100085 国家地震局地壳应力研究所)

主题词 地震危险性；地震活动；地震综合预报

极值理论由于其不需要一段时间内的全部观测资料,而只需各个短时间段内的极大值,故在利用历史资料及现代资料分析地震活动性中有其独特的优越性.

冈贝尔(Gumbel)的极值统计方法在地震活动性研究中得到了广泛的应用.其中,冈贝尔 I 型( $G_1(x)$ )和冈贝尔 III 型( $G_{\text{III}}(x)$ )已被国外许多专家多次用于研究地震活动性.

陈培善和林邦慧(1973)在 $G_1$ 的基础上考虑到地震的发生具有上限,提出了一种修正型极值分布函数,国内有多人使用陈培善修正型极值分布函数分析地震趋势(周冉、林邦慧, 1992).

Fahmi 和 Al-Abbasi 在使用  $G_1$  及  $G_{\text{III}}$  型极值分布函数研究地震活动性时发现,在震级大的区间上,  $G_1$  及  $G_{\text{III}}$  都不能很好地拟合实际资料,于是提出了一种将  $G_1$  及  $G_{\text{III}}$  有限混合起来的冈贝尔混合分布函数  $G_{\text{mix}}$ . 并将其用于研究伊拉克和埃及的地震活动性中. 结果表明,  $G_{\text{mix}}$  能较好地拟合实际地震资料 (Fahmi and Al-Abbasi, 1990).

目前,还没有人将冈贝尔混合极值分布函数用于分析中国大陆地震危险性.对此,我们将中国大陆按地震活动性分布分成 14 个分区,并将混合冈贝尔极值分布函数用于预测这 14 个地震区今后 5 年(1994—1998)的地震危险性.同时,为验证混合极值分布函数预测中国大陆地震危险性结果的可靠性及适用性,对这 14 个分区 1964 年至 1993 年底按 5 年一个间隔分成 6 个区间做了预测检验,并给出每个地震区的综合预报准确率及危险阈值.针对各个地震区地震活动的不均匀性,对预测结果的影响进行了分析与讨论.

## 1 极值分布函数

冈贝尔极值分布函数用于地震活动性分析具有如下几种形式:

$$(I) \quad G_1(x) = \exp\{-\exp[A(x-B)]\} \quad (1)$$

其中,  $A, B$  都是统计常数, 在这种形式的极值分布中, 极值既无上限, 又无下限.

$$(II) \quad G_{\text{III}}(x) = \exp\{-[(W-X)/(W-U)]^K\} \quad (2)$$

可以看到,  $-\ln[-\ln(G(x))]$  相对震级是线性分布,而  $-\ln[-\ln(G_{\text{III}}(x))]$  则是一条向上凸的曲线.当震级较大时  $G_1(x)$  分布相对观测值偏高,而  $G_{\text{III}}(x)$  分布相对观测值偏低,于是 Fahmi 等提出将  $G_1$  与  $G_{\text{III}}$  线性组合的冈贝尔混合型分布函数.图 1 给出了青藏西部的实际资料及  $G_1$ ,  $G_{\text{III}}$ ,  $G_{\text{mix}}$  拟合结果.可以看出,  $G_{\text{mix}}$  能更好地拟合实际资料.

(III)  $G_{\text{mix}}$  的形式

\* 1994 年 3 月 14 日收到初稿, 1994 年 7 月 11 日收到修改稿并决定采用.

$$G_{\text{mix}}(x) = P_1 G_1(x) + P_2 G_{\text{III}}(x)$$

由于  $P_1 + P_2 = 1$

$$G_{\text{mix}}(x) = P_1 G_1(x) + (1 - P_1) G_{\text{III}}(x) \quad (4)$$

$$= P_1 \exp\{-\exp[-A(x - B)]\} + (1 - P_1) \exp\{-[(W - x)/(W - U)]^K\} \quad (5)$$

其中,  $A, B, W, U, K$  均为统计常数,  $P_1$  为混合比例因子. 可以看到,  $G_1$ ,  $G_{\text{III}}$  及  $G_{\text{mix}}$  参与拟合的参数个数分别为 2, 3 和 6 个.

## 2 参数的估计方法

有多种方法可用于估计极值分布参数, 其中最小二乘法和最大似然法是最常用的. 目前, 国内地震研究中都采用最小二乘法来估算极值分布参数, Abbasi 将最大似然法和最小二乘法同时用于估计伊拉克地震的极值分布参数, 并证实, 最大似然法估计从总体上看拟合效果比最小二乘法好 (Al-Abbasi and Fahmi, 1985). 本文采用最大似然法估计  $G_{\text{mix}}$  极值分布函数中的 6 个参数.

最大拟然法的基本思路为, 已发生的随机事件 (如地震) 一定来自使它们出现概率最大的那个总体, 用于求极值分布函数中的参数时, 首先要给出极值分布函数的最大似然函数, 将最大似然函数相对其分布参数求导, 令其导数为零, 通过迭代而求得. 对于  $G_{\text{mix}}(x)$  其最大似然函数为 (Fahmi and Al-Abbasi, 1990).

$$L = \prod_{j=1}^n [p_1 g_1(x) + (1 - P_1) g_2(x)] \quad (6)$$

其中

$$g_1(x) = G_1'(x) = A \exp[-A(x - B)] - \exp[-A(x - B)] \quad (7)$$

$$g_2(x) = G_{\text{III}}'(x) = [k/(W - U)] [(W - x)/(W - U)]^K \exp\{-[(W - x)/(W - U)]^K\}$$

## 3 统计预报及危险性计算

实际运用时, 将一段时间内的观测资料按某个单位时间划分 (如 1 年), 并选取各个单位时间内的一个最大地震值, 如震级为  $x_j (j=1, 2, \dots, n)$  共  $n$  个, 把这  $n$  个观测值按震级大小排列起来

$$x_1 \leq x_2 \leq \dots \leq x_j \leq \dots \leq x_n$$

第  $j$  位观测值是个顺序统计量, 也是一个随机变量, 其分布函数为  $G(x_j)$ , 如果  $G(x)$  的值已知, 则利用  $x_j$  值及  $G(x_j)$  的估计值, 利用最大似然法进行拟合, 从而可确定其分布函数中的各个参数.

冈贝尔提出的  $G(x_j)$  无偏估计值为

$$G(x_j) = \frac{j}{n+1} \quad (j = 1, 2, \dots, n) \quad (8)$$

当分布函数中各个参数确定后, 可利用式(8)确定震级  $\geq x$  的地震平均复发周期为

$$T(x) = (1 - G(x))^{-1} \quad (9)$$

在  $D$  年内的地震危险性, 也就是说  $D$  年内发生震级  $\geq x$  的概率为

$$R_D(x) = 1 - G_D(x)$$

对于冈贝尔混合极值分布函数,  $D$  年内的地震危险性为

$$R_{\text{mix}}(x | D) = 1 - (G_{\text{mix}}(x))^D \quad (10)$$

考虑到某地区一个震级为  $x$  的地震发生后, 该地区中的  $M \leq x$  的地震能量得到释放, 从这一时间

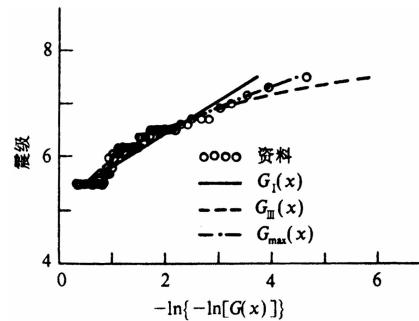


图 1 青藏西部(1989—1993 年)的拟合结果



ACTA SEISMICA

ACTA SEISM

开始重新积累, 同一地区再发生震级为  $x$  地震的危险性随时间的增大而增加(不考虑震群或强余震的情况), 故在计算未来几年中震级为  $x$  的地震的危险性时, 时间段  $D$  的选取是这样考虑的: 选取上次大于等于  $x$  级地震发生的时间到所要预测的时间间隔作为  $D$ , 我们用此方法检验了 1964 年到 1993 年间的地震危险性, 并与实际发生的地震对照, 结果表明此方法是可行的.

#### 4 对中国大陆各个地震区的地震危险性预测检验

中国大陆各个地震区的划分见图 2. 我们对各个地震区 1964 年到 1993 年底, 按 5 年一个间隔分成 6 个区间. 利用最大似然法分别计算每个地震区在 6 个区间内的混合极值分布参数, 再利用式(10)计算了其地震危险性, 并与实际地震发生情况进行了对比, 如: 我们利用帕米尔地区 1890 年至 1963 年底的地震资料, 算其极值分布参数并算出其未来 5 年(1964 年至 1968 年底)的地震危险性, 并与帕米尔地区在 1964 年至 1968 年底期间实际发生的地震进行对比. 以此类推, 我们算出了各个地震在 6 个时间区间内的地震危险性的估计值. 表 1 中列出了几个有代表性地区的计算结果与实际发震情况.

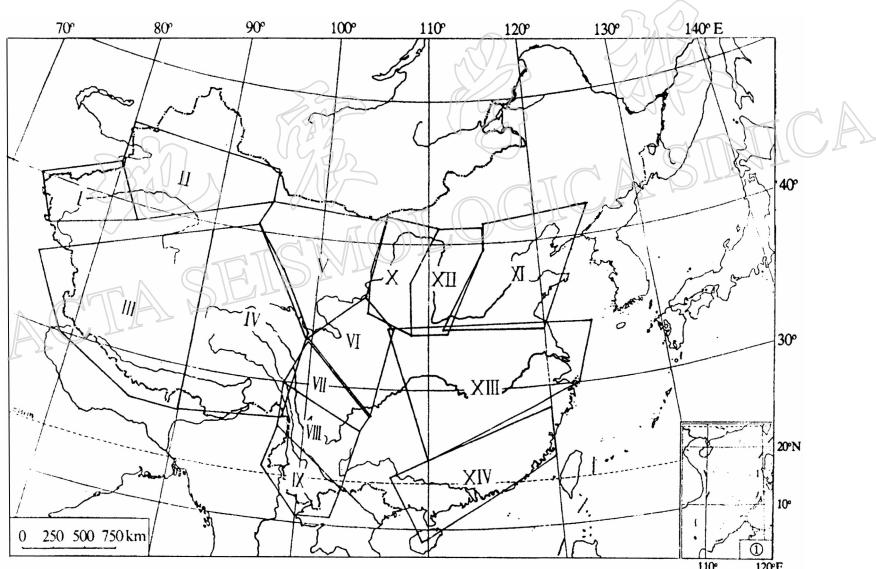


图 2 中国大陆各个地震区的划分. 图中①为南海诸岛

将表 1 中的地震危险估计值与实际发震情况对比, 我们可以得到一个危险阈值, 当所算出的危险概率大于该值时, 认为有地震危险; 反之, 则地震危险较小. 这样就可用于中长期地震预报. 不同地震区其地震危险阈值是不同的, 有的地区较高, 有的地区较低, 其原因主要是: ① 各个地震区地震活动性强弱不同, 地震活动性强, 其地震复发周期短, 地震危险概率大; 反之, 则小. ② 资料截取的长短对危险性估计也有影响. 资料截取偏短, 则计算结果偏大; 反之, 则偏小. 故从表 1 可以看到, 有些地区平均概率值很高(如青藏东部), 而有的地区平均概率值较低, 并不意味着就没有地震危险(如南北带中段). 故合理地判断不同地震区其地震危险阈值, 对预测未来地震危险性是很重要的. 在判定各个地震区地震危险阈值时, 我们对同一地区大致选取几个值, 并算出这些值下的  $n_{11}, n_{10}, n_{00}, n_{01}$ , 将其代入式(11), 求出综合预报准确率  $f$  选取使  $f$  为最大的值作为危险阈值. 这样得到了各个地震区危险阈值及综合预报准确率值(见表 2). 采用陈培善和林邦慧(1973)提出的综合衡量预报准确率公式

$$f = \frac{f_0 - pq + 0.25}{1 - pq + 0.25} \quad (11)$$

其中

$$f_0 = \frac{n_{11}}{n_{11} + n_{10}} \cdot \frac{n_{11}}{n_{11} + n_{01}} q + \frac{n_{00}}{n_{00} + n_{10}} \cdot \frac{n_{00}}{n_{00} + n_{01}} p$$

$$q = \frac{n_{10} + n_{00}}{n} \quad p = \frac{n_{11} + n_{01}}{n}$$

$n_{11}$ 为报有且有震的次数;  $n_{10}$ 为报有而无震的次数, 即虚报数;  $n_{00}$ 为报无且无震的次数;  $n_{01}$ 为报无而有震的次数, 即漏报数;  $n = n_{11} + n_{10} + n_{00} + n_{01}$ 即每个地震区的总预报数.

表 1 中国大陆各个地震区预测检验结果

I 青 藏 东 部							
检验时间 $t/a$	资料选取时间 $t/a$	在检验时间间隔内地震危险性估计及实际发生情况					
		5.5	6.0	6.5	7.0	7.5	8.0
1989—1993	1890—1988	0.94Y	0.79Y	0.55Y	0.32N	0.76N	0.45N
1984—1988	1890—1983	0.93Y	0.78Y	0.66N	0.85Y	0.71N	0.41N
1979—1983	1890—1978	0.95Y	0.82Y	0.84Y	0.78N	0.67N	0.38N
1974—1978	1890—1973	0.95Y	0.82Y	0.66N	0.69N	0.61N	0.39N
1969—1973	1890—1968	0.94Y	0.81Y	0.75Y	0.54N	0.55N	0.30N
1964—1968	1890—1963	0.89Y	0.74Y	0.53Y	0.33N	0.47N	0.26N
II 川 滇 西 部 北 区							
检验时间 $t/a$	资料选取时间 $t/a$	5.5	6.0	6.5	7.0	7.5	8.0
		0.91Y	0.76Y	0.74Y	0.47N	0.31N	
1989—1993	1900—1988	0.87N	0.72Y	0.56N	0.32N	0.26N	
1984—1988	1900—1983	0.93N	0.80Y	0.63N	0.37Y	0.15N	
1979—1983	1900—1978	0.80Y	0.61Y	0.40N	0.21Y	0.08N	
1974—1978	1900—1973	0.76Y	0.64Y	0.41N	0.51N	0.22Y	
1969—1973	1900—1968	0.97Y	0.88Y	0.68Y	0.41N	0.17N	
1964—1968	1900—1963						
III 山 西 带							
检验时间 $t/a$	资料选取时间 $t/a$	在检验时间间隔内地震危险性估计及实际发生情况					
		5.0	5.5	6.0	6.5	7.0	7.5
1989—1993	1800—1988	0.43Y	0.46Y	0.30Y	0.73N		
1984—1988	1800—1983	0.42Y	0.31N	0.23N	0.75N		
1979—1983	1800—1978	0.49Y	0.28Y	0.13N	0.72N		
1974—1978	1800—1973	0.64N	0.44N	0.60Y	0.68N		
1969—1973	1800—1968	0.42N	0.31N	0.59N	0.71N		
1964—1968	1800—1963	0.43Y	0.40Y	0.43N	0.63N		

注: Y 代表发生了震级  $\geq x$  的地震; N 代表没发生震级  $\geq x$  的地震.

综合衡量预报准确率  $f$ , 综合考虑了报对数, 包括有震报对和无震报对( $n_{11}$  和  $n_{00}$ )、虚报数( $n_{10}$ )、漏报数( $n_{01}$ )以及地震发生次数多少的因素, 并将“瞎碰准确率”归到同一水平, 当  $f=0.2-0.25$  时, 则认为这种方法还没有找到与地震的关系, 报对也是偶然碰上而已.  $f$  并不代表报对的地震比率, 它是衡量预报有效程度的一个综合指标.

对 14 个地震区的预测检验结果汇总于表 2. 可以看到: 综合预报准确率较高的地区有青藏东部(0.79)、青藏西部(0.76)、南北带中段(0.71). 这三个地区地震活动的特点是无明显活跃期与平静期. 这与极值理论假定事件的发生为随机事件的条件相符合. 综合预报准确率最低的地区是山西带(0.26)与南北带北段(0.26), 这两个地区的地震活动特点为平静期与活跃期很明显. 在活跃期内大震连续发生, 平静期内中强震都很少发生. 对这种地震在时间轴上分布极不均匀的地区, 极值理论无法给出满

意的结果. 预测检验结果比较好的地区还有川滇西部北区(0.64)、天山地区(0.58)、东南沿海(0.53).

表 2 各个地震区检验情况汇总

地 震 区	危险阈值	报有有震	虚报数	报无无震	漏报数	综合预报准确率
I. 帕米尔地区	0.60	14	9	11	2	0.50
II. 天山地区	0.75	8	5	21	2	0.58
III. 青藏西部	0.75	9	1	12	2	0.76
IV. 青藏东部	0.73	15	2	17	2	0.79
V. 祁连山地区	0.60	4	6	23	3	0.38
VI. 南北带中段	0.40	8	2	24	2	0.71
VII. 川滇西部北区	0.60	12	3	12	3	0.64
VIII. 川滇西部中区	0.60	6	3	20	7	0.44
IX. 川滇西部南区	0.50	10	0	5	9	0.44
X. 南北带北段	0.55	1	2	29	4	0.26
XI. 华北平原渤海	0.46	5	4	15	6	0.38
XII. 山西带	0.40	7	11	4	2	0.26
XIII. 长江中下游	0.60	2	2	12	2	0.40
XIV. 东南沿海	0.50	6	5	23	2	0.53

综上所述, 在我们所研究的 14 个地震区内, 除 山西带和南北带北段完全不适用以外, 此方法在其它 12 个区中都有不同程度的预测能力, 在大约半数的区域, 该方法的预测效果还是比较好的.

## 5 中国大陆未来 5 年(1994—1998 年)中强地震趋势估计

利用直至 1993 年底的地震记录, 计算了中国大陆未来 5 年中强震危险, 结果示于表 3, 可以看到: 帕米尔地区、天山地区、青藏东区在未来 5 年里发生 7 级以上地震的可能性较大. 华北平原渤海地区、东南沿海、川滇西部北区有发生 6.5 级地震的可能. 祁连山地区发生 6.0 级以上地震, 以及长江中下游地区发生 5.5 级地震的可能性都不大.

表 3 中国各个地震区 1994—1998 年地震危险性估计

地 震 区	资料选取时间 <i>t/a</i>	震 级						危 险 阈 值
		5.5	6.0	6.5	7.0	7.5	8.0	
帕米尔地区	1890—1993	0.91	0.71	0.56	0.49	0.87	0.51	0.55
天山地区	1890—1993	0.83	0.58	0.33	0.83	0.72	0.42	0.75
青藏西区	1890—1993		0.78	0.47	0.67	0.62		0.75
青藏东区	1890—1993	0.95	0.80	0.64	0.56	0.81	0.50	0.73
祁连山地区	1800—1993	0.51	0.32	0.27	0.14	0.34	0.20	0.60
南北带中段	1800—1993	0.77	0.53	0.53	0.29	0.35	0.27	0.40
川滇西部北区	1900—1993	0.93	0.84	0.64	0.60	0.40		0.60
川滇西部中区	1700—1993	0.63	0.75	0.51	0.51	0.27	0.58	0.60
川滇西部南区	1800—1993	0.62	0.45	0.29	0.16	0.07		0.50
华北平原渤海	1800—1993	0.61	0.68	0.59	0.37	0.21		0.46
长江中下游	1800—1993	0.38	0.25	0.33				0.60
东南沿海	1800—1993	0.34	0.62	0.65	0.29	0.37		0.50

陈培善教授对本文进行了审阅并提出宝贵意见, 在此深表感谢.

## 参 考 文 献

- 陈培善、林邦慧, 1973. 极值理论在中长期地震预报中的应用. 地球物理学报, **16**, 6—24.
- 周冉、林邦慧, 1992. 用极值理论对京-津-唐地区及 7 个主要地震带未来地震活动的分析. 地震学报, **14**, 增刊, 710—715.
- Al-Abbasi, J. N. and Fahmi, K. J. , 1985. Estimating maximum magnitude earthquakes in Iraq using extreme values statistics. *Geophys. J. R. astr. Soc.*, **82**, 535—548.
- Fahmi, K. J. and Al-Abbasi, J. N. , 1987. Mixture distributions—an alternative approach for estimating maximum magnitude earthquake occurrence. *Geophys. J. R. astr. Soc.*, **89**, 741—747.
- Fahmi, K. J. and Al-Abbasi, J. N. , 1990. Application of a mixture distribution of extreme values to earthquake magnitudes in Iraq and conterminous regions. *Geophys. J. Int.*, **109**, 209—217.

地 省 学 院 研 究 所  
ACTA SEISMOLOGICA SINICA