

6-17
基于均生函数的数学预测模型在地震
中期预报中的应用马文静¹, 张晓清², 张晓东³(1. 青海省地震局, 青海 西宁 810001; 2. 中国科学技术大学地球和空间科学系,
安徽 合肥 230026; 3. 中国地震局分析预报中心, 北京 100036)

摘要:利用均生函数数学建模方法拟合地震的震级或能量序列, 进行地震中期趋势预测. 由于引入了“双评分准则”, 即在拟合时考虑了数量预报和趋势预报的平衡问题, 选取合适的调整系数, 从而提高了地震中期趋势预测的准确率. 另外采用了分级拟合的思路, 利用残差序列逐级提高拟合的精度和速度. 利用该方法对中国大陆地区未来1~3年的地震活动趋势进行了预测.

关键词: 中期预报; 趋势预报; 周期分析; 数学模型; 均生函数**中图分类号:** P315.75 **文献标识码:** A **文章编号:** 1000-0844(2000)01-0062-06

0 前言

周期分析是地震中长期预测的主要方法之一. 归纳起来可分为2类: 一类是对已发生的地震序列进行定性分析, 以经验划分周期进行外推预测; 另一类就是从地震序列本身出发, 利用各种数学方法拟合并进行预测. 在以往用数学模型拟合时存在2个比较突出的问题, 一是达到一定的精度要求, 但需要多级拟合, 就是说自变量的数量很大; 二是内符检验效果很好, 但外推结果与实际情况差别较大. 针对上述问题, 本研究借鉴气象学中用于中长期预测的数学方法, 结合地震预测在某些情况下只要求趋势准确而非数值精确的特点, 提出了一种新的地震趋势预测的数学方法.

1 方法介绍

利用均生函数和方差分析的数学方法建模, 进行周期性外延对时间序列作出多步外推. 拟合时不仅考虑了数量预报和趋势预报平衡(双评分准则), 同时还考虑了残差与波数的平衡, 后者根据能否提高分级拟合率而定.

1.1 用均生函数构造时间序列

设时间序列

$$x(t) = \{x(1), x(2), \dots, x(N)\}$$

式中: N 为样本量. 其均生函数定义为:

收稿日期: 1999-04-15

基金项目: 中国地震局“九五”攻关项目 95-04-06 项目资助

作者简介: 马文静(1964-), 女(汉族), 副研究员, 现主要从事地震预报研究工作.

$$\overline{x_L}(i) = \frac{1}{n_L} \sum_{j=0}^{n_L-1} x(i+jL) \quad (1)$$

$$(i = 1, 2, 3, \dots, L, \quad 1 \leq L \leq M)$$

式中: n_L 为满足 $n_L \leq \frac{N}{L}$ 的最大整数, M 为不超过 $\frac{N}{2}$ 的最大整数.

当 L 的值取 1 到 M 时, 可产生一系列的均值生成函数, 可用一个三角矩阵来表示:

$$H = \begin{bmatrix} \overline{x_1}(1) & \overline{x_2}(1) & \overline{x_3}(1) & \cdots & \overline{x_M}(1) \\ & \overline{x_2}(2) & \overline{x_3}(2) & \cdots & \overline{x_M}(2) \\ & & \overline{x_3}(3) & \cdots & \overline{x_M}(3) \\ & & & \ddots & \\ & & & & \overline{x_M}(M) \end{bmatrix} \quad (2)$$

由此可见, 均值生成函数是由时间序列按一定的时间间隔计算均值而派生出来的.

取不同的周期 ($L = 1, 2, \dots, M$), 用均值生成函数循环外推构造延拓序列, 即为在一定周期下构造的新序列:

$$x_L(t) = \overline{x_L} \left[i - \text{INT} \left(\frac{i-0.5}{L} \right) > L \right] \quad (3)$$

$$i = 1, 2, \dots, N, \quad L = 1, 2, \dots, M$$

1.2 拟合周期的选取

用不同周期的均生函数所构造的序列不同, 如何选取合适的周期, 使拟合的效果达到最佳是一个关键的问题. 本文从 2 个方面入手来解决这一问题.

1.2.1 双评分准则

对于地震的中长期预测要求是“大趋势”或“大方向”基本准确. 用传统的 F 检验、AIC 和 BIC 准则等方法是很难达到这种要求的. 双评分准则 CSC (Couple Score Criterion) 是从数量预报评分和趋势预报评分 2 个方面来权衡变量的筛选或阶数的确定, 采用隐式表达, 即 $\text{CSC} = S_1 + S_2$, 其中: S_1 表示数量评分, 称之为精评分; S_2 表示趋势评分, 即粗评分. 双评分准则旨在使精评分和粗评分之和达到最小. 该准则有 3 种不同的表达式, 本文采用第一种, 即 CSC_1 准则:

$$\text{CSC}_1 = \left(1 + \lambda \frac{N}{N_k} \right) Q_k \quad (4)$$

其中: N 为样本量; Q_k 为平均残差平方和; N_k 为趋势评分.

$$Q_k = \frac{1}{N} \sum_{t=1}^N [x(t) - \overline{x_L}(t)]^2 \quad (5)$$

如果预报考虑升平降 3 种趋势, 则趋势评分定义为:

$$N_k = \sum_{t=2}^N \left\{ 1 - \frac{1}{2} \left| C_u[\Delta x(t)] - C_u[\Delta \overline{x_L}(t)] \right| \right\} \quad (6)$$

式中: $\Delta x(t) = x(t+1) - x(t)$, $\Delta \overline{x_L}(t) = \overline{x_L}(t+1) - \overline{x_L}(t)$

$$C_u[\Delta x(t)] = \begin{cases} 1 & \Delta x(t) > u \\ 0 & |\Delta x(t)| \leq u, \quad u > 0 \\ -1 & \Delta x(t) < -u \end{cases} \quad (7)$$

u 是判定不同趋势的标准, 可根据预报问题凭专业知识确定. λ 为调整系数, 表示数量预报和趋势预报的权重. $\lambda = 1.0$ 时, 二项权重相等; $\lambda > 1.0$ 时, 趋势预报占的比重较大, 这一点对地

震的中长期预报而言是非常重要的. 在进行拟合时, 取不同的周期, 可构造不同的序列, 根据双评分准则, 计算其 CSC_1 值. 取 CSC_1 值最小的周期作为一级拟合的周期.

1.2.2 利用残差序列逐步提高拟合的精度

为了提高拟合的精度, 采用分级拟合, 即在第一级拟合中选取 CSC_1 值最小的周期 L 作为第一级的拟合周期, 拟合率为:

$$D_L^{(1)} = \frac{N_L^{(1)}}{N} \quad (8)$$

然后构造残差序列

$$x^{(1)}(t) = x(t) - \overline{x_L}(t) \quad (9)$$

对残差序列重复上述步骤, 选取第二级拟合周期, 求出 $D_L^{(2)}$. 若 $D_L^{(2)} > D_L^{(1)}$, 则继续进行下一级的拟合; 当 $D_L^{(i)} \leq D_L^{(i-1)}$ 时, 计算结束, 以 $x_L^{(1)}(t), x_L^{(2)}(t), \dots, x_L^{(i-1)}(t)$ 序列相加值作为 $x(t)$ 的拟合序列, $D_L^{(i-1)}$ 为历史拟合率.

1.2.3 最优周期的选择

若一时间序列 $x(t) = \{x(1), x(2), \dots, x(N)\}$ 存在长度为 L 的周期, 那么由此周期构成的新序列是由 $x(1), x(2), x(3), \dots, x(L), x(L+1), x(L+2), x(L+3), \dots, x(2L), \dots$ 加上随机项构成的. 显然, 序列中 $x(L)$ 与 $x(L+1), x(2L+1), \dots$ 数据相近, $x(2)$ 与 $x(L+2), x(2L+2), \dots$ 相近. 依此类推, 不同周期相同序号上的数据之间的方差称为组内方差, 同一周期内数据之间的方差称为组间方差. 根据方差分析的基本原理, 可用 F 检验来判断是否存在长度为 L 的周期:

$$F^{(L)} = \frac{S^{(L)}/(L-1)}{S/(N-L)} \quad (10)$$

$$S^{(L)} = \sum_{i=1}^L n_L [\overline{x_L}(i) - \overline{x}]^2$$

$$S = \sum_{i=1}^L \sum_{j=1}^{n_L} \{x[i + (j-1)L] - \overline{x_L}(i)\}^2$$

其中: n_L 为满足 $n_L \leq \frac{N}{L}$ 的最大整数. 对于给定的置信水平 α , 如果 $F^{(L)} > F_\alpha$ (自由度为 $L-1, N-L$), 则认为隐含长度为 L 的优势周期.

对地震序列而言, 可能存在多个周期, 计算不同周期下的 $F^{(L)}$ 值, 将通过检验的 $F^{(L)}$ 值从大到小排序, 根据实际需要选取所需的主要优势周期.

2 预测模型在地震趋势预测中的应用

在地震中期趋势预测(1~3年尺度)中, 需要关注的问题是某一区域内地震的强度和释放总能量, 以此确定该区是处于活跃期、平静期, 还是过渡期. 因此可以用某区域年最大震级或年释地震总能量构造序列, 用上述方法对其进行拟合和外推预测.

2.1 预测模型检验

为验证该方法的稳定性和确定各项参数的取值范围, 选取中国大陆地区年最大震级构成的时间序列进行试算, 其结果如下:

- (1) 选取不同的 λ 值, 即 $\lambda = 1.0 \sim 1.5$, 计算结果表明, 当 $\lambda = 1.0$ 时的拟合效果最好.
- (2) 选取不同的 u 值(判定趋势的阈值), 分别取 $u = 0.2, 0.5, 1.0, 2.0$, 计算结果表明, u

$= 0.5$ 是比较合适的值。

(3) 取不同的资料段进行试算, 结果基本稳定, 但所取资料时段越长, 预测的效果越好。

2.2 展望性检验的结果分析

以 1900~1978 年的年最大震级为原始序列进行展望性检验, 即预测未来 3 年的年最大震级, 增加 1 年的数据, 依次向后推 1 年预测未来 3 年的值, 共得到 20 组展望性检验的结果, 即 1979~1998 年年最大震级预测值, 见表 1。

表 1 展望性检验结果

序号	资料时段	预测时段	趋势预测			数值预测误差			拟合 级数	拟合率	M	$M_{y,1}$	$\overline{M}_y$	ΔM
			1	2	3	1	2	3						
1	1900~1978 年	1979~1981	✓	•	/	0.7	-0.6	-0.2	6	0.98		6.1		
2	1900~1979 年	1980~1982	✓	/	•	0.2	-0.4	-1.5	5	0.99		6.4		
3	1900~1980 年	1981~1983	✓	•	/	0.4	1.7	0.3	5	0.98	6.9	7.3	7.2	0.3
4	1900~1981 年	1982~1984	•	•	•	-1.5	-0.5	-1.0	5	0.96	6.6	8.2	8.1	1.5
5	1900~1982 年	1983~1985	•	•	/	0.1	-1.2	-0.8	5	0.98	6.8	6.7	7.0	0.2
6	1900~1983 年	1984~1986	/	/	•	0.1	0.2	-1.6	2	0.82	6.2	6.1	6.9	0.7
7	1900~1984 年	1985~1987	/	•	/	0.2	-1.6	0.4	2	0.82	7.4	7.2	7.5	0.1
8	1900~1985 年	1986~1988	•	✓	/	-1.3	-0.2	0.2	6	0.99	6.5	7.9	8.0	1.5
9	1900~1986 年	1987~1989	✓	✓	/	0.8	0.2	-0.9	6	0.99	6.4	5.5	6.0	-0.4
10	1900~1987 年	1988~1990	•	•	/	1.9	0.3	-1.5	5	0.98	7.6	5.7	6.8	-0.8
11	1900~1988 年	1989~1991	✓	/	/	0.7	0.7	0.3	6	0.98	6.7	7.4	7.1	0.4
12	1900~1989 年	1990~1992			/	1.3	0.3	-1.6	7	0.99	7.0	5.7	7.3	0.3
13	1900~1990 年	1991~1993	✓	/	/	-0.5	-2.0	0.2	5	0.99	6.5	7.0	6.5	0.0
14	1900~1991 年	1992~1994	✓	/	/	0.1	-0.2	-1.0	6	0.99	6.9	6.8	8.1	1.2
15	1900~1992 年	1993~1995	✓	/		-0.2	-1.0	1.0	6	0.99	6.6	6.8	6.7	0.1
16	1900~1993 年	1994~1996	✓		•	-0.7	1.2	-0.6	5	0.99	7.3	8.0	8.2	0.9
17	1900~1994 年	1995~1997	•	•	•	1.4	-0.4	0.2	5	0.99	7.3	5.9	6.1	-1.2
18	1900~1995 年	1996~1998	/	/	•	0.9	0.2	-1.4	2	0.84	7.1	6.2	7.1	0.0
19	1900~1996 年	1997~1998	/	•		0.2	-1.4		2	0.84	7.5	7.4	7.4	-0.1
20	1900~1997 年	1998	✓			0.5			5	0.98	6.6	6.4	7.5	0.9

注: 以震级差 0.5 为标准; ✓ 表示与实际趋势一致; • 表示相反; 1、2、3 为外推年份; $M_{y,1}$: 外推第 1 年的值;

$\overline{M}_y$: 3 次预测的平均值; M : 实际发生的地震的最大震级; $\Delta M = \overline{M}_y - M$, 即数值预测误差

对预测结果从 2 个方面检验, 一是判断趋势是否正确, 将下一年的趋势分为上升、持平、下降 3 种(以 $\Delta M = 0.5$ 为划分标准); 二是计算预测值与实际值的差值。从表 1 的统计结果来看, 外推第 1 年的报准率为 70%, 外推第 2 年为 45%, 外推第 3 年为 55%, 其震级差分别为 0.70、0.75 和 0.77, 符合外推预测的规律。对每一年而言, 有 3 个预测值, 取这 3 个值的均值 $\overline{M}_y$ 作为最后的预测值, 其误差可减小到 0.58。另外, 如果 $\overline{M}_y \geq 8.0$, 可能是超常的野点, 那么要视具体情况而定。如在 3 次预测的值都很大的情况下, 其预测值有可能偏离实际值; 如果第 1、2 次(未来 3 年、2 年)的预测值偏大, 而第 3 次(未来 1 年)预测值较低, 则以第 3 次的预测值为主(如表 1 中第 14 组数据)。用此标准进行修正剔除野点后, 其误差可降低到 0.33, 对震级序列而言, 这种误差是比较小的。当然这种标准只是来自有限次的检验统计结果, 是否具有普适性还有待于进一步的检验。

关于野点的出现, 笔者认为可能是采用多级拟合而产生的。对于已有的拟合数据, 通过实际值与拟合值残差的逐级拟合, 使其误差达到最小, 但对于预测值而言, 由于无实际值对照, 分级拟合的误差有可能叠加, 从而产生野点。

图1为全部数据的拟合曲线(内符检验).由图可见,该方法内符检验的效果较好.

图2给出了20次展望性检验的预测结果和原始值曲线的对比.从图2可以看出,在展望结果中,除了有4组与实际趋势不同外,其它各组的趋势预测正确.另外,剔除野点之后3次预测值的均值的预测效果更佳.

在计算过程中发现,该方法的拟合速度较快,一般通过5~6级的拟合,历史拟合率可达98%,也就是说所需自变量的数目较少.

从以上展望性检验的结果来看,该

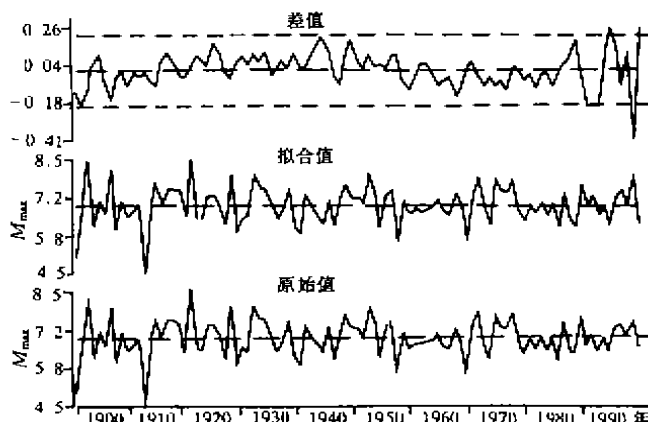


图1 内符检验的拟合曲线

Fig 1 Fitting curve of internal check.

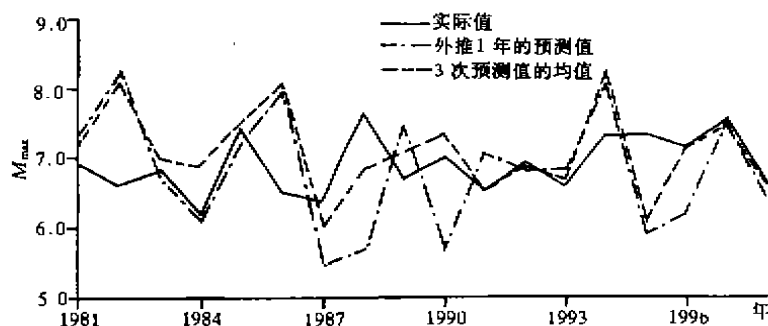


图2 展望性检验结果与原始曲线的对比

Fig 2 Comparison between extrapolated check results and original curve

方法内符效果相当好,外推预测也有一定的报准率,可用于地震的中期趋势预测.

2.3 对中国大陆地区未来1~3年地震活动趋势的预测

2.3.1 最大震级预测

取3组不同时间段的数据对未来1~3年(1999~2001年)的地震趋势进行预测,结果见表2和图3.

从预测结果来看,对1999年的3次预测值都比较相近,稳定性较好.因此,1999年中国大陆地区的地震活动最高水平为6.6级左右,与1998年持平.对2000年的前2次预测值明显偏高,可能属于野点,对其趋势的判定要等第3次的预测值,即在1999年结束以后.

表2 中国大陆地区1999~2001年年最大震级预测结果

资料时段	趋势预测			数值预测		
	1999年	2000年	2001年	1999年	2000年	2001年
1900~1996年	下			6.2		
1900~1997年	平	上		6.6	8.8	
1900~1998年	平	上	下	6.8	9.5	6.9
均值	平			6.6		

2.3.2 年释放总能量预测

考虑到1920年以前中国西部地区5级左右地震有可能漏记,对总释放能量有影响,故取1920以后的年释放总能量进行拟合.同样取3组不同时段的数据对未来1~3年的年释放总能量($\lg E$)进行预测,内符检验的最大误差为0.41.预测结果表明,1999年总释放能量与1998

年持平,2000年继续保持这种较低的活动水平,2001年有所回升。

2.4 最优周期的选定

取不同的数据长度,挑选出的优势周期大致相同。根据 F 值的大小排列,得到中国大陆地区地震活动的优势周期依次为 30、38 和 7 年。当然还有一些周期可以通过 F 检验,但与上述基本周期呈倍数关系,因此不再单独列出。

3 结语

本文所采用的方法在拟合速率方面有较大的提高,一般通过 5~6 级的拟合就可达到较高的精度(对内符而言)。对外推预测的精度也有所提高,报准率可达 70%,剔除一些异常点后,对震级的预测误差为 0.33。对大地震的趋势预测不稳定,有时与实际水平相差较大。但如果依照展望性检验所总结出的规律加以修正,可以提高预测结果的准确率。由于数学方法的局限性,对于一些超常的事件(原序列未包含的信息)很难作出正确的预测,这也从另一方面说明了地震活动的复杂性。

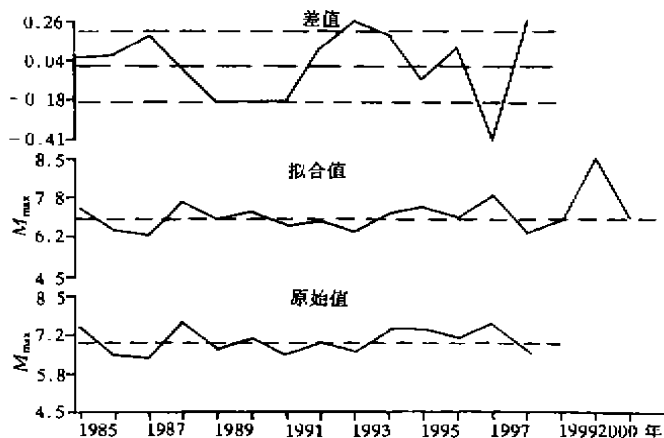


图3 中国大陆地区未来1~3年年最大震级预测结果

Fig.3 Prediction of maximum earthquake magnitude of Chinese continent in next three years.

[参考文献]

- [1] 魏凤英,曹鸿兴.长期预测的数学模型及其应用[M].北京:气象出版社,1990.
- [2] 成都地质学院《概率论与数理统计》编写组.概率论与数理统计[M].北京:地质出版社,1981.

APPLICATION OF MATHEMATICAL MODEL BASED ON MEAN GENERATION FUNCTION IN EARTHQUAKE MEDIUM-TERM PREDICTION

MA Wen-jing¹, ZHANG Xiao-qing², ZHANG Xiao-dong³

(1. Seismological Bureau of Qinghai Province, Xining 810001, China;

2. University of Science and Technology of China, Hefei 230026, China;

3. Center for Analysis and Prediction, CSB, Beijing 100036, China)

Abstract: The earthquake magnitude and energy sequence are fitted by using mathematical model based on mean generation function and prediction for medium-term earthquake trend of next three years are made. The method has following features: couple score criterion is applied for raising the fitting accuracy and rate; and balance between numerical prediction and trend prediction is considered; the multistage fitting of difference sequence for raising accuracy and speed of fitting is used. Besides, the optimum period of the time sequence is given. The earthquake trend of Chinese continent in next 1~3 years is forecasted by using the method.

Key words: Medium-term prediction; Trend prediction; Periodic analysis; Mathematical model; Mean generation function