

南北地震带年最高震级 序列的建模和预测

赵卫明

(宁夏地震局)

摘 要

本文应用 A.G.Ivakhnenko 提出的数据处理的分组方法的基本思想, 以生物有机体演化的方式, 给出了一种自动选择模型的最佳形式和参数, 构造高阶自回归模型的方法, 并用此方法对南北地震带年最高震级序列做了建模和预测检验, 效果较好。利用上述方法, 本文还对南北地震带未来 5 年最高震级作了预测。

关键词: 南北地震带; GMDH 方法; 震级预报

一、引言

对一个时间序列 $\{X_1, X_2, \dots\}$ 的预测, 需要首先建立一个模型

$$\hat{X}_{k+1} = f(X_k, X_{k-1}, \dots)$$

, 再用此模型做外推或预测。目前对地震时间序列的预测模型有 ARMA 模型^[1]和灰色模型^[2]等。但这些模型在具体应用时存在一定困难, 如常常需在建模前估计序列的一些特征, 这在数据序列较短时是难以做到的; 并且一般模型的具体形式和参数也是固定的, 这些都不利于对多种形式的时间序列建模和预测。

A.G.Ivakhnenko 于 1966 年提出的“数据处理的分组方法”(简称 GMDH 方法)^[3], 以生物有机体演化的方式构造高阶自回归模型。使用 GMDH 方法在建模前并不需要了解序列的一些特征, 在建模时也不必事先设置模型的形式和参数, 由每一代的计算结果来自动判断, 最终获得的模型为一高阶多项式, 这与传统的建模方法是有所不同的。该方法比较适合对变化较大的序列建模和预测。沈心焯等^[4]用此方法对我国大陆年最高震级序列做了建模尝试。本文应用 GMDH 方法建模的基本思想, 对南北地震带年最高震级序列做了建模试验和分析, 得到了一种比较好的建模方法。用该方法对南北地震带年最高震级序列进行了建模和预测尝试。

二、方法概述

1. 设有时间序列

$$X_1, X_2, \dots, X_{*T/m}, X_{m+1}, \dots, X_{m+n}, \quad (1)$$

首先将序列标准化。设序列的均值和标准差分别为 $\bar{X}$ 和 S , 标准化后的序列为 $\{X'_i\}$, 有

$$X'_i = (X_i - X) / S, \quad i = 1, 2, \dots, m+n. \quad (2)$$

为了方便, 以下标准化后的序列仍用 $\{X_i\}$ 表示。

2. 设序列 (2) 的第 i 个输出值为 i 之前 p 和 q 两个输入值的二次函数,

$$\hat{X}_i = a + bX_{i-p} + cX_{i-q} + dX_{i-p}^2 + eX_{i-q}^2 + fX_{i-p}X_{i-q}, \quad (3)$$

$$i = m+1, \dots, m+n; p = 1, 2, \dots, m, q = 1, 2, \dots, m, p \neq q.$$

使它们对序列中的后 n 个数据做最佳拟合。对于给定的 m , 共有 $m(m-1)/2$ 个这样的回归方程组, 每个回归方程组有 n 个方程。要使每个回归方程组有解, 必须取 $n \geq 6$ 。

求出每个回归方程组的解和剩余残差平方和 r_j , 并计算出后 n 个数据点的多项式值 $Y_{m+1, j}$ ($i = 1, 2, \dots, n, j = 1, 2, \dots, m(m-1)/2$)。由原序列 (1) 的前 m 个数据和所有计算出的后 n 个数据点组成 $m(m-1)/2$ 个新一代的序列 $\{Z_{ij}\}$ 。

3. 将新一代的序列看作是比上一代序列有更好的预报能力的变量。由于每一代序列数和上一代序列数相比是以 $m(m-1)/2$ 倍增加的, 并且一般残差平方和大的序列生成的下一代序列的残差平方和也大, 故为了减少计算量仅保留残差平方和较小的那些序列。经实际验算我们采取了两种筛选方法。当给定在每一代所需保留的序列个数 w 后, 在第一代时保留 w 个残差平方和较小的序列, 其后各代则先从上一代每个序列生成的 $m(m-1)/2$ 个序列中挑出 $w/3$ 个残差平方和较小的序列, 然后再在已初步挑选出的 $w^2/3$ 个序列中挑选出 w 个残差平方和较小的序列。这样做既可减少计算量, 又尽可能使保留下来的 w 个序列来自于上一代的不同序列, 以避免各代“单传”情况的出现。

4. 求出每一代保留下来的 w 个序列的残差平方和的最小值 R_{mn} 。如果 R_{mn} 的值比上一代的 R_{mn} 值小, 则用保留下来的 w 个序列分别代入 (3) 式的右端, 用原序列 (1) 式代入 (3) 式的左端, 从步骤 2 开始重复下一代的计算; 若 R_{mn} 值比上一代的大, 则认为以后各代都不能再提高预报能力, 终止下一代的计算, 并利用上面各代的回归系数和原序列做外推预测。

若迭代至第 k 代 R_{mn} 值达最小, 则用计算第 k 代的残差平方和最小的那个序列的二次函数循原路径往回代, 就可得到一个最高阶次为 $2k$ 的非线性回归模型 (在实际计算时不必求出此模型), 我们记此模型为 GMDH (m, n, k)。

需要说明的是, 步骤 1 并非必需, 但可以大大提高模型的稳定性, 所以最好不要省去。在求解回归方程组时, 最好用求解精度比较高的直接消去法、奇异值分解等^[5]方法, 这样不仅可以减少迭代次数, 也可以提高模型的稳定性。

在实际计算时为了节省计算机的存储空间和提高计算速度, 我们在每一代求解出了回归方程组和解的残差平方和后, 并不马上生成每个新序列, 先按步骤 3 挑出残差平方和较小的 w 组回归系数, 再用这些回归系数生成 w 个新序列, 并在每个新序列的后面补上 1 个外推值。当找到 R_{mn} 值达最小的那一代后, 其残差平方和最小的序列的后 1 个值就为最终求得的预测值。

三、南北地震带年最高震级的建模预测

南北地震带位于青藏板块的东部边缘, 为我国大陆东西两部分在地貌、地质构造、地球物理场及地震活动特征等方面的分界线。由于南北地震带主要受隆起的青藏高原的挤压

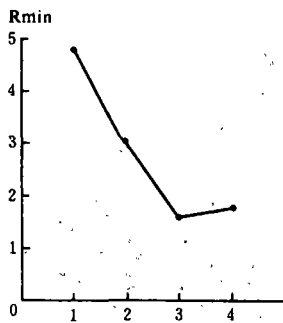


图1 GMDH (5, 16, 3)
模型的各代 R_{min} 值曲线

Fig.1 The R_{mn} value in the Strata
of the GMDH (5, 16, 3) model

和华北板块的阻挡, 沿该带中强地震频繁发生, 且有明显的时空分布特征。本文选取 1965 年以来, 东经 95° – 107° , 北纬 22° – 40° 范围内的年最高震级时间序列, 用上述方法做了 GMDH 建模和预测尝试。

首先用 1965–1985 年南北地震带年最高震级序列做了建模和预测试验, 得到了 GMDH (5, 16, 3) 模型。图 1 为各代的 R_{mn} 值曲线, 图 2 中给出了模型的拟合值、预测值和原序列值。可以看出, 虽然最终获得的模型只有 3 代, 但 R_{mn} 值却由第 1 代的 4.81 减小到了 1.58, 表明最小残差平方和的减小是比较显著的。模型的拟合值和实测值之差都在 0.5 级以内。用此模型对南北地震带 1986 年到 1990 年年最高震级的预测结果见表 1。实测值和预测值之差除 1988 年达 0.7 级外, 其余都在 0.5 级以内。由于南北地震带 1977 年到 1987 年的年最高震级都小于 7.0 级, 其中 1987 年的年最高震级仅为 5.8 级, 故 1988 年澜沧 7.6 级地震在此序列中为一变化大的序列点。1988 年的预测值虽然和实测值相差大一些, 但仍然达 6.9 级。

表 1 南北地震带 1986 年到 1990 年年最高震级预测结果

年份	预测震级	实际发生最大地震				
		时间	经度	纬度	震级	参考地名
1986	6.5	8 月 26 日	$37^{\circ} 42'$	$101^{\circ} 34'$	6.4	青海门源
1987	6.0	1 月 8 日	$34^{\circ} 10'$	$103^{\circ} 14'$	5.8	甘肃迭部
1988	6.9	11 月 6 日	$22^{\circ} 50'$	$99^{\circ} 43'$	7.6	云南澜沧
1989	6.7	4 月 16 日	$29^{\circ} 56'$	$99^{\circ} 15'$	6.7	四川巴塘
1990	6.4	4 月 26 日	$36^{\circ} 07'$	$100^{\circ} 08'$	6.9	青海共和

我们分别用 1969–1991 年和 1967–1991 年两个不同时段的数据建模, 对南北地震带 1992 年到 1996 年的年最高震级进行了预测, 结果见图 3 至图 6 及表 2。两个模型的预测

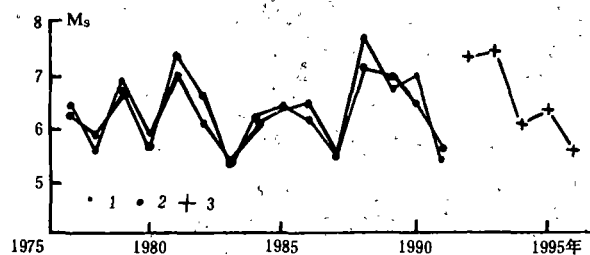


图2 GMDH (5, 16, 3) 模型

1. 实测值; 2. 拟合值; 3. 预测值

Fig.2 The GMDH (5, 16, 3) model

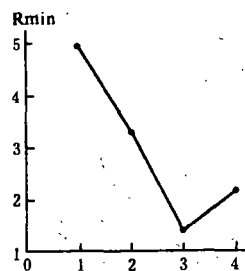


图3 GMDH (8, 15, 3)

模型的各代 R_{min} 值曲线Fig.3 The R_{min} value in the strata

of GMDH (8, 15, 3) model

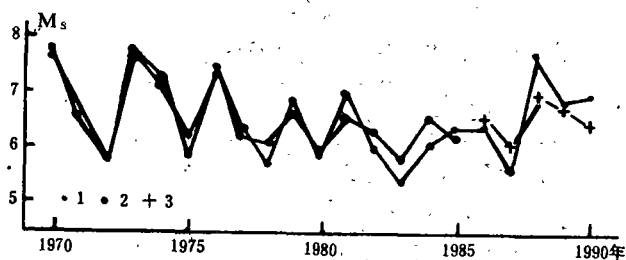


图4 GMDH (8, 15, 3) 模型

1. 实测值; 2. 拟合值; 3. 预测值

Fig.4 The GMDH (8, 15, 3) model

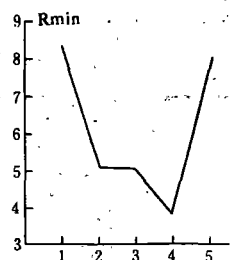


图5 GMDH (7, 18, 4)

模型的各代 R_{min} 值曲线Fig.5 The R_{min} value in the strata

of GMDH (7, 18, 4) model

表2 南北地震带 1992 年至 1996 年年最高震级预测结果

年 份	1992	1993	1994	1995	1996
GMDH (8, 15, 3) 模型	7.2	7.3	6.0	6.3	5.5
GMDH (7, 18, 4) 模型	7.3	6.8	6.1	6.0	5.8

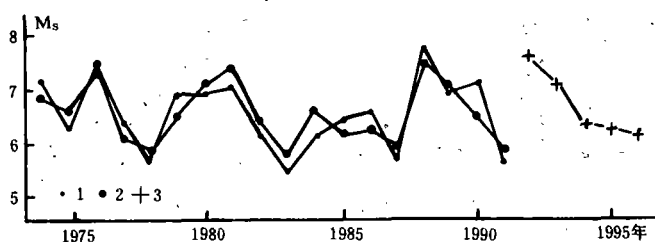


图6 GMDH (7, 18, 4) 模型

1. 实测值; 2. 拟合值; 3. 预测值

Fig.6 The GMDH (7, 18, 4) model

结果是比较一致的, 都表明 1992 和 1993 年度将会发生 7.0 级左右地震, 而 1994 到 1996 年度的年最高震级将会降至 6.0 级左右。1992 年 4 月 23 日在中国和缅甸交界附近发生的 $M_s 6.9$ 地震, 在一定程度上验证了上述预测结果的可靠性, 其余预测结果还有待实践检验。

(本文 1992 年 8 月 26 日收到)

参考文献

- [1] 黎令仪、刘德富, 震级序列的门限建模, 地球物理学报, Vol.28, No.3, 1985.
- [2] 黄雪范, 华北强震的灰色预测, 中国地震, Vol.3, No.2, 1987.
- [3] Stanley J. Farlow, Self-Organizing Method in Modeling, Dekker, 1984.
- [4] 沈心焯、朱照宣等, 震级序列的 GMDH 建模, 地球物理学报, Vol.35, No.1, 1992.
- [5] 吴明熙、王鸣等, 1985 年禄劝地震部份余震的精确定位, 地震学报, Vol.12, No.2, 1990.

ESTABLISHING MODEL AND PREDICTION ON THE HIGHEST YEARLY MAGNITUDE SEQUENCE OF THE NORTH-SOUTH SEISMIC BELT

Zhao Weiming

(Seismological Bureau of Ningxia, Yinchuan, China)

Abstract

In this paper, by using the basic idea of the method of dividing into groups in data handling proposed by A.G. Ivakhnenko, according to the way of creature-organism evolution, we have given a method which can select the best form and parameter of the model automatically and construct the higher order autoregression model. Using this method, a trial of establishing model and prediction on the highest yearly magnitude sequence of the North-South Seismic Belt is taken. And the highest magnitude of the North-South Seismic Belt is predicted in the future 5 years.

Key Words: North-South Seismic Belt; GMDH Method; Magnitude Prediction