

37-43

新疆主要地震区 pp 回归综合预报模型研究

赵翠萍, 周仕勇, 朱令人

(新疆维吾尔自治区地震局, 新疆 乌鲁木齐 830011)

p315.7

摘要: 应用基于 pp 回归理论的数值型地震综合预报软件系统, 选择了 b 值、 C_b 值、 A_b 值等 15 个测震学参量, 建立了它们与响应变量(未来任意时段内最大地震震级)的 pp 回归模型. 对新疆 4 个主要地震区分别给出了各区的建模岭函数及综合预报模型. 回顾检验及实际预报结果均表明该模型的预报效果是较为理想的.

主题词: 新疆; 地震综合预报; pp 回归模型; 岭函数

中图分类号: P315.7; P315.8 **文献标识码:** A **文章编号:** 1000-0844(1999)01-0037-07

0 引言

在实际地震预报过程中, 对于各单项异常的判定带有很强的人为性. 同时, 各单项手段预报因子之间的相关问题, 预报因子非正态、非线性等问题使一些综合预报方法存在难以克服的困难. 朱令人等^[1]应用投影寻踪回归(简称 pp 回归)理论提出了一种数值型综合预报方法. 该方法利用计算机对数据进行投影寻优降维分析, 客观地审视数据结构, 充分获取有用信息, 并以数值函数描述其结构后再用于预测, 故它可以自动排除无关因子的干扰, 解决了众多自变量之间的相关问题. 应用该方法不需做人为的异常识别, 故可以排除单项异常判别带来的人为干扰. 同时由于 pp 回归是以数据的线性投影为基础, 通过寻找多个有意义方向上岭函数的叠加, 找出线性投影中的非线性结构, 最终实现非线性回归, 因而所建立的自变量与预测变量之间的关系更接近客观实际. 自 1994 年以来在新疆地区地震的预报实践中该方法显示出较强的实用性, 而且利用该方法可以进行时序预报. 为了对该项综合预报方法进行推广和应用, 本文选择了新疆 4 个主要地震区利用该方法建立综合预报模型, 并在实际预报中进行了检验.

1 原理简述及软件构成

1.1 pp 回归原理

设 (X, Y) 是一对随机变量, 其中 X 是 p 维的自变量, Y 是 q 维的响应变量(预测量), 且令

$$F(x) = E(Y/X) \quad (1)$$

回归问题就是要用 (X, Y) 的大量观测样本 $(X_1, Y_1), \dots, (X_n, Y_n)$ 来估计回归函数 $F(x)$.

pp 回归采用一系列的岭函数的和来逼近回归函数, 即令

收稿日期: 1997-07-17

第一作者简介: 赵翠萍, 女, 1967 年 1 月生, 助理研究员, 主要从事地震分析预报工作.

$$F(x) = \sum_{m=1}^M G_m(A_m^T \cdot x) \quad (2)$$

式中: $G_m(A_m^T \cdot x)$ 表示第 m 个岭函数; A_m 表示第 m 个岭函数 $p \times 1$ 的投影矩阵, 可以用增大 M 的办法来减少模型误差.

可以看出, 当 $M = 1$ 时, $G_1(t) = ct$, 式(2)呈线性回归. 因此 pp 回归是线性回归更一般的形式.

式(1)的分立式可写为如下形式:

$$F_i(x) = E[Y_i/X_1, X_2, \dots, X_p] = Y_i + \sum_{m=1}^m \beta_{im} \cdot G_m\left(\sum_{j=1}^p A_{jm} \cdot x_j\right) \quad (3)$$

式中: $Y_i = EY_i$; $EG_m = 0$; $EG_m^2 = 1$; $\sum_{j=1}^p A_{jm}^2 = 1$; A_{jm} 表示第 m 个岭函数的第 j 个自变量线性系数向量; β_{im} 为第 m 项岭函数对第 i 个响应变量的影响系数; G_m 是第 m 项岭函数. 对这些岭函数和系数的估计有多种方法, 本文采用最小二乘法, 以便使 pp 回归损失最小. 即令

$$L_i = \sum_{i=1}^q W_i \cdot E\left[Y_i - \bar{Y}_i - \sum_{m=1}^m \beta_{im} \cdot G_m(A_m^T \cdot x)\right]^2 = \min \quad (4)$$

式中: q 为响应变量的个数; W_i 为第 i 个响应变量的权重.

在本项工作中采用多重平滑回归技术对回归模型的参数进行估计, 即根据观测资料估计各项岭函数及系数, 进而估计回归函数 $F(x)$. 具体过程见文献[1], 本文不再赘述.

1.2 建模参量及软件构成

1.2.1 建模参量的选取

我们初步选定了 15 个自变量作为建模参量(还可以根据需要增减建模参量的计算模块), 它们是: 描述地震活动空间分布状态的特征量 C_1 , C_b 和 H 以及描述地震活动时间分布状态的特征量 B , AM 等. 其中 C_1 为地震空间集中度^[2]; C_b 为地震带状集中度^[2]; H 为地震活动图象信息熵^[4]; B 为某时段的 b 值; AM 为某时段平均震级; MM 为某时段的最大震级; I_q 为地震时间分布缺信量^[5]; M_1 为所选时间窗内后三分之一时段的最大震级; N_2 为 2.0 级以上地震的月频次; A_b 为地震活动度; D 为地震危险度^[6]; TN_2 为本时段 2.0 级以上地震频次与上时段 2.0 级以上地震频次之差; T_b 为某时段 b 值与上时段 b 值之差; TA_b 为某时段 A_b 值与上时段 A_b 值之差; TD 为某时段 D 值与上时段 D 值之差.

选取响应变量时, 首先选择一个研究区, 预测在研究区内未来某时段中将发生的 strongest 地震的震级, 即 $Y = M$, 并且给出该时段分别发生 5~7 级以上地震的概率.

显然以上所选的 15 个参量与响应变量的关系有的较为密切, 有的或许关系不大甚至还可能完全无关. 为了减少建立 pp 回归模型的计算量, 首先要对所选的 15 个自变量进行优选. 应用投影原理, 将初选的 15 个参量分别投影到每一个特征自变量与响应变量(预测量)构成的低维空间上. 观察投影图, 如果投影图上的数据点构成一条与坐标轴平行的直线, 说明该自变量与预测量无关, 可剔除; 若投影图上的数据点分散, 但呈一定形状分布, 则说明该特征量与预测量有关, 应选用. 据此, 对初选的 15 个自变量进行判别, 择优选出其中一部分自变量作为最终建模用的预报参量, 并根据投影分析各预报参量与预测量的相关关系, 计算出各预报参量的建模权重. 将上述程序编制成软件, 图 1 为软件系统的构成框图.

由于各单项自变量因子与地震的相关性往往因分区不同而有差异, 因而所选区域不同, 自变量的最终优选结果也不尽相同.

1.2.2 模型建立与检验

对优选后的建模参量进行平滑处理,然后建立 pp 回归模型.预留一段原始资料对模型进行检验.如果检验效果不好,则必须重新选择特征自变量,重新调整建模控制参数,直到有满意的效果为止.检验时约定,预报震级与实际震级的绝对误差不超过 0.5 级或相对误差不大于 10%,则为预测结果合格.

通过预留检验确认所建模型后,即可应用于实际预报.从而可得到相应的预测震级.

1.3 强震发生的概率预测

根据已有的观测资料,给出研究区未来某段时间范围内所发生最大地震的可能震级是本系统的主要功能.但由于在多数情况下人们所关注的是未来时段内某个震级范围内的强震发生的概率,故在本系统中还设计了对 $M_s \geq 5.0, 6.0, 7.0$ 地震发生概率的预测功能.其设计原理是:

设有 n 个观测时段,其中第 i 个观测时段研究区内实际发生的最大地震的震级为 M_{ai} , 根据软件建立的模型所得到的第 i 个观测时段内的最大地震的震级预测值为 M_{pi} .

$$\Delta M_i = M_{pi} - M_{ai} \quad (i = 1, \dots, n) \quad (6)$$

式中: ΔM_i 显然为预测偏离,由于预测模型是一种统计模型,预测偏离可以被看成是随机变量,且当 n 值较大时, ΔM_i 的变化应服从正态分布,因此其 μ 及 σ 值可按式估计:

$$\mu = E(\Delta M_i) = \sum_{i=1}^n \Delta M_i / n \quad (7)$$

$$\sigma^2 = \sum_{i=1}^n (\Delta M_i - \mu) / (n - 1) \quad (8)$$

在应用模型作实际预报时,如果得到某研究区未来发生最大地震震级的预测值为 M_p , 这时该研究区未来时段中将发生 5 级、6 级和 7 级地震的概率如下:

$$P(M_s \geq 5.0) = P[\Delta M \leq (M_p - 5.0)] = \int_{-\infty}^{(M_p - 5.0)} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \quad (9)$$

$$P(M_s \geq 6.0) = P[\Delta M \leq (M_p - 6.0)] = \int_{-\infty}^{(M_p - 6.0)} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \quad (10)$$

$$P(M_s \geq 7.0) = P[\Delta M \leq (M_p - 7.0)] = \int_{-\infty}^{(M_p - 7.0)} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \quad (11)$$

2 新疆部分地区建模研究

众所周知,新疆是中强地震频发的地区,因而也是检验预报方法好坏的研究实验区.在工作中,我们重点对 4 个地区进行了建模研究,各区范围见表 1.所用地震资料取自新疆维吾尔自治区地震局所建地震目录库.

特征自变量的计算采用的时间窗长度为 12 个月,滑动步长为 1 个月.响应变量 M 为未来 3 个月内发生在研究区内的最大地震的震级.以上参数还可以根据需要加以调整,如调整窗

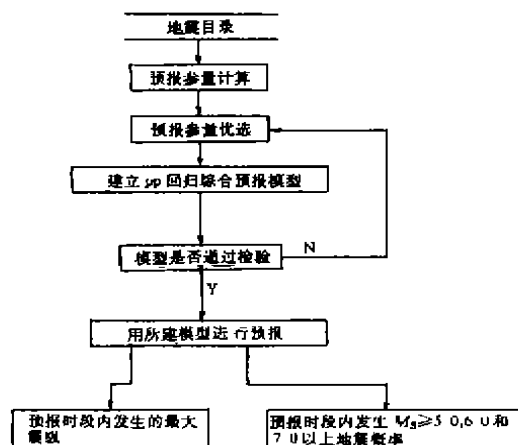


图 1 软件系统的构成框图

Fig.1 The sketch map of the projection pursuit regression system.

长、滑动步长以及根据预报的需要调整响应变量以达到最优建模的目的。

2.1 乌鲁木齐地区 pp 回归综合建模研究

利用 1985 年 1 月~1996 年 2 月共 12 年的资料进行建模. 经过对初选 15 个自变量的优选后, 确定 12 个建模参量, 见表 2.

利用以上 12 个自变量与 1 个响应变量 M , 1985 年 12 月~1994 年 6 月共 106 组数据建成该区 pp 回归综合地震预报模型(图 2). 用 1994 年~1996 年 5 月 23 组数据进行检验, 结果为: 预报震级与实际震级的绝对误差 < 0.5 级的占 70%, 相对误差 $\leq 10\%$ 的占 61%, 符合约定检验标准, 所建模型由相对权重分别为 1.00 和 0.838 的 2 个岭函数构成, 其数值表达式为:

$$M_{\text{响}} = 1.00G_1(A_1^T \cdot X) + 0.838G_2(A_2^T \cdot X)$$

式中: $X = [X_1, X_2, \dots, X_{12}]$.

$A_1 = [0.095\ 2, -0.085, -0.176\ 8, 0.407\ 7, 0.138\ 1, -0.005\ 7, -0.002\ 4, 0.065\ 9, -0.101\ 8, -0.184\ 7, -0.001]$

$A_2 = [-0.435\ 6, 0.151\ 3, 0.451\ 0, 0.263\ 4, 0.056\ 4, 0.012\ 1, 0.009\ 9, 0.119\ 5, -0.300\ 8, 0.633\ 5, 0.001]$

表 2 乌鲁木齐地区 pp 回归综合建模参量

自变量	特征量	相对权重
X_1	N_2	1.00
X_2	IN_2	0.56
X_3	C_b	0.50
X_4	M_1	0.18
X_5	B	0.17
X_6	C_1	0.13
X_7	D	0.13
X_8	I_q	0.10
X_9	MM	0.09
X_{10}	A_b	0.06
X_{11}	H	0.04
X_{12}	AM	0.01

研究区	区域范围	起算震级(M_1)
乌鲁木齐区	42°30'~45°N, 87°~89°E	2.0
北天山区	42°30'~45°N, 81°~87°E	2.0
南天山	41°~42°30'N, 81°~87°E	2.0
柯坪区	39°~42°30'N, 76°~81°E	2.0

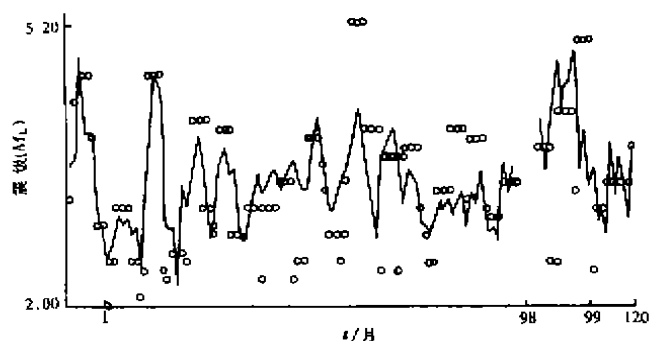


图 2 乌鲁木齐地区 pp 回归综合预报模型(图中的圈代表预报时段内实际发生的最大地震的震级, 在本文中为 3 个月内的最大震级, 下同)

Fig. 2 The comprehensive forecasting model by projection pursuit regression method of Urumqi area.

2.2 北天山地区 pp 回归综合建模研究

采用 1984 年 6 月~1995 年 12 月共 12 年的资料进行建模研究. 经对初选的 15 个特征自变量进行优选后, 最终确立了 11 个建模参量, 见表 3.

利用以上 11 个自变量与 1 个响应变量 M , 1985 年 6 月~1994 年 2 月共 105 组数据建立了该区 pp 回归综合预报模型(图 3). 用 1994 年 3 月~1995 年 10 月的 20 组数据进行检验, 结果为: 预报震级与实际震级的绝对误差 < 0.5 级的占 65%, 相对误差 $\leq 10\%$ 的占 50%. 所建模型由相对权重分别为 1.00 和 0.974 的 2 个最优岭函数构成, 其数值函数表达式为:

$$M_{\text{响}} = 1.000G_1(A_1^T \cdot X) + 0.974\ 26G_2(A_2^T \cdot X)$$

式中: $X = [X_1, X_2, \dots, X_{11}]$.

$A_1 = [0.301\ 5, 0.168\ 7, -0.736\ 1, -0.252\ 3, -0.496\ 2, -0.003\ 0, -0.003\ 3, -0.067\ 4, 0.089\ 4, -0.053, -0.1]$

$A_2 = [0.231\ 3, -0.082, 0.358\ 7, 0.018\ 4, 0.575\ 2, 0.002\ 8, -0.002\ 9, -0.542, 0.021\ 8, -0.684\ 4, -0.1]$

根据模型预测 1995 年 11 月~1996 年 1 月和 1995 年 12 月~1996 年 2 月 2 个时段内该区发生的最大地震的震级分别为 4.58 和 4.59, 2 时段内发生 5 级以上地震的概率均为 58%, 发生 6 级以上地震的概率为 19%. 实际情况是 1996 年 1 月 9 日发生沙湾南 5.2 级地震, 预测效果较好.

表 3 北天山地区 pp 回归建模参量

自变量	特征量	相对权重
1	B	1.000
2	N_2	0.827
3	A_b	0.727
4	MM	0.726
5	C_1	0.258
6	M_1	0.220
7	TN_2	0.214
8	H	0.193
9	C_6	0.187
10	I_q	0.119
11	A_m	0.049

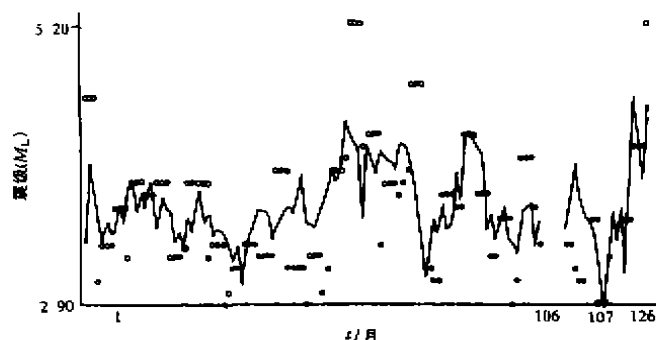


图 3 北天山地区 pp 回归综合预报模型

Fig. 3 The comprehensive forecasting model by projection pursuit regression method of the west segment in the northern Tianshan seismic belt.

2.3 南天山东段 pp 回归综合建模研究

采用 1985 年 1 月~1995 年 12 月共 12 年的资料在该区进行建模研究. 经对初选的 15 个特征自变量进行优选后, 最终确立了 12 个特征自变量作为建模参量, 见表 4.

利用以上 12 个特征量与一个响应变量 M , 1985 年 12 月~1994 年 2 月共 99 组数据建成该区 pp 回归综合地震预报模型(图 4), 用 1994 年 3 月~1995 年 12 月共 20 个数据进行检验, 结果为: 预报震级与实际震级的绝对误差 < 0.5 级的占 70%, 相对误差 $\leq 10\%$ 的占 60%, 符合

表 4 南天山东段 pp 回归建模参量

自变量	特征量	相对权重
1	MI	1.00
2	CI	0.87
3	A_b	0.69
4	TN_2	0.64
5	I_q	0.59
6	C_6	0.57
7	MM	0.43
8	N_2	0.41
9	b	0.38
10	H	0.29
11	D	0.22
12	AM	0.04

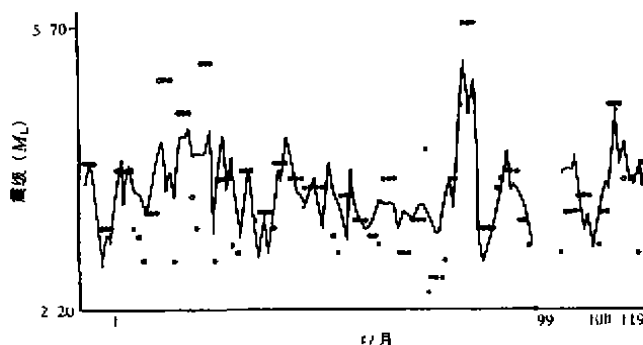


图 4 南天山东段 pp 回归综合预报模型

Fig. 4 The comprehensive forecasting model by projection pursuit regression method of the east segment in the southern Tianshan seismic belt.

约定检验标准. 所建模型由相对权重分别为 1.00, 0.73 和 0.599 的 3 个岭函数构成, 其数值函数表达式为:

$$M_{\text{响}} = 1.00G_1(A_1^T \cdot X) + 0.59859G_2(A_2^T \cdot X) + 0.73G_3(A_3^T \cdot X)$$

式中: $X = [X_1, X_2, \dots, X_{12}]$.

$A_1 = [0.1738, -0.0349, -0.8995, 0.1816, 0.1719, 0.0020, -0.001, 0.0033, 0.1176, 0.2088, 0.1]$

$A_2 = [0.1967, -0.0402, 0.0095, -0.6017, 0.1352, -0.0038, 0.0036, 0.056, -0.3755, 0.3642, 0.01]$

$A_3 = [-0.2849, 0.036, 0.0801, 0.0738, -0.4303, 0.0022, -0.0028, -0.0437, 0.4617, -0.3591, -0.1]$

2.4 柯坪区 pp 回归综合建模研究

采用 1985 年 1 月~1996 年 2 月共 11 年的资料进行该区建模研究. 经过对 15 个特征自变量进行优选后, 最终确立 10 个特征量为该区建模参量, 见表 5.

由以上 10 个特征量与 1 个响应变量 M , 1985 年 12~1994 年 4 月共 101 组数据建成该区 pp 回归综合预报模型(图 5), 用 1994 年 5 月~1995 年 12 月共 18 组数据进行检验, 结果为: 预报震级与实际震级的绝对误差 ≤ 0.5 级以及相对误差 $\leq 10\%$ 的比例均高达 80%, 表明所建模型的检验效果很好. 根据模型对 1996 年 1 月~1996 年 3 月和 1996 年 2 月~1996 年 4 月 2 时段所发生的地震的概率进行预测, 结果为: 在以上时段中该区发生 5 级以上地震的概率为 92%, 发生 6 级以上地震的概率为 82%. 实际情况是于 1996 年 3 月 19 日在该区发生伽师 6.9 级地震, 可见预报是较为准确的. 所建模型由相对权重分别为 1.0000, 0.97426 的 2 个岭函数构成, 其数值表达式为:

$$M_{\text{响}} = 1.00G_1(A_1^T \cdot X) + 0.9806G_2(A_2^T \cdot X)$$

式中: $X = [X_1, X_2, \dots, X_{10}]$.

$A_1 = [0.6488, -0.0058, -0.1717, 0.6197, 0.3233, 0.0010, 0.0555, -0.2385, 0.0286, -0.0121]$

$A_2 = [0.6489, -0.0116, 0.0200, -0.3867, -0.4360, -0.0006, -0.1958, 0.2794, 0.2731, 0.2184]$

表 5 柯坪区 pp 回归建模参量

自变量	特征量	相对权重
1	TN_2	1.000
2	B	0.36
3	C_1	0.29
4	C_6	0.26
5	H	0.23
6	N_2	0.11
7	I_9	0.07
8	MM	0.07
9	M_1	0.06
10	A_M	0.02

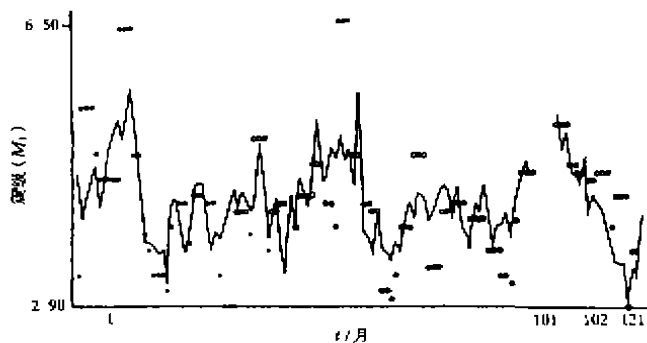


图 5 柯坪区 pp 回归综合预报模型

Fig. 5 The comprehensive forecasting model by projection pursuit regression method of Keping area.

3 结语

本文运用 pp 回归数值型综合地震预报系统,对新疆 4 个地区建立了未来 3 个月地震短趋势预报 pp 回归综合预报模型.所建模型检验合格率均达 60% 以上,最好的是柯坪地区,达 80%.对近期新疆地区中强地震的试验性预报效果较好.针对实际预报的需要还给出了每一预报时段分别发生 5, 6 和 7 级以上地震的概率.在具体应用时,预报时段可以根据需要加以调整,如建立 1 个月、6 个月或者 1 年时间尺度的预报模型,对各分区可以进行时序性预报.

pp 回归综合预报模型基于其独到的数据处理优势,具有广阔的开拓前景,如预报参量还可包括各种前兆,响应变量可以考虑时间和地点等,这些都是值得进一步研究的.

本文的发表得到新疆维吾尔自治区地震局科技监测处的支持,在此深表感谢.

参考文献

- [1] 朱令人,周仕勇,邓传玲.地震综合预报的新方法——投影寻踪回归[J].地震学报,1994,16(增刊):1~9.
- [2] 顾方琦.地震空间集中度的另一种表示方法及其在地震预报中的应用[J].地震,1985,(4):1~9.
- [3] 周仕勇,王海涛.地震空间分布的带状集中度 C_b 值[J].地震,1991,(2):59~64.
- [4] 朱令人,周仕勇.信息量在地震活动研究中的应用[J].西北地震学报,1991,13(3):15~20.
- [5] 王伟.地震时间间隔的统计分布及其地震危险度 D 值在华北地震前的异常变化[J].地震学报,1987,9(2):125~135.

STUDY ON THE SEISMOMETRY COMPREHENSIVE FORECASTING MODEL ESTABLISHED BY USING SEISMOMETRY INDICES AND PROJECTION PURSUIT REGRESSION METHOD OF MAIN SEISMIC REGIONS IN XINJIANG

ZHAO Cui-ping, ZHOU Shi-yong, ZHU Ling-ren

(Seismological Bureau of Xinjiang Uygur Autonomous Region, Urumqi Xinjiang 830011)

Abstract

Based on the projection pursuit regression (PPR) principle, the PPR model has been set up with 15 independent variables such as b value, C_b value, A_b value etc., which are selected from seismometry parameters and one dependent variable representing maximum magnitude of earthquake that would occur in coming several months. By using the model, maximum magnitude of coming earthquake and probability of earthquake to occur in the future can be predicted. The PPR model of four mainly seismic regions in Xinjiang and their ridge functions are showed. The prediction effects of reservation test and practical test are satisfactory.

Key words: Xinjiang; Multi-disciplinary earthquake prediction; Projection pursuit regression model; Ridge function