

用多元混合门限回归进行汛期 降水量预测试验

孙 贞^{1,2} 徐晓亮²

(1. 南京大学大气科学系, 210093; 2. 青岛市气象局)

提 要

在综合考虑预测对象的周期变化和前期外部因子的共同作用后, 给出带有周期分量的多元混合门限回归模型。通过青岛汛期降雨量的7年预报试验表明, 该模型具有较为稳定的预报能力, 值得进一步研究应用。

关键词: 混合门限回归 周期分量 预测试验

引 言

在非线性模型中, 门限自回归是一种常用的非线性统计预测方法, 1994年严华生等将门限自回归中的基本思想运用到多元分析中, 提出一种多元门限回归模型^[1]。文献[2]中指出将多元门限回归和门限自回归结合起来, 可以构造出新的多输入单输出混合门限回归模型。本文根据文献[2]中混合门限回归建模思想, 提出带有周期分量的多元混合门限回归模型, 并用青岛汛期6~8月的汛期降水量进行了对比试验研究。

1 预测模型和基本建模步骤

气候资料的时间序列反映了曾经发生过的所有因果关系和结构关系的影响, 而且大气运动的多频振动的现象在气候资料的时间序列中也得到充分的体现^[3, 4]。因此气候资料的时间序列中蕴涵不同时间尺度的特征, 往往表现为气候系统的周期性变化。这种周期性反映了时间序列自身的历史演变规律, 影响着预测对象的变化。因此我们设想将混合门限回归模型中自回归部分用预测对象的若干显著周期分量来取代, 与影响预测对象的预报因子共同组成带有周期分量的多元混合门限回归模型。该模型反映了影响预测对象的外部因素和其自身变化的综合作用。

设有预测对象 $y(k) = \{y(1), y(2), \dots, y(N)\}$, N 为样本长度, 建立如下预测模型:

$$\hat{y}(t) = a_0^{(j)}(t) + \sum_{i=1}^{k_j} a_i^{(j)}(t) f_i^{(j)}(t) + \sum_{i=1}^{p_j} b_i^{(j)}(t) x_i^{(j)}(t)$$
$$w_{t-d} \in R_j \quad j = 1, 2, \dots, l$$

其中: $w_{t-d} = \{x_{t-d}, f_{t-d}\}$ 。

式中 $\hat{y}(t)$ 为预测对象, $f(t)$ 为预测对象的显著周期分量, $x(t)$ 为外部预报因子, $a_i^{(j)}, b_i^{(j)}$ 为第 j 个区间的回归系数, k_j 是显著周期分量个数, p_j 是外部因子个数, l 为门限区间个数, d 为门限时滞, $w_{t-d} = \{x_{t-d}, f_{t-d}\}$ 表示 $x_{(t)}$ 或 $f_{(t)}$ 门限变量, R_j 为第 j 个区间的门限值。式中的第二项为周期项, 它反映了时间序列自身周期性的历史演变规律, 第三项为外部影响因子项, 体现了外部因子对预测对象的作用。

对于上述预测模型应用文献[2]提出的最优分割法对 $x(t)$ 和 $f(t)$ 组成的样本序列寻找门限变量和门限值。在确定门限变量和门限值后, 将门限变量按小于、大于和等于门限值分为两段, 其余样本序列以及预测对象也相应地分段, 然后逐段作逐步回归, 建立分段回归预测方程。当样本数足够多时也可对 $x(t)$ 和 $f(t)$ 分别确定各自的门限变量 x_k, f_k 和门限值 r_x, r_f , 然后按下列形式进行组合, 确定样本分组条件。根据组合条件将样本以及预测对象相应地分段后分别逐段作逐

步回归,建立分段回归预测方程即完成建模过程。

$$\begin{cases} r_x < x_k, r_f < f_k \\ r_x \geq x_k, r_f < f_k \\ r_k < x_k, r_f \geq f_k \\ r_x \geq x_k, r_f \geq f_k \end{cases}$$

在具体作预报时,首先将入选方程的周期分量向前一步计算出预测值,与外部影响因子共同组成预报因子;然后根据门限变量和门限值判定其属于哪一段,最后用该段回归模型代入显著周期分量和外部预报因子实时资料后,即可作出预报。

2 降水周期分量的计算

目前常见的时序序列周期分析有方差分析、谱分析以及 ARMA 等。为了提取汛期降水中的隐含周期,本文采用均值生成函数^[5]。均值生成函数是由时序序列按一定的时间间隔计算均值而派生出来的。现对均值生成函数简要介绍如下:

设有时间序列 $y(k) = \{y(1), y(2), \dots, y(N)\}$, 其均值生成函数定义为:

$$\bar{y}_l(i) = \frac{1}{n_l} \sum_{j=0}^{n_l-1} y(i+jl)$$

$$i = 1, 2, \dots, l \quad 1 \leq l \leq m$$

其中: $n_l = \text{INT}(N/l)$, $m = \text{INT}(N/2)$ 表示最大周期长度, l 为均值生成函数的时间间隔长度, N 为样本长度, INT 表示取整。

$$\hat{y} = \begin{cases} -376.777 + 0.693f_6 + 0.433f_{11} + 0.217f_{12} + 0.349f_{13} + 15.911x_5, & x_{12} < 227.7 \\ 3224.206 - 1.018f_{10} - 269.775x_5 - 89.865x_{27}, & x_{12} \geq 227.7 \end{cases}$$

其中第一段 $F=5.12$ 、第二段 $F=6.8$, 均通过 $\alpha=0.01$ 的显著性检验。分析两个预测方程可以看到入选的预报因子除了 x_5 外,其余入选的预报因子均不相同,两个方程的系数也不一样。这可以解释为气候系统中出现的异常信号导致了预测关系的改变。入选方程因子中降水周期分量因子所占比重较大,表明夏季青岛降水多寡的趋势直接与其自身长期振动有密切关系。这些周期分量因子分别代表了汛期降水 6 年、10 年、11 年、12 年、13 年的周期,分析可能与太阳黑子 11 年

对均值生成函数 $\bar{y}_l(i)$ 作周期型延拓,为此按公式:

$$f_i(t) = \bar{y}_l(i) \quad t = i(\text{mod}(l))$$

$$t = 1, 2, \dots, N$$

这里 mod 表示同余。由此构造出 $m-1$ 个外延的均生函数,记为 $f_i, i=1, 2, \dots, m-1$ 。此 $m-1$ 个外延均生函数即作为降水周期的备选因子。

3 汛期降水量试验

用青岛 1961 年至 1996 年 6~8 月的汛期降水量为预报对象,1960 年~1996 年亚洲地区 500hPa 月平均西风环流指数,北半球 500hPa 月平均极涡中心强度等资料作为预报因子,另外还用了同期月平均气温、日照时数两类前期气象要素作为待选预报因子(信度均超过 0.05)。根据目前气候可预报性研究指出的预报上限为 6~12 个月的结论,我们确定预报因子时段选取为上年 7 月至当年 3 月。1997~2003 年资料作为独立样本,用于模型的试报检验。

限于资料样本数的限制,本文只对模型进行单门限变量的试验。按上述建模步骤,确定出门限变量为 x_{12} (上年 9 月日照时数),门限值为 227.7,其对应的 F 值为 15.99,通过 $\alpha=0.001$ 的显著性检验。为了提高预测的稳定性,限制进入方程的最大因子个数不超过 5 个。于是得出青岛汛期降水量的门限回归预报方程为:

周期活动有关。因子 x_{27} 是上年 12 月 500hPa 极涡中心强度,这与文献[6]所指出的夏季环流异常与前春和前冬环流异常的联系密切是相一致的。对于前期气象要素因子入选模型(上年 11 月平均气温 x_5 和 x_{12}),我们认为可以解释为时间尺度较大的前期气象要素的高低在一定程度上反映了长期天气过程或大气环流形势的差异,是相应时段内一系列特定大气环流形势组合下的综合“效应”^[7]。

用上述预测方程对 1997~2003 年的独

立样本进行试报,计算结果见表1。为了比较这个回归方程的预测能力,我们还用同样资料建立了仅有周期分量的门限自回归模型和仅有外部预报因子的多元门限回归模型,计算结果列在表1中。表中所采用的评分办法是长期预报业务标准即预报值与实况值距平百分率符号相同、两者符号相异或其中之一为零但两者误差绝对值 $\leq 20\%$ 时,预报正确,反之错误。从表中可见,三种方法预测结果的评定分别为:6/7、4/7、5/7,带有周期分量

的多元混合门限回归的结果要明显好于其他两个回归模型。从平均相对误差来看,三种模型的拟合值分别为18.7%、12.6%、23.7%,7年的预报平均相对误差分别为19.1%、26.1%、28.0%。比较来看只有多元混合门限回归的预报值与拟合值较为接近,而其他两个回归模型相差较大,表明带有周期分量的多元混合门限回归方程具有较好的预测稳定性,准确率也更高。

表1 1997~2003年青岛汛期降雨量试报结果

年份	实况 (距平百分率)	y ₁				y ₂				y ₃			
		预测值 /mm	距平 百分率	相对 误差	评定	预测值 /mm	距平 百分率	相对 误差	评定	预测值 /mm	距平 百分率	相对 误差	评定
1997	243.5(-39.9)	298.1	-26.4	22.4	√	287.3	-29.1	18.0	√	301.1	-25.7	23.7	√
1998	490.7(21.1)	467.2	15.3	4.8	√	413.5	2.0	15.7	√	453.7	11.9	7.5	√
1999	261.0(-35.6)	336.8	-16.9	29.0	√	332.7	-17.9	27.5	√	329.5	-18.7	26.2	√
2000	451.5(11.4)	370.5	-8.6	17.9	√	300.8	-25.8	33.4	×	494.1	21.9	9.4	√
2001	601.0(48.3)	471.8	16.7	21.3	×	728.5	79.7	21.2	×	371.5	-8.3	38.2	×
2002	202.7(-50.0)	258.2	-36.3	27.4	√	318.2	-21.5	57.0	×	342.5	-15.5	68.9	×
2003	410.2(1.2)	452.1	11.5	10.2	√	369.8	-8.8	9.8	√	481.9	18.9	17.5	√
平均值				19.1				26.1				28.0	

注:1. 距平百分率、相对误差的数值为百分数;2. y₁ 为带有周期分量的多元混合门限回归; y₂ 为仅有周期分量的门限自回归; y₃ 为多元门限回归。

为了更加清晰看到试报结果,将7年的试报结果绘制在图1中。从图1中可以看到除2001年的差距较大外,其余年份的预测值和实况值的差距较小,并且整体预测趋势与实况趋势基本一致。在具体分析2001年预测差距时,我们发现该年份汛期期间青岛遭遇一场多年来少见的登陆北上台风暴雨。这次过程造成青岛8月1日降雨量达219.1mm。由此可见,影响青岛汛期降水的因素很复杂,有必要引进更多资料如海温场等作进一步试验。

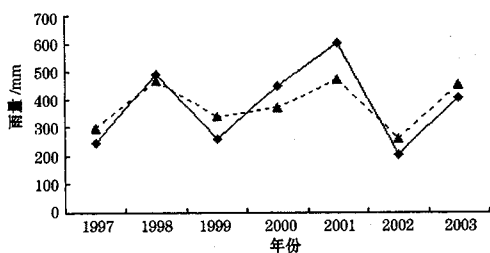


图1 青岛汛期降雨量试报结果曲线
实线为实况值,虚线为预测值

4 结语

(1) 由于既考虑到预测对象自身的周期性变化规律,同时又兼顾前期外部因素的影

响,带有周期分量的多元混合门限回归模型的汛期降水量预测较为稳定,效果也较好,值得进一步研究应用。同时也提示我们在作气候预测时应该充分考虑气候系统的自身变化和外部影响两类因素。

(2) 本次试验的总体拟合率和试报结果的精度还不是很高,对个别异常年份的预测能力还有待提高,因此建立的预测模型还有待更多资料的进入,以期得到更深入的探讨和检验。

参考文献

- 1 严华生,曹杰.多元门限回归的一种建模方案.大气科学,1994,18(2):194~199.
- 2 严华生,谢应齐,曹杰.非线性统计预报方法及其应用.昆明:云南出版社,1998:58~62.
- 3 周家斌.一种新的时间序列预报方法.科学通报,1983,28:1250~1252.
- 4 魏凤英,曹鸿兴.逐步回归周期分析.气象,1983,(2):2~4.
- 5 魏凤英,曹鸿兴.长期预测的数学模型及其应用.北京:气象出版社,1990:49~90.
- 6 孙力.东亚地区春冬季与夏季大气环流异常相互关系的研究.应用气象学报,2002,13(6):650~661.
- 7 周洪祥.灾害性天气的预测方法.北京:气象出版社,2002:120~121.

Prediction Test with Multivariate Mixed Threshold Regression Model for Rainy Season Precipitation

Sun Zhen Xu XiaoLiang

(Qingdao Meteorological Bureau, Shandong Province, 266003)

Abstract

Considering the effects of the outside factor and the periodic change of the predictand, the multivariate mixed threshold regression model with periodic components is developed. It shows that the model is better and steadier on prediction capacity in the 7-year experiments for rainy season precipitation.

Key Words: mixed threshold regression periodic components prediction test