

门限回归模型预测粤中西部冬春气温^{*}

彭 端¹, 温坚培¹, 周 静¹, 郭媚媚¹, 陈创买²

(1. 广东省肇庆市气象局, 广东 肇庆 526040;

2. 中山大学大气科学系, 广东 广州 510275)

摘 要: 用门限自回归模型和混合门限回归模型对广东省中西部的冬春季平均气温做预测。选取北半球 500 hPa 高度场和太平洋海温场的分区分季节的多个气候背景场的前若干个主分量作为因子(共 1792 个), 这些因子包含着气候背景场的主要信息, 通过对预报量和预报因子群的相关筛选, 选取相关系数最大的因子建立混合门限非线性回归模型。并初步探讨了气候背景场对预报量的影响。结果表明: 混合门限回归模型优于门限自回归模型; 前者能较好地拟合冬春气温序列, 预报效果较理想。粤中西部冬季气温场与东北半球南部 3—4 月高度场第 5 个主模态相关最密切, 当 3—4 月 500 hPa 高度场东亚大陆为低槽控制, 日本海太平洋附近为高压脊, 有利于当年冬季气温偏低; 春季气温场则与北半球冬季 500 hPa 高度场第 8 个主模态相关密切, 当冬季 500 hPa 高度场乌拉尔山附近地区为明显的阻塞高压形势, 有利于次年春季气温偏低。

关键词: 主分量因子; 混合门限回归模型; 冬春气温; 广东中西部

中图分类号: P468.0⁺21 **文献标识码:** A **文章编号:** 0529-6579(2005)S2-0262-04

门限回归预测模型是具有有效地描述非线性振动现象, 解析各种类型的稳定循环, 能反映出类似共振、跳跃的突变功能的非线性回归模型^[1,2], 在很多科学领域中得到了广泛的应用。门限自回归模型是基于分段样本建立自回归模型而得, 根据外加因子序列与原序列之间的关系建立分段回归模型, 我们称之为混合门限自回归模型。因此, 此模型既可以研究序列本身内部性质及变化规律, 也可以通过研究外加相关物理因子与序列内部性质之间的非线性的关系来预测序列未来状态。

广东省中西部位于西江流域的下游沿岸一带, 地形大部分由冲积平原、河漫滩平原、谷地平原组成, 土地肥沃、物产丰富, 素有“小三角洲”之称, 其冬春气温的变化对其农业经济作物的生长发育有着重要的影响。本文拟用混合门限回归模型对其进行预测, 并初步探讨其影响因子的背景, 为短期气候预测提供依据。

1 资料来源

1.1 预报量场资料及处理

取广东中西部 10 市县自建站至 2002 年的月平均气温。并用序列差值订正法^[3]进行插补, 把各站系列向前延长至 1955 年。形成各站 1955~2002 年月气温资料。并作区域平均, 产生粤中西部月平均气温资料。

1.2 因子资料

取 1951—2002 年北太平洋海温场 5 个区春夏秋冬 4 季、6 个自然季节、前、后汛期、夏季半年和冬季半年等总共 14 个季节的气候场的前 8 个主分量。以及取北半球 500 hPa 高度场 11 个区 14 个季节气候场的前 8 个主分量, 两者合共 1792 个主分量, 将这些主分量(又称为主模态)作为预报因子。

2 门限自回归模型与混合门限回归模型

据文献[2], 门限自回归模型具有有效地描述非线性振动现象, 能反映类似共振、跳跃的突变性能。其模型为:

$$X_t = \alpha_0^{(j)}(t) + \sum_{i=1}^{k_j} \alpha_i^{(j)}(t)X_{(t-i)} + \epsilon_i^{(j)}(j) \quad (1)$$

$$X_{(t-d)} \in R_j, \quad R_j = [r_j - 1, r_j], \quad j = 1, 2, \dots, L$$

其中, X_t 是 t 时刻带有随机噪声的气象观测值, $\{\epsilon_i^{(j)}\}$ 是 L 个相互独立的正态白噪声序列, d 为时滞, X_{t-d} 为门限变元; $r_0 < r_1 < \dots < r_L$, $r_1, r_2, \dots, r_{L-1}$ 为门限值。门限值将实轴分为 L 段, 当 X_{t-d} 落入第 j 段时, 采用上式中第 j 个自回归模型描述, 阶数为 k_j 。

* 收稿日期: 2005-10-02

作者简介: 彭端(1973 年生), 女, 工程师, E-mail: peng075@163.com

如果有另一时间序列 $\{Y_t\}$ 与预报量序列 $\{X_t\}$ 相关性很好, 那么就可以考虑使用混合门限自回归模型。设 $\{X_t, Y_t, t=1, 2, \dots\}$ 是一对时间序列, d 为正整数, 系列作两段, r 是实的门限值, 如果 $\{X_t\}$ 可表为 $X_t =$

$$\begin{cases} a_0^{(1)} + \sum_{i=1}^{p_1} a_i^{(1)} x_{t-i} + \sum_{i=0}^{q_1} b_i^{(1)} y_t - i + \epsilon_t^{(1)}, & \text{当 } y_{t-d} \leq r \\ a_0^{(2)} + \sum_{i=1}^{p_2} a_i^{(2)} x_{t-i} + \sum_{i=0}^{q_2} b_i^{(2)} y_t - i + \epsilon_t^{(2)}, & \text{当 } y_{t-d} \geq r \end{cases} \quad (2)$$

其中, p_1, q_1, p_2, q_2 是正整数, $a_i^{(j)}, i=1, \dots, p_j, b_i^{(j)}, i=1, \dots, p_j, j=1, 2$ 是常数, $\{\epsilon_t^{(j)}\}$ 是零均值, 方差为 $\sigma_{\epsilon_t^{(j)}}^2$ 的白噪声 ($j=1, 2$), 且 $\{\epsilon_t^{(1)}\}$ 和 $\{\epsilon_t^{(2)}\}$ 相互独立, 那么 $\{x_t, y_t\}$ 称为混合门限自回归模型。 r 称为门限值, d 称为延迟参数, $(p_1, q_1), (p_2, q_2)$ 分别为第一、第二分段的阶数, 按一定规则, 将时间序列分为 2 段, 并用 AIC 准则进行建模, 即可对序列作出预报。

3 预报因子的选取

北半球 500 hPa 高度场是反映东亚气候变化的主要因素之一, 太平洋海温场也是全球气候异常的强信号——厄尔尼诺现象的产生源地。因此这两个气候场备受气候学家的关注, 它们对东亚气候产生重要的影响。那么, 如何获取这两个气候场对局地气候变化的影响信息呢? 研究表明, 一种最有效从因子场中获取含高信息量的因子的方法, 就是对气候场的不同区域和不同季节进行主分量分析, 提取其前若干个主分量作为待选预报因子。这些场的主分量含有很高的信息量, 并且具有明确的关于场分布的物理背景。由前述从对这两个场的分区分季节的主分量分析可得到 1792 个待选预报因子。

由于预报因子的数量很大, 进入回归模型的因子是通过对预报量与所有待选预报因子进行相关筛选而得。设第 k 个预报量 y_k 与第 l 个预报因子 F_l 的相关系数为:

$$\gamma_{kl} = \frac{1}{m} \sum_{i=1}^m y_{ki} - \overline{y_k} \overline{f_l} - \overline{f_l} \left(S_k \circ S_l \right) \quad (3)$$

其中, $\overline{y_k}$ 和 $\overline{f_l}$ 为平均值, S_k 和 S_l 为标准差。给定一显著性水平 α ($\alpha < 0.05$), 找出几十个可能预报因子称因子群。取其相关系数最大者作为混合门限回归模型的最相关因子。

4 平均气温的门限回归预测模型

4.1 门限自回归预测

对广东省中西部气温作区域平均, 取其冬春季

平均气温序列进行门限自回归预测, 最大延迟参数为 4, 模型的最大阶数为 4, 模型只有限作 2 个分段。从得出的预报拟合曲线(图略)分析, 冬季平均气温门限自回归预报曲线基本与实况曲线趋势一致; 但春季平均气温门限回归预报拟合效果不好, 预报曲线变化幅度很小, 基本在平均值附近作小的振动。说明春季平均气温序列本身的相关性较差。

4.2 混合门限自回归模型

由第 3 节所述可知, 粤中西部冬、春平均气温序列都能找到相关性好的物理气候背景的主分量因子群。建模时, 从各因子群中取出与预报量序列单相关系数最好的一个因子建立回归方程, 最大延迟参数为 4, 模型的最大阶数为 4, 得以下预报模型。

(1) 春季平均气温预报模型

TARSO (2, (4, 4), (1, 2)), 最大后延参数取 $D=1$, 门限值 $TR=-0.49$, $NAIC=-0.93$, $AIC1=-23.94$, $AIC2=-17.03$, 预报方程为:

$$X_t = \begin{cases} = 22.34 + 0.42x_{t-1} + 0.43x_{t-2} - 0.40x_{t-3} - 0.43x_{t-4} - \\ 0.22 - 0.15y_{t-1} + 0.08y_{t-2} - 0.04y_{t-3} - 0.26y_{t-4} + \epsilon_t^{(1)} \\ (y_{t-1} \leq -0.49) \\ = 16.37 + 0.25x_{t-1} - 0.06 + 0.02y_{t-1} - 0.05y_{t-1} + \epsilon_t^{(2)} \\ (y_{t-1} > -0.49) \end{cases}$$

其中, y 为冬季北半球 500 hPa 高度场第 8 个主模态因子, 其与预报量的单相关系数 $r=-0.497$ 。

(2) 冬季平均气温预报模型

TARSO (2, (3, 1), (4, 4)), 最大后延参数取 $D=2$, 门限值 $TR=1.85$, $NAIC=-0.52$, $AIC1=2.08$, $AIC2=-25.01$, 预报方程为: X_t

$$\begin{cases} = 12.83 + 0.19x_{t-1} + 0.34x_{t-2} - 0.49x_{t-3} - 0.26 + \\ 0.07y_{t-1} + \epsilon_t^{(1)}, & (y_{t-1} \leq 1.85) \\ = 26.60 + 0.29x_{t-1} - 0.78x_{t-2} - 0.33x_{t-3} - 0.19x_{t-4} - \\ 0.73 - 0.28 - y_{t-1} - 0.11y_{t-2} - 0.35y_{t-3} + 0.22y_{t-4} + \epsilon_t^{(2)} \\ (y_{t-1} > 1.85) \end{cases}$$

其中, y 为 3—4 月东北半球南部 500 hPa 高度场第 5 个主模态因子, 其与预报量的单相关系数 $r=-0.481$ 。

图 1 给出冬春季平均气温的混合门限回归拟合曲线图, 它的拟合效果比门限自回归模型好。通过对门限自回归模型与混合门限回归模型作粤中西部平均气温预测的有关统计量对比以及 2003 年的预报检验(表 1), 发现混合门限回归模型从联合方差、拟合误差、拟合最大误差、2003 年预报误差都优于门限自回归模型。2003 年春季平均气温预报误差达到 0.16°C , 预报效果是相当理想的; 混合门限回归模型作 2002—2003 年冬季气温绝对误

差仅为 0.15℃，而门限自回归模型也只有 0.16℃，表明冬季气温本身序列相关性比春季气温好。

表 1 门限自回归模型与混合门限回归模型作
粤中西部平均气温预测对比

Tab 1 The comparative prediction to temperature in mid-west GD by two threshold model					℃
项目	门限自回归模型		混合门限回归模型		
时距	春季	冬季	春季	冬季	
时滞值	1	3	4	4	
联合方差	0.92	0.14	0.24	0.47	
拟合误差	0.82	1.23	0.49	0.69	
拟合最大绝对误差	2.13	2.67	1.36	1.83	
2003 年预测值	21.25	14.46	23.26	14.15	
2003 年实况值	23.1	14.3	23.1	14.3	
预报绝对误差	1.85	0.16	0.16	0.15	

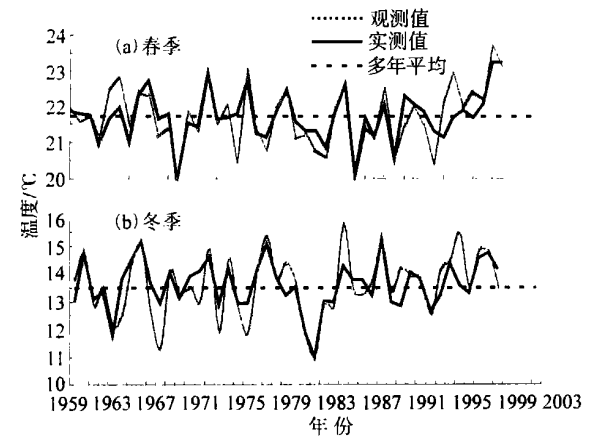


图 1 粤中西部平均气温混合门限回归拟合及试报曲线
Fig 1 Fitting and prediction curve of temperature in mid west
GD by the mixed threshold model

5 北半球 500 hPa 高度场对粤中西部气温的影响

各预报模型所选到的气候背景场的因子，都是高度场。冬季气温是与东北半球南部 3—4 月 500 hPa 高度场第 5 个主分量相关最密切，呈反相关关系。根据主分量与特征向量场相配合的原则，当主分量为正值时，反映的气候背景场是东亚大陆为低槽控制，槽中心在华北地区，日本海太平洋附近为高压中心（如图 2），再根据冬季气温与东北半球南部 3—4 月 500 hPa 高度场第 5 个主分量的相关系数为-0.481，那么，上述这种形势将有利于当年冬季气温偏低；反之，当主分量为负值时，反映的是东亚大陆为高压值控制，日本海太平洋附近为低槽的形势，有利于当年冬季气温偏高。春季气温与北半球冬季 500 hPa 高度场第 8 个主模态相关较密

切，反映的是北半球两槽三脊（三槽两脊）的环流分布（图 3），由图 3 可知，当主分量为正值时，反映在特征向量场上为乌拉尔山附近地区为明显的阻塞高压形势，中国大陆为低槽控制，再根据春季气温与北半球冬季 500 hPa 高度场第 8 个主分量的相关系数为-0.497，那么，上述这种形势将有利于次年春季气温偏低；反之，当主分量为负值时，反映的是乌拉尔山附近地区为明显的低槽形势，中国大陆为高压脊控制，这种形势将有利于次年春季气温偏高。

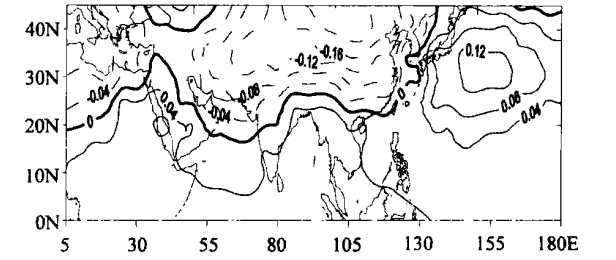


图 2 3—4 月东北半球南部高度场第 5 特征向量场
Fig 2 The fifth principal component of the height in the south
of the north-eastern hemisphere in March to April season

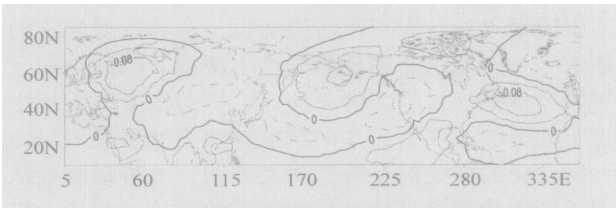


图 3 冬季北半球高度场第 8 特征向量场
Fig 3 The eighth principal component of the north hemisphere
of the 500 hPa geopotential height in winter

6 结 论

- (1) 门限回归模型可以作气温的预报，其预报误差较小。外加因子的混合门限回归模型优于门限自回归模型。
- (2) 粤中西部冬季气温是与东北半球南部 3—4 月 500 hPa 高度场第 5 个主模态相关最密切，当 3—4 月 500 hPa 高度场东亚大陆为低槽控制，日本海太平洋附近为高压脊，有利于当年冬季气温偏低；反之，当东亚大陆为高压脊控制，日本海太平洋附近低槽的形势有利于当年冬季气温偏高。
- (3) 粤中西部春季气温与冬季北半球 500 hPa 高度场第 8 个主牛态相关最密切，当冬季 500 hPa 高度场乌拉尔山附近地区为明显的阻塞高压形势，有利于次年春季气温偏低；反之，当乌拉尔山附近地区为明显的低槽形势，有利于次年春季气温偏

高。

参考文献:

[1] 余锦华, 王振堂 编著. 经济预测方法程序和实例[M] . 广州: 中山大学出版社, 1996: 366—375.

[2] 杨明, 张方诚, 沙文生. 具有时变参数的门限自回归模型及其在气候预报中的应用[J] . 热带气象学报, 1996, 12(3): 224—229.
[3] 么枕生, 丁裕国. 气候统计[M] . 北京: 气象出版社, 1990: 786—788.

Application of Threshold Regression Modeling to Prediction of the Temperatures in Mid-western Guangdong

PENG Duan, WEN Jian pei, ZHOU Jing, GUO Mei mei, CHEN Chuang mai

(1. Zhaoqing Meteorological Observatory, Zhaoqing 526040, Guangdong, China;
2. Department of Atmospheric Science, Sun Yat sen University, Guangzhou 510275, China)

Abstract: In this paper, the threshold autoregressive model is improved by analyzing the principal component of the seasonal, sectorized 500 hPa geopotential height in the northern hemisphere and the sea surface temperature field of Pacific from 1951 to 2002 to get the prognostic factor group (1792 factors) consisted of the main information of the climatological background field, and choose the factor with the highest correlation to set up the mixed threshold nonlinear model. The mean temperature in both winter and spring in the mid-western Guangdong are forecasting trial by the authors. In the mean time, the influence of the climatological background field to the forecasting variable is initially discussed. Computational results show that the mixed threshold model is better than the threshold autoregressive one, and it can forecast the series of the temperature with perfect effect. The temperature of winter in the mid-western Guangdong has the most closed relationship with the fifth principal component of the height in the south of the north-eastern hemisphere in March to April season. The temperature of spring has most closed relationship with the eighth principal component of the north hemisphere of the 500 hPa geopotential height in winter.

Key words: factors of principal component; mixed threshold regressive model; temperature in winter and spring; mid-west Guangdong