

文章编号: 1004-4965 (2009) 02-0196-09

气候场的主分量逐步回归预测模型及应用

陈创买¹, 周文², 林爱兰³

(1. 中山大学环境科学与工程学院大气科学系, 广东 广州 510275; 2. 香港城市大学物理暨材料系, 香港;
3. 中国气象局广州热带海洋气象研究所, 广东 广州 510080)

摘 要: 提出一种气候场的主分量逐步回归预测模型, 该模型将气候场的预测变成对该场主分量的预测。提取对气候场变化有重要影响的 500 hPa 高度、太平洋及印度洋海温和全球海平面气压等含不同区域不同季节的因子场的前若干个含高信息量的优质主分量因子。通过相关筛选和双重检验逐步回归, 建立气候场的主分量的预测方程, 由此就建立气候场与多个因子场之间的联系, 但是模型却仍然保持着两个场相关的特点。根据气候场的特征向量的近似不变性, 将其与主分量配合进行反算, 从而得到气候场的预测。以广东汛期降水预测作为试验例子, 计算了它的主分量, 分析了它的时空分布特征及分类, 对 2003—2005 年的分布作预测并与实测分布作比较及进行误差检测, 发现预测效能显著。有关预测的年限、区域范围、要素和因子的物理背景等问题有待进一步研究。

关键词: 气候统计预测; 气候要素场; 主分量分析; 双重检验逐步回归; 广东汛期降水

中图分类号: P468.0

文献标识码: A

1 引 言

对气候要素场(以下简称气候场)的预测, 在日常气象业务中十分重要, 不少学者对此进行过研究。徐瑞珍等^[1]作了两个场的相关分析, 谢炯光等^[2]把奇异值分解方法用于广东季降水预测, 张尧庭等^[3]和肯德尔^[4]对两个场进行典型相关分析。但这些方法仅局限于考虑两个场之间的相关。实际上影响气候场变化的因素决不会仅仅是单一的因子场, 例如广东夏季降水决不会仅受海温场的影响, 或仅受某一高度场的影响。因此对气候场的预测如何与多个影响因子场相联系是一个值得研究的问题。Hannack 等^[5]对美国季降水量的预测, Broccoli 等^[6]对北太平洋气压场的预测, Michael 等^[7]对雷暴天气的预测, 都从不同角度应用了主分量。金菊良等^[8]用遗传门限自回归模型对气象序列作预测。主分量分析能将随时间变化的气候场分解成为空间函数(特征向量矩阵)部分和时间函数(主分量矩阵)部分^[9], 空间函数部分概括场的不随时间变化的地域分布特点, 而随时间变化的主分量的前个占有原空间点变量的总

方差的很大部分。根据主分量的这些特点, 可以找到将对气候场的预测与多个因子场的影响相联系的途径。首先, 通过主分量分析, 可以将对气候场的预测变成对气候场的前若干重要主分量的预测。其次, 预测因子可取自对气候变化有重要影响的 500 hPa 高度场、太平洋及印度洋海温场和全球海平面气压场等含不同区域不同季节要素场的前若干个主分量, 它们集中了各因子场的最重要信息。通过相关筛选大量地收集那些含高信息量的预测因子, 用逐步回归分析来建立气候场与各气候因子场的广泛联系。因为预测是对气候场的主分量来进行的, 而且各种气候场及因子场的主分量都与其特征向量即空间分布型相对应, 因此预测量与预测因子的相关, 实际上就是气候场的某种空间分布型与某因子场的一定空间分布型的相关。这说明利用因子场的主分量作为预测因子具有十分明确的气候背景, 物理意义十分奥妙。由于特征向量矩阵的近似不变性^[10], 所以当完成对主分量矩阵的预测后, 可利用该两个矩阵作回算, 实现对原始场和未来气候场分布的预测。以广东汛期(4—9月)降水为例进行探讨。

收稿日期: 2007-10-08; 修订日期: 2007-12-12

基金项目: 广东省科技计划项目“广东干旱年景预测技术研究”(2005B32601007); 国家自然科学基金(40675051)共同资助

通讯作者: 陈创买, 男, 教授(已退休), 主要从事气象学、气候学和统计气候预测的教学和研究。E-mail: chencm@pub.guangzhou.gd.cn

2 预测模型的数学原理

2.1 概 述

将要预测的气候场看作一个整体, 根据主分量分析的特点, 对该场的主分量进行预测。首先对气候场作标准化处理, 而后进行主分量计算得到特征向量矩阵和主分量矩阵, 通过相关分析和逐步回归计算, 对该场的标准化主分量逐个建立预测方程。预测因子可利用各种因子场的主分量。然后对该气候场的标准化主分量的拟合和预测结果进行回算, 得到标准化变量和原始变量的气候场的拟合和预测, 并作误差估计和评估。

2.2 数学原理

2.2.1 对气候场进行标准化处理

设原始气候场是含有 p 个空间点和 m 个时间点的 $p \times m$ 阶矩阵 $X_{p \times m}$, 则其标准化变量矩阵为,

$$Z_{p \times m} = \begin{bmatrix} Z_{11} & \cdots & Z_{1m} \\ \vdots & \ddots & \vdots \\ Z_{p1} & \cdots & Z_{pm} \end{bmatrix} \quad (1)$$

于是原始气候场可以写成,

$$X_{p \times m} = S_p Z_{p \times m} + E_p' \quad (2)$$

其中 $E_p = (\bar{E}_1, \bar{E}_2, \dots, \bar{E}_p)$ 为 p 阶平均值行矩阵,

$S_p = (s_1, s_2, \dots, s_p)$ 为标准差对角矩阵。

2.2.2 对气候场作主分量计算

对标准化气候场 $Z_{p \times m}$ 作主分量计算^[8],

$$Z_{p \times m} = V_{p \times r} W_{r \times m} \quad (3)$$

其中 V 为 $p \times r$ 阶特征向量阵, W 为 $r \times m$ 阶主分量阵, (r 为 p 与 m 的最小值)。再取占 96% 以上方差的前 K (≤ 25) 个主分量进行标准化,

$$Y_{r \times m} = \Lambda^{-\frac{1}{2}} W_{r \times m} \quad (4)$$

其中 $\Lambda^{-\frac{1}{2}}$ 为以特征值开方为分母的对角阵。

2.2.3 对气候场的标准化主分量进行相关筛选和逐步回归预测

由于主分量预测因子数量大, 用于建立回归方程的因子是通过对该预测量与所有预测因子进行相关筛选而得。第 k 个预测量 y_k 与第 l 个预测因子 f_l 的相关系数为,

$$r_{kl} = \frac{1}{m} \sum_{i=1}^m \frac{y_{ki} - \bar{y}_k}{S_k} \frac{f_{li} - \bar{f}_l}{S_l} \quad (5)$$

其中 \bar{y}_k 与 \bar{f}_l 为平均值, S_k 与 S_l 为标准差。给定显著性水平 α , 相应显著相关系数为 r_α , 若 $r_{kl} \geq r_\alpha$, 则

相关显著, 该因子可选出作为可能预测因子供逐步回归计算。

对预测场的前 K (一般 $K \leq 25$) 个标准化主分量逐个建立回归方程, 则第 k 个标准化主分量的回归预测模型为,

$$\tilde{y}_{kt} = C_{k0} + \sum_{i=1}^{L_k} C_{ki} F_{kit} + e_k \quad (6)$$

($k=1, 2, \dots, K; \quad t=1, 2, \dots, m$)

其中 C_{k0} 为常数, C_{ki} 为回归系数, F_{ki} 为预测因子, e_k 为误差, L 为可能预测因子数。

对预测场第 k 个的标准化主分量进行逐步回归计算, 假设有 L_k 个预测因子进入回归方程, 由式 (6) 可得,

$$\hat{y}_{kt} = C_{k0} + \sum_{i=1}^{L_k} C_{ki} f_{kit} \quad (7)$$

($t=1, 2, \dots, m; \quad k=1, 2, \dots, K$)

利用上述对气候场的前 K 个主分量的逐步回归计算, 可对该场的标准化主分量进行拟合和预测,

$$\hat{y}_{k \times (m+s)} = (\hat{y}_{k \times m} \quad \hat{y}_{k \times s}) \quad (8)$$

其中 $\hat{y}_K = (\hat{y}_{k1}, \hat{y}_{k2}, \dots, \hat{y}_{k(m+s)})'$, 由式 (7) 计算得出。

2.2.4 对标准化变量场和原始气候场的预测

首先对气候场标准化主分量的拟合值和预测值进行反算, 使其变成普通主分量场,

$$\hat{W}_{k \times (m+s)} = \Lambda^{\frac{1}{2}} \hat{Y}_{k \times (m+s)} = \Lambda^{\frac{1}{2}} (\hat{Y}_{k \times m} \quad \hat{y}_{k \times s}) \quad (9)$$

其中 $\Lambda^{\frac{1}{2}}$ 为 K 阶特征值开方对角阵。

假定在很小的 s 时间内预测场的空间函数阵 $V_{p \times r}$ 是不变的^[10], 将式 (9) 代入式 (3) 右边得,

$$\hat{Z}_{p \times (m+s)} = V_{p \times K} \hat{W}_{K \times (m+s)} = V_{p \times K} (W_{k \times m} \quad W_{k \times s}) = (\hat{Z}_{p \times m} \quad \hat{Z}_{p \times s}) \quad (10)$$

其中 $\hat{Z}_{p \times (m+s)} = (\hat{z}_1, \hat{z}_2, \dots, \hat{z}_p)'$ 为对各站的标准化变量的拟合和预测。

其次, 将式 (10) 代入式 (2) 的右端, 得到原始气候场的拟合和预测,

$$\hat{X}_{p \times (m+s)} = S_p \hat{Z}_{p \times (m+s)} + E_p' \quad (11)$$

其中 E_p 及 S_p 的意义与式 (2) 相同。

2.2.5 对预测结果的误差估计

由预测场和原始场可以估计预测误差。第 k 个变量的绝对误差估计为,

$$D_k = \left[\frac{1}{m} \sum_{i=1}^m (\hat{x}_{ki} - x_{ki})^2 \right]^{\frac{1}{2}} \quad (k=1, 2, \dots, p) \quad (12)$$

相对误差估计为,

$$E_{rk} = \frac{D_k}{x_k} \quad (k=1,2,\dots,p) \quad (13)$$

式中 \bar{x}_k 为第 k 个变量的平均值。总平均相对误差为,

$$\bar{E}_r = \frac{\sum_{k=1}^p E_{rk}}{P} \quad (14)$$

3 广东汛期降水的时空变化特征

3.1 资料及处理

预测场取广东 1954—2005 年 48 站月降水量为基本资料, 其中 1954—2002 年资料用于统计, 2003—2005 年资料用于试报。因子场取太平洋海温场 1951—2003 年 286 个 5×10 格点月平均海温资料和北半球 500 hPa 高度场 1951—2003 年 576 个 5×10 梅花格点月平均高度资料。全球海平面气压场 1951—2003 年 1 368 个 5×10 网格点月平均气压。将太平洋海温场分为 5 个区, 北半球 500 hPa 高度场分为 11 个区, 全球海平面气压场分为 7 个区。对每个区各取 14 种季节。每一个区的每一种季节构成一个因子场, 总共可形成 322 个因子场。对它们进行主分量分析, 各取其前 8 个主分量作为预测因子。

14 个季节包括: 春、夏、秋、冬, 6 个自然季节: 1—2 月、3—4 月、5—6 月、7—8 月、9—10 月和 11—12 月, 前、后汛期、夏半年和冬半年, 实际上还可以设立更多的季节。

北半球 500 hPa 高度场的 11 个区设为: 1 区: 东北半球北部 ($5 \sim 180^\circ\text{E}$, $45 \sim 85^\circ\text{N}$); 2 区: 东北半球南部 ($5 \sim 180^\circ\text{E}$, $10 \sim 45^\circ\text{N}$); 3 区: 西北半球北部 ($185 \sim 360^\circ\text{E}$, $45 \sim 85^\circ\text{N}$); 4 区: 西北半球南部 ($185 \sim 360^\circ\text{E}$, $10 \sim 45^\circ\text{N}$); 5 区: 东北半球 ($5 \sim 180^\circ\text{E}$, $10 \sim 85^\circ\text{N}$); 6 区: 西北半球 ($180 \sim 360^\circ\text{E}$, $10 \sim 85^\circ\text{N}$); 7 区: 整个北半球 ($5 \sim 360^\circ\text{E}$, $10 \sim 85^\circ\text{N}$); 8 区: 太平洋地区南部 ($90 \sim 270^\circ\text{E}$, $10 \sim 45^\circ\text{N}$); 9 区: 太平洋地区北部 ($90 \sim 270^\circ\text{E}$, $50 \sim 85^\circ\text{N}$); 10 区: 北半球副热带高压区 ($5 \sim 360^\circ\text{E}$, $10 \sim 45^\circ\text{N}$); 11 区: 北半球极涡区 ($5 \sim 360^\circ\text{E}$, $50 \sim 85^\circ\text{N}$)。

北太平洋海温场的 5 个区设为: 1 区: 北太平洋西部 ($110 \sim 180^\circ\text{E}$, $10^\circ\text{S} \sim 15^\circ\text{N}$); 2 区: 北太平洋东北部 ($180 \sim 270^\circ\text{E}$, $15 \sim 50^\circ\text{N}$); 3 区: 赤道东太平洋 ($180 \sim 270^\circ\text{E}$, $10^\circ\text{S} \sim 15^\circ\text{N}$); 4 区: 全北太平洋 ($110 \sim 285^\circ\text{E}$, $15^\circ\text{S} \sim 50^\circ\text{N}$); 5 区: 太平洋热带区 ($110 \sim 285^\circ\text{E}$, $10^\circ\text{S} \sim 35^\circ\text{N}$)。

全球海平面气压场的 7 个区设为: 1 区: 东北

半球区 ($0 \sim 180^\circ\text{E}$, $0 \sim 40^\circ\text{N}$); 2 区: 东南半球区 ($0 \sim 180^\circ\text{E}$, $-40 \sim 0^\circ\text{N}$); 3 区: 西北半球区 ($180 \sim 360^\circ\text{E}$, $0 \sim 40^\circ\text{N}$); 4 区: 西南半球区 ($180 \sim 360^\circ\text{E}$, $-40 \sim 0^\circ\text{N}$); 5 区: 东半球区 ($0 \sim 180^\circ\text{E}$, $-40 \sim 40^\circ\text{N}$); 6 区: 西半球区 ($180 \sim 360^\circ\text{E}$, $-40 \sim 40^\circ\text{N}$); 7 区: 全球区 ($5 \sim 360^\circ\text{E}$, $-40 \sim 40^\circ\text{N}$)。

3.2 广东汛期平均降水及标准差的分布

广东汛期平均降水分布 (图 1) 有 3 个大降雨中心分别位于海陆丰地区、清远-佛冈和阳江地区, 其中心雨量均达到 1 800 mm 以上。小雨中心出现在东北部、北部、西部和雷州半岛地区, 雨量在 1 100 mm 以下。图 2 为广东汛期降水标准差分布, 它和平均分布场非常相似, 也有 4 个大小值分布中心。

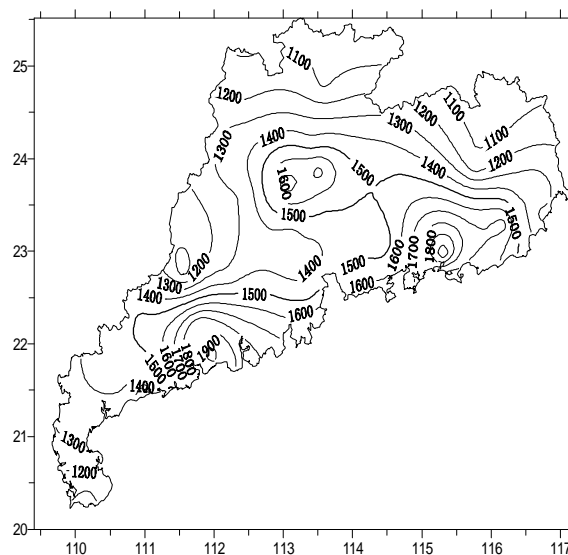


图 1 广东汛期平均降水量分布

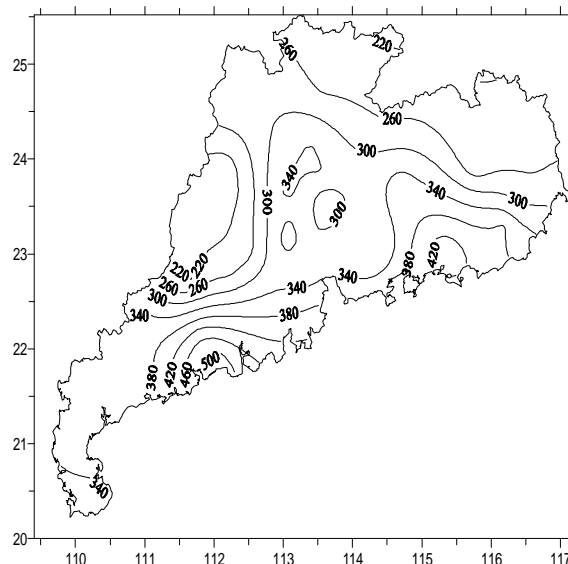


图 2 广东汛期降水标准差分布

3.3 广东汛期降水的主分量计算

对广东汛期(4—9月)降水进行主分量计算得到特征向量矩阵和主分量矩阵,其前25个主分量所占总方差百分数如表1,显然前4个主分量已占总方差约69%,前25个达97.1%。以下对前4个主分量和特征向量作分析。

表1 广东汛期降水前25个主分量所占方差

主分量	特征值	占总方差/%	累计方差/%
1	24.02	46.20	46.20
2	5.63	10.82	57.03
3	3.48	6.70	63.74
4	2.74	5.28	69.02
5	1.87	3.61	72.63
6	1.49	2.87	75.50
7	1.20	2.30	77.81
8	1.11	2.13	79.94
24	0.23	0.45	96.65
25	0.23	0.44	97.09

3.4 广东汛期降水的时空分布特征

3.4.1 第1特征向量及主分量

广东汛期降水第1特征向量分布如图3。全省均为正值,大值中心分布在珠江三角洲及至西北和北部,东南伸至汕头。小值中心分布在梅州、清远和西南部及雷州半岛。这是全省性的涝(旱)分布(I)型。当第1主分量为正(负)时出现全省性的涝(旱)。从表1可知这一分布型占总方差的46.2%。最严重的旱、涝年份都与这一型分布有关。

图4是第1主分量变化曲线,从图4可见,1973、2001和1961年主分量有最大值属于严重的涝年;1956、1963和1991年主分量有最小值属于严重的旱年。例如1973年广东汛期降水全为正距平分布(图5),即是涝,有4个大值区分别出现在潮汕、珠海至阳江、雷州半岛和粤北地区。小值区分别出现梅州、郁南至信宜地区。这与平均分布有比较大的差异。1956年广东汛期降水与1973年完全相反为负距平分布,即是干旱,严重干旱出现在汕头、珠海、韶关和信宜,均在-40%以上(图6)。

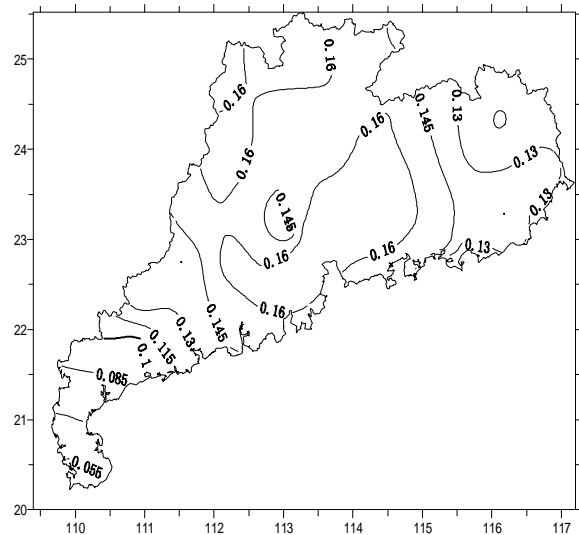


图3 广东汛期降水第1特征向量分布

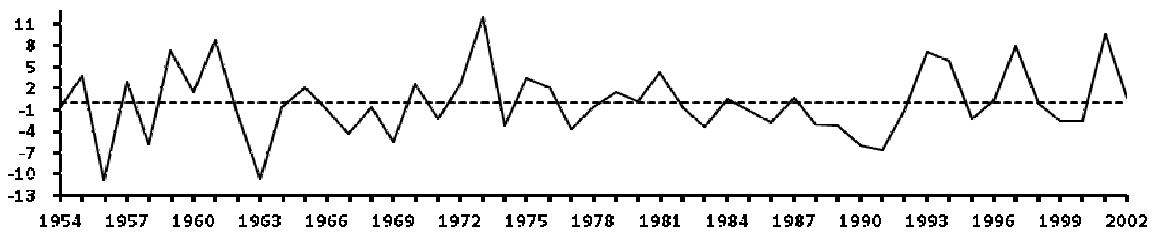


图4 广东汛期降水第1主分量变化曲线

横坐标为年份。

3.4.2 第2特征向量及主分量

图7是广东汛期降水第2特征向量分布,全省呈正和负两种相态,零值等值线从珠江口至西部的封开将广东分成两半,东至中北部为负,南至西南至雷州半岛为正,在西北部的连州还有一块正值区。显示一种东至中北部旱(涝)、南至西南至雷州半岛涝(旱)分布(II)型。当第2主分量为正(负)时出现东至中北旱(涝)、南至西南至雷州半岛涝(旱)。第2主分量变化曲线最大值出现在2002、1976和1985年,最小值出现在1961、1977和1983年。

3.4.3 第3特征向量及主分量

图8是广东汛期降水第3特征向量分布,它具有两种相态,零值等值线一条从深圳至和平,另一条从阳江西至化州北将广东分为3部分。中西北部为正,东部和西南至雷州半岛为负。它属于中西北部涝(旱),东部和西南至雷州半岛旱(涝)分布(III)型。当第3主分量为正(负)时中西北部涝(旱)而东部和西南至雷州半岛旱(涝)。第3主分量变化曲线最大值出现在1964和1998年,最小值出现在1990年。

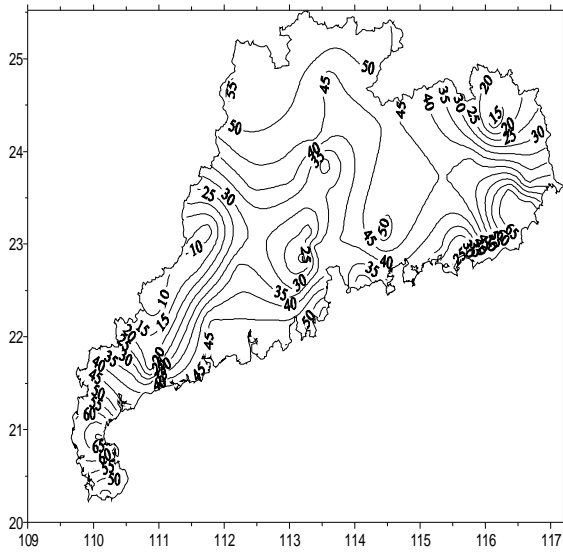


图5 1973年广东汛期降水分布

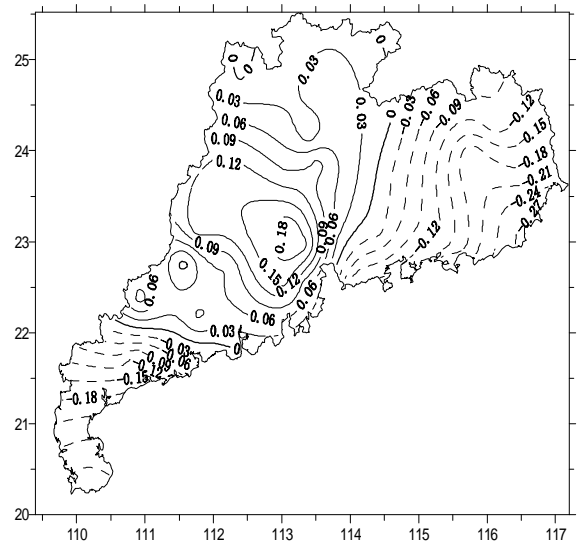


图8 广东汛期降水第3特征向量分布

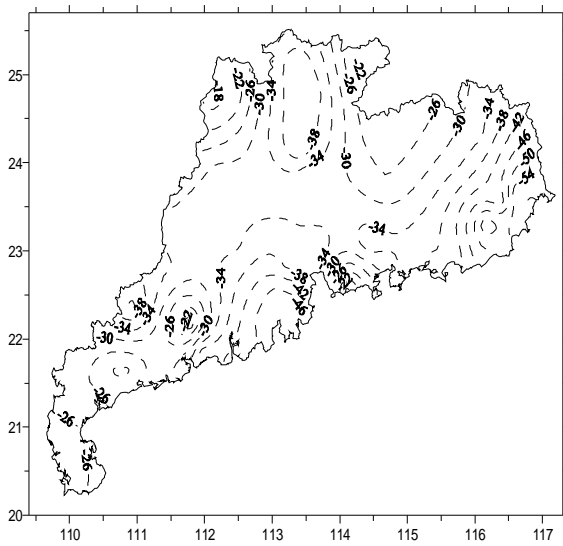


图6 1956年广东汛期降水分布

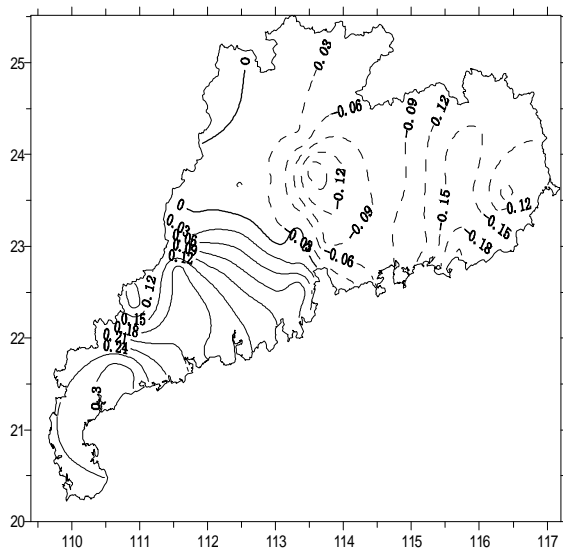


图7 广东汛期降水第2特征向量分布

3.4.4 第4特征向量及主分量

图9是广东汛期降水第4特征向量分布，可见出现两种相态，零值等值线从东北部的蕉岭经紫金至郁南的西部，将广东分成南与北两部分。北面为负、南面为正。它属于北旱（涝）、南涝（旱）分布（IV）型。当第4主分量为正（负）时，表现为北旱（涝）、南涝（旱）分布。第4主分量变化曲线1965年有一个极大值出现北旱、南涝分布，1962和1968年有极小值出现相反分布。

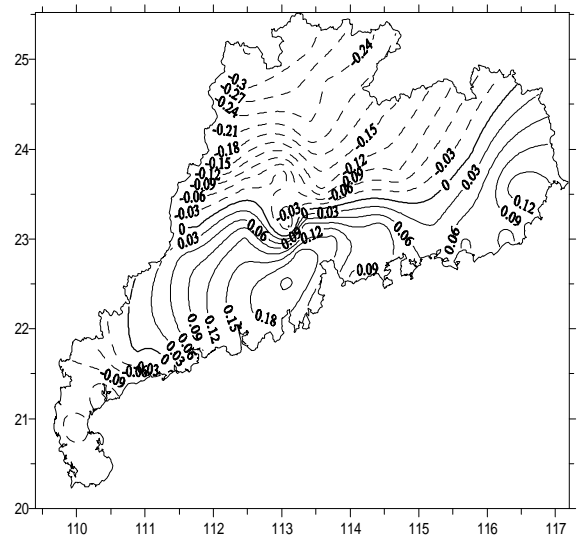


图9 广东汛期降水第4特征向量分布

3.5 广东汛期降水的分类

把广东汛期降水的历年分布看成是由该年前4个主分量的方差贡献所组成^[21-22]，分布类型就可用前4个主分量确定，因为主分量有正和负，从量上反映了特征向量随时间演变的规律。其绝对值最大

者表示该特征向量场的贡献最大, 从而定出该年和历年的分布类型分出 I、II、III、IV。将主分量的正和负分别记为 a 和 b, 得 I_a 、 I_b 、 II_a 、 II_b 、 III_a 、 III_b 、 IV_a 、 IV_b 。由于第 I 型是全广东汛期降水涝(旱)型, 它的方差贡献较大, 属于 I 型的年数也多, 但并不是所有属于 I 型年份都会有象典型的 I 型那样的分布, 为了区分那些汛期降水严重涝和严重旱的年份有必要再细分。为此定义一个数据结构函数 YC , 它是气候场第 1 主分量中前第 k 绝对最大值其累积方差占第 1 主分量总方差的 30% 以上。以 T_1 表示第 1 主分量, 若 $T_1 \geq YC$ 为 a, $T_1 \leq -YC$ 为 b,

$0 \leq T_1 < YC$ 为 c, $-YC < T_1 < 0$ 为 d, 得 I_a 、 I_b 、 I_c 、 I_d , 分别表示广东汛期降水严重涝、严重旱、偏涝、偏旱 4 种状况。本文计算的 $YC = 6.438$, 分型结果见表 2。

经统计出现 I 型的有 26 年占 53.2%, 其中 I_a 为 6 年占 12.3%, I_b 为 2 年占 4.1%, $I_a + I_b = 8$ 年占 16.4%, 这些是属于全省性严重涝和严重旱的年份, 以 1973、2001 年和 1956、1963 年为代表。 I_c 为 6 年占 12.3%, I_d 为 12 年占 24.6%, 出现 II、III、IV 型的分别有 9、6 和 8 年。这 3 个型合共占 46.8% 是属于涝和旱分布不均匀的年份。

表 2 广东汛期(4—9 月)降水分布的类型

年份	I	II	III	VI	类型	年份	I	II	III	VI	类型
1954	-0.58	-0.64	2.09	-0.25	IV_b	1979	1.51	-0.57	-0.49	-0.65	IV_a
1955	3.69	-4.2	2.75	-1.56	II_b	1980	0.23	1.72	-0.36	1.4	IV_b
1956	-10.85	0.52	1.38	0.12	I_b	1981	4.25	3	2.88	1.36	I_c
1957	2.96	-2.54	-0.49	-0.93	I_c	1982	-0.67	-0.58	-0.4	-2.03	IV_a
1958	-5.77	0.55	-1.92	0.23	I_d	1983	-3.46	-3.74	-1.78	-0.63	II_b
1959	7.34	-2.47	-0.06	0.5	I_a	1984	0.5	1.45	-1.07	1.4	IV_b
1960	1.4	-2.15	-2.66	2.73	III_b	1985	-1.11	3.76	-2.09	1.21	II_a
1961	8.66	-4.5	0.18	2.41	I_a	1986	-2.71	-0.21	-0.77	-0.14	I_d
1962	-1.39	-0.8	2.47	0.79	IV_b	1987	0.71	-0.06	2.71	-1.28	III_a
1963	-10.72	2.45	-0.31	0.31	I_b	1988	-3.07	-1.72	1.66	-0.45	I_d
1964	-0.6	-0.5	3.13	1.17	III_a	1989	-3.3	-0.44	1.65	2.9	I_d
1965	2.11	2.29	2.41	0.79	IV_a	1990	-6.01	-3.63	-5.03	0.52	I_d
1966	-0.93	-2.34	-0.2	-1.02	II_b	1991	-6.71	0.55	-0.5	-1.67	I_d
1967	-4.3	1.03	0.21	1.19	I_d	1992	-1.1	1.01	-0.99	1.04	I_d
1968	-0.67	-3.02	-1.24	-2.78	IV_b	1993	7.1	-1.68	1.66	0.21	I_a
1969	-5.46	0.04	0.3	-2.04	I_d	1994	5.76	3.39	-0.89	-1.11	I_c
1970	2.56	0.84	0.58	-1.17	I_c	1995	-2.21	1.08	-1.13	-1.34	I_d
1971	-2.19	0.87	2.36	-0.06	III_a	1996	0.33	0.13	2.82	0.77	III_a
1972	2.49	2.46	-0.67	-0.2	I_c	1997	7.96	-0.56	-1.23	-3.26	I_a
1973	11.9	0.2	-2.94	1.13	I_a	1998	-0.17	1.69	3.15	-1.05	III_a
1974	-3.3	0.97	-0.57	1.1	I_d	1999	-2.64	-1.01	-0.13	-0.33	I_d
1975	3.32	-0.8	1.76	-0.14	I_c	2000	-2.54	-2.62	-1.53	2.28	II_b
1976	2.07	4.05	-1.17	-0.12	II_a	2001	9.47	2.86	-2.23	-0.11	I_a
1977	-3.75	-4.36	-0.8	0.02	II_b	2002	0.59	5.21	-2.16	-1.68	II_a
1978	-0.7	3.06	-0.32	0.44	II_a						

4 广东汛期降水的主分量逐步回归预测

根据第 2 节预测模型数学原理的步骤对广东汛期降水进行主分量逐步回归预测。

4.1 预测因子的相关筛选

由于预测因子的数量大, 逐步回归运算有困难, 必须将气候场的前 25 个主分量逐个与所有因子进行相关筛选, 给定显著性水平 $\alpha = 0.05$, 适当调整 α 的值, 通常可选出 70 个高相关因子供逐步回归使

用。各个因子的单相关系数在 $\gamma = 0.300$ 以上, 最大达 0.545。

4.2 双重检验逐步回归

对广东汛期降水场的前 25 个主分量逐个利用相关筛选所得高相关因子进行双重检验逐步回归计算, 设定 F -信度检验^[11]的临界值 $P_a = 95\%$, 得到该场的前 25 个主分量的回归方程。下面仅对第 1 主分量及其对应的特征向量进行分析。

广东汛期降水第 1 主分量的双重检验逐步回归预测方程为,

$$\begin{aligned}
y = & -0.50386 + 0.48469x_1 + 0.24140x_2 - \\
& 0.40935x_3 - 0.24692x_4 + 0.37167x_5 - \\
& 0.24958x_6 + 0.26343x_7 + 0.20079x_8 + \\
& 0.18482x_9 - 0.19442x_{10}
\end{aligned}
\tag{15}$$

其中 x_1 为太平洋东北部7—9月海温场第5主分量; x_2 为东北半球4—6月500 hPa高度场第2主分量; x_3 为太平洋地区南部夏季500 hPa高度场第7主分量; x_4 为北太平洋1—2月海温场第4主分量; x_5 为东北半球北部10—12月500 hPa高度场第7主分量; x_6 为西北半球

冬半年500 hPa高度场第6主分量; x_7 为赤道东太平洋5—6月海温场第4主分量; x_8 为热带太平洋9—10月海温场第4主分量; x_9 为西北半球北部7—9月500 hPa高度场第7主分量; x_{10} 为全太平洋4—6月海温场第7主分量。复相关系数为 $R=0.8759$ 。图10是广东汛期降水第1主分量的预测曲线。可见预测值对实测值的拟合程度相当高。

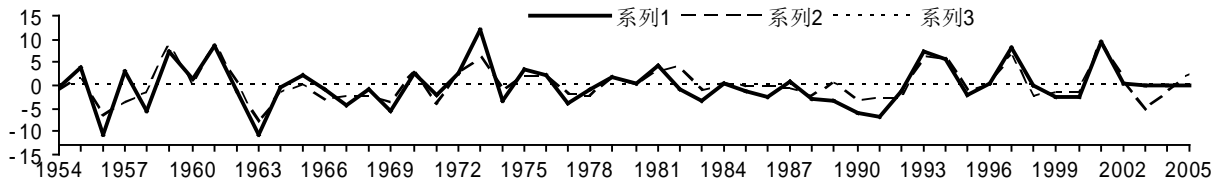


图10 广东汛期降水第1主分量预测曲线 横坐标为年份。
实线为原值,点线为平均值,短划线为预测值,2003—2005年仅有预测值。

5 广东汛期降水分布的预测

对2003年广东汛期降水分布的预测(图11)结果显示,除了雷州半岛地区外,全省都为负距平,汛期降水严重偏少10%~20%,东南部沿海达30%。从实测结果(图12)看,全省绝大部分地区均为负距平,但是在梅汕地区、海陆丰地区、珠江口西地区各有一块小范围正距平,与预测的情形相反。雷州半岛为负距平也与预测结果相反。总的来说2003年的预测基本上准确。

2004年广东汛期降水分布的预测(图13)表明,全省大部分地区为负距平,西北部和西南部的部分地区为正距平。实测结果(图14)则是,全省除了广州至东莞地区有一小片正距平外,其余地区均为

负距平。预测结果基本上准确。

由2005年广东汛期降水分布预测(图15)结果可知,广东中南部至中北部及至东北部为正距平,东南部的小片地区和西至西南部地区为负距平。西部的零值等值线从封开西南部经郁南西部、罗定、阳春到阳西。实测结果(图16)表明,大致西南部地区为负距平,其余地区为正距平,零值等值线从镇海湾经顺德沿纬线向西至罗定的北部,然后在湛江有个小正值区,在清远至佛冈有一小的负值区。因此预测准确率大于70%。

从以上预测试验例子来看,主分量逐步回归预测模型对于一个省范围的降水的预测是成功的。

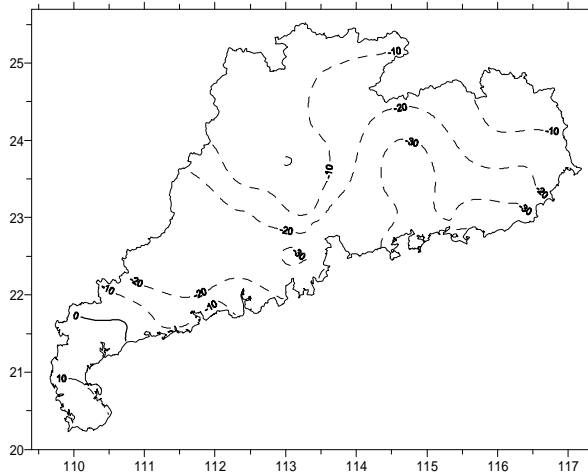


图11 2003年广东汛期预测降水距平分布

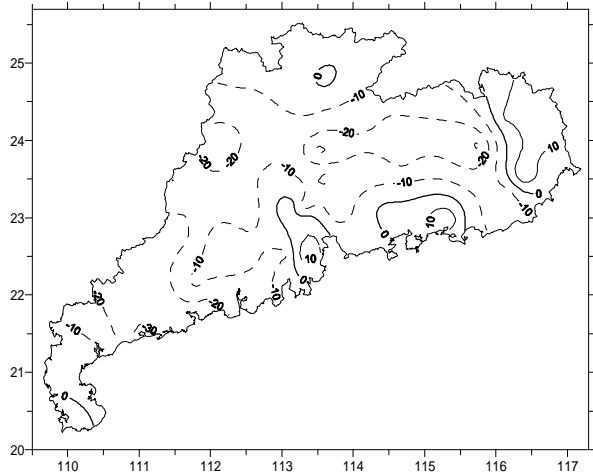


图12 2003年广东汛期实测降水距平分布

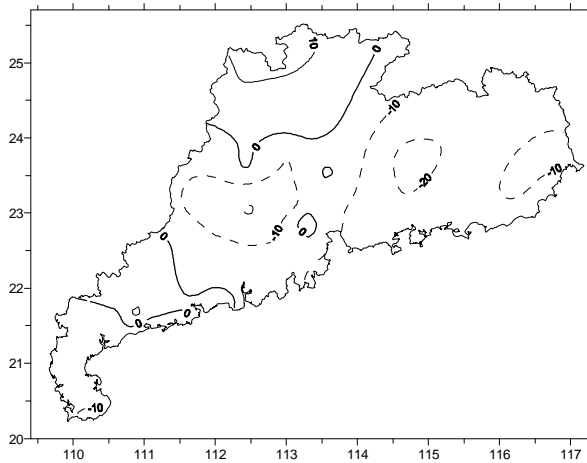


图 13 2004 年广东汛期降水预测距平分布

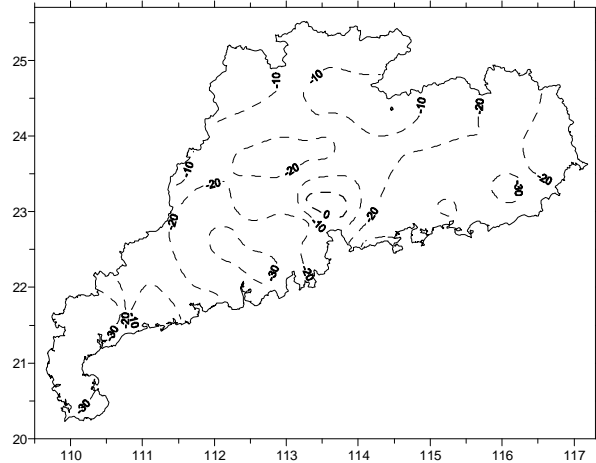


图 14 2004 年广东汛期降水实测距平分布

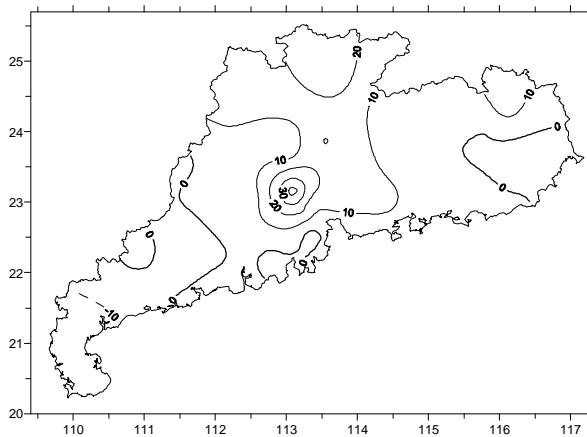


图 15 2005 年广东汛期降水预测距平分布

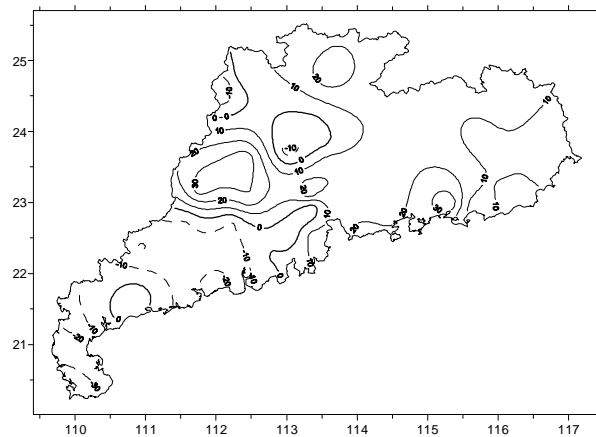


图 16 2005 年广东汛期降水实测距平分布

为了与业务预测质量检验一致, 利用李辑等^[12]短期气候预测评估方法: 按 6 级预测评分公式进行评估,

$$P_6 = \frac{N_0 + f_1 N_1 + f_2 N_2}{N + f_1 N_1 + f_2 N_2} \times 100.0 \quad (16)$$

其中 N 为测站总数, N_0 为预测与实况距平符号相同的站数与符号不同但相差只有一级的站数之和; N_1 为预测与实况同为偏级的站数与实况为特级的站数之和。 N_2 为预测与实况同为特级的站数及实况为偏级的站数之和。 $f_1 = 0.5$ 为偏级预测正确加分权重系数。 $f_2 = 1.0$ 为特级预测正确的加分权重系数。对 2003—2005 年广东汛期降水的预测结果按式(16)进行评分, 它们的得分分别为: 89.0%、94.4%和 91.3%。这个结果明显比李辑等对辽宁省近 10 年短期气候预测质量评估检验结果好^[12]。

为了进一步检验预测效果, 本文还对所预测的年份分别计算了预测场与实测场的相关系数。结果 2003—2005 年两个场的相关系数分别为: 0.342 7、

0.699 0 和 0.657 4, 它们都分别超过显著性水平为 $\alpha = 0.02$ 和 $\alpha = 0.01$ 的相关显著性检验。

6 结论和讨论

(1) 气候场的主分量逐步回归预测模型, 将气候场的预测变成对气候场的主分量的预测。

(2) 预测因子取自对气候场有重要影响的 500 hPa 高度、太平洋及印度洋海温和全球海平面气压等含不同区域不同季节的诸多因子场的前若干主分量, 可为气候场的预测提供大量的高信息量的优质因子。

(3) 利用相关筛选和双重检验逐步回归分析, 建立气候场的变化与多种因子场的变化的联系, 这样它就超越了典型相关分析、奇异值分解等方法仅考虑两个场之间相关的限制。

(4) 因为预测对象和预测因子都是主分量, 所以两者的关系反映了气候场的某种特定分布型与预

测因子场的某一特定分布型之间的关系,因此模型仍然保持着两个场相关的特点。

(5) 对广东汛期降水的预测试验,基本时空分布和变化特征清晰,预测效果显著。

(6) 关于预测年限的问题有待进一步深入研究。本文所取预测年限为3年,即提前3年起报。事实上还可做许多不同年限的预测试验,比如1年,2年,……,10年等。

(7) 关于预测因子的气候物理背景问题十分重要,有待另文分析研究。

(8) 关于预测区域的问题,目前已做过的区域有:广东省、珠江流域、中国,以一个省的范围效果最好。

(9) 关于预测要素的问题,已做过的试验有:区域性的降水、温度、水资源;单站的降水序列,温度序列等。这些都是实际业务预测最需要的内容。

参 考 文 献:

- [1] 徐瑞珍,张先恭. 经验正交函数在两个气象场相关分析中的应用[J]. 气象学报, 1982, 40(1): 117-122.
- [2] 谢炯光,秦冰冰,王静渊. 奇异值分解及其在季降水预报中的应用[J]. 气象学报, 1997, 55(1): 117-124.
- [3] 张尧庭,方开泰. 多元统计分析引论[M]. 北京:科学出版社, 1983: 302-315.
- [4] 肯德尔 M[英]. 中国科学院计算中心概率统计组译,多元分析[M]. 北京:科学出版社, 1983: 76-89.
- [5] HARNACK R P, LANZANTE J R. Specification of United States seasonal precipitation[J]. Mon Wea Rev, 1985, 113(3): 319-325.
- [6] BROCCOLI A J, HARNACK R P. Predictability of monthly North Pacific sea level pressure from monthly sea surface temperature for the period 1933-1976[J]. Mon Wea Rev, 1981, 109(10): 2107-2117.
- [7] MICHECL P J, CLUCCI S J. Modification of MOS-derived thunderstorm probabilities over complex terrain with continental scale upper air data: Preprints 10th conf on weather forecasting and analysis[J]. Amer Met Soc, 1984, 65(1): 160-164.
- [8] 金菊良,杨晓华,金保明,等. 遗传门限自回归模型在气象时间序列预测中的应用[J]. 热带气象学报, 2001, 17(4): 415-422.
- [9] 黄嘉佑. 气象统计分析与预报方法(第三版)[M]. 北京:气象出版社, 2004: 66-67.
- [10] GRAY B H. On the stability of temperature eigenvectors pattern[J]. J Climatology, 1981, 1: 273-281.
- [11] 陈创买,周文. 关于因子显著性问题的F-信度检验法[J]. 热带气象学报, 2001, 17(4): 679-682.
- [12] 李辑,金巍,赵连伟. 辽宁省近10年短期气候预测质量评估检验[J]. 气象, 2007, 33(4): 82-87.

MODEL OF PREDICTION OF PRINCIPAL COMPONENT-STEPWISE REGRESSION OF CLIMATIC FIELD AND ITS APPLICATIONS

CHEN Chuang-mai¹, ZHOU Wen², LIN Ai-lan³

(1. Department of Atmospheric Sciences, SUN YAT SEN University, Guangzhou 510275, China;

2. Depart of Physics and Materials, City University of Hong Kong, China;

3. Guangzhou institute of Tropical Marine Meteorology, CMA, Guangzhou 510080, China)

Abstract: A prediction model has been presented of principal component stepwise regression of the climate field. The model changes the prediction of the climate field to that of the principal component of that field. From the 500 hPa height and the sea surface temperature of Pacific and Indian ocean and the sea-level pressure of global in various seasons and regions, excellent principal component factors have been extracted with high information content and important influence on the climate field. Through correlation screening and double test stepwise regression, a prediction equation is developed for the principal component of the climate field, and the relation of the climate field with multiple fields of factors has been established. But the model keeps the correlation between the two element fields. According to the approximate invariability of eigenvectors of the climate field, the prediction of climate field is obtained by return computation, together with the principal component. A test example is predicting the flood period rainfall in Guangdong. The principal component and the characteristics of special-temporal distribution and its classification are computed and analyzed. The prediction of field for 2003—2005 is made and comparisons with the field of observations and errors test are established. The results show that the predictive efficacy is remarkable. Questions concerning the predicted length (years), regional coverage, physical background of elements and factors will need further research.

Key words: climatic statistical prediction; climatic element field; principal component analysis; double test step-wise regression; flood period rainfall in Guangdong