

典型相关方法在低纬高原汛期降水 预报中的应用试验^①

张万诚 郑建萌

解明恩

(云南省红河州气象局, 蒙自 661100)

(云南省气象台)

提 要

利用典型相关方法作云南汛期(6~8月)降水的预报试验,选取有实际预报意义的北半球500hPa场1~4月平均环流特征量、部分关键区的高度和距平资料及太阳黑子数作为共同影响因子,建立典型相关模型。并取所有的和前10个典型变量分别作汛期雨量预报对比。对1998年、1999年的汛期雨量预报结果表明:取前10个典型变量为预报因子作汛期雨量预报的准确率优于取所有典型场为预报因子的预测模型。从实况检验来看,该模型物理意义清晰,效果较好。并作了2000年的汛期预报,其准确率为10/16。

关键词: 典型相关 汛期雨量 预报试验

引 言

目前用于短期气候预测的统计模型大多是基于残差平方和(RSS)准则下的点对点预测模型,由于要素场水平分布的复杂性,逐点预测没有考虑相邻各地及面上各点之间的物理联系,也不具有连续性,即使把各点预报都作出来,也不能表征场的水平分布的物理联系。因而研究要素场的水平分布预报方法很有必要,云南地处低纬高原,在雨量场的分布预报中,单点的雨量观测值随机性很大,又受局地环境影响,往往缺乏代表性。因此,对雨量的预报采用场预报方法更加适宜。

典型相关方法作短期气候预测在低纬高原地区尚不多见,本文将其用作云南汛期降水预报,并作了两年预报检验和2000年的预报。从实况检验来看,该方法可以用作降水预报。

1 方法和预报原理

设某一气象要素场为预报场有 q 个空间点,因子场有 p 个空间点, n 个样本。预报量的协方差阵为

$$S_{22} = \frac{1}{n} \mathbf{Y} \mathbf{Y}' = \frac{1}{n} \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} [Y'_1, Y'_2, \dots, Y'_n],$$

因子变量的协方差阵为

$$S_{11} = \frac{1}{n} \mathbf{X} \mathbf{X}' = \frac{1}{n} \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix} [X'_1, X'_2, \dots, X'_n]$$

预报量和预报因子之间协方差阵为

$$S_{12} = \frac{1}{n} \begin{bmatrix} X_1 Y'_1, & X_1 Y'_2, & \dots, & X_1 Y'_q \\ X_2 Y'_1, & X_2 Y'_2, & \dots, & X_2 Y'_q \\ \dots, & \dots, & \dots, & \dots \\ X_p Y'_1, & X_p Y'_2, & \dots, & X_p Y'_q \end{bmatrix}$$

其中 $\mathbf{X}_k \mathbf{Y}'_j = \sum_{i=1}^n (x_{ki} - \bar{x}_k)(y_{ji} - \bar{y}_j)$, $S_{12} = S_{21}$, 将两个变量组合为 $p+q$ 个变量, $p+q$ 个变量的协方差阵为

$$\mathbf{S} = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix}$$

典型相关的基本思想是对两组变量分别

^① 本文得到国家“九五”重中之重科技攻关项目云南专题“云南短期气候预测系统的研究(96-908-05-08)”资助

作线性组合,构成新的一对变量 $U_1、V_1$,使它们之间有最大的相关系数;再分别作 $U_1、V_1$ 独立的线性组合 $U_2、V_2$,使它们之间有其次大的相关系数,如此进行下去。则 $U_i、V_i (i = 1, 2, \dots, q)$ 称为典型变量。

当求出典型变量 $U_i、V_i$ 及相关系数 r_i 后,则整个预报对象场为(具体求解见文献[1]、[3]):

$$Y_{n \times q} = S_{22} \times d_{m \times q} \times \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_q \end{pmatrix}^{1/2} \times c'_{p \times m} \times X_{n \times m} \quad (1)$$

通过典型相关场分析,可直接用整个预报因子场预测预报对象场(此时 $m = q$),也可按典型相关系数的显著性取前 m 个的典型相关场来预报整个预报对象场,本文将对两种方法作对比预报试验。

2 预报试验结果

取云南 16 个地州站 1960~1997 年汛期雨量(6~8 月雨量)为预报对象,预报因子为 1960~1997 年 1~4 月北半球 500hPa 的 1~4 月平均环流特征量、部分关键区高度和距平资料如极涡强度、印缅低槽等资料为共同影响因子,因子数共计为 16 个,按前述方法进行典型相关,可得典型相关系数、典型变量值。

应用式(1)可得 16 站预测模型为:

$$Y_{\text{昭通}} = -0.11x_1 + 0.47x_2 + 0.27x_3 + 0.47x_4 + 0.26x_5 + 0.57x_6 + 0.43x_7 + 0.57x_8 + 0.56x_9 + 0.10x_{10} + 0.34x_{11} + 0.08x_{12} + 0.04x_{13} + 0.26x_{14} - 0.03x_{15} + 0.18x_{16} + 404 \quad (2)$$

$$Y_{\text{昆明}} = -0.03x_1 - 0.27x_3 - 0.69x_4 - 0.31x_5 - 0.24x_6 - 0.08x_7 - 0.59x_8 - 0.48x_9 - 0.16x_{10} - 0.19x_{11} + 0.41x_{12} - 0.15x_{13} - 0.03x_{14} - 0.3x_{15} - 0.09x_{16} + 571 \quad (3)$$

$$\vdots \quad \dots\dots$$

$$Y_{\text{蒙自}} = -0.15x_1 + 0.08x_2 + 0.21x_3 + 0.3x_4 + 0.5x_5 + 0.1x_6 + 0.08x_7 +$$

$$0.21x_8 + 0.49x_9 + 0.22x_{10} + 0.15x_{11} - 0.32x_{12} + 0.45x_{13} + 0.16x_{14} - 0.2x_{15} + 0.26x_{16} + 446 \quad (4)$$

表 1 为应用上式(即取整个场典型相关场)作 16 站汛期雨量的预测距平百分率,评分标准为预报值与实况值距平百分率符号相同为正确,反之错误(以下同)。从表中可看出,1998 年模型预报准确率为 9/16,1999 年预测准确率为 10/16。

表 1 1998~1999 年 16 站汛期雨量实况及预报值距平百分率(%)

站名	1998 年			1999 年			拟合率
	预测	实况	评定	预测	实况	评定	
昭通	34.7	22.5	正确	-8.2	55	错误	35/38
沾益	-26.1	-8.6	正确	10.0	27.9	正确	33/38
昆明	-13.0	62.9	错误	6.1	82.7	正确	35/38
玉溪	5.9	29.0	正确	2.0	107.8	正确	35/38
楚雄	0.0	14.7	正确	4.1	77.6	正确	34/38
大理	10.6	-8.2	错误	-12.3	34.6	错误	34/38
丽江	-52.5	49.4	错误	13.2	44.9	正确	33/38
中甸	-6.4	35.1	错误	-4.9	25.8	错误	33/38
泸水	-19.0	-23.7	正确	-5.2	15.9	错误	37/38
保山	40.5	-32.9	错误	-22.3	18.7	错误	35/38
潞西	-40.6	-12.8	正确	27.4	60.5	正确	36/38
临沧	-28.6	1.5	错误	14.4	21.2	正确	34/38
思茅	20.2	12.8	正确	-14.5	19.5	错误	34/38
景洪	21.5	26.3	正确	-3.1	-14.3	正确	35/38
蒙自	-0.9	7.4	错误	1.6	28.3	正确	37/38
文山	21.6	8.3	正确	-9.4	-6.9	正确	33/38

取前 10 个典型相关作为预报因子(即上述公式中取 1~10 个变量)作 16 站汛期雨量趋势预测,表 2 为汛期雨量预测距平百分率,从表中可看出,1998 年模型预报准确率为 9/16,1999 年预测准确率为 12/16。

将表 2 的预报和实况值绘在地理底图上,除了 1998 年滇中昆明、玉溪预报报错外,预测结果分布与实况吻合很好。报错的站点基本上是滇西的个别站,在地理底图上反映出大片偏多或偏少中的孤立点,这可能与云南地处低纬高原且地形复杂,常有中小尺度局地单点大雨有关,对这些奇异点,短期气候预测是很难报的。用前 10 个典型相关建立的预报模型作 2000 年 16 站汛期雨量趋势预

测,其准确率为 10/16(表略)。

表2 1998~1999年16站汛期雨量实况及预报值距百分率(%)

站名	1998年			1999年			拟合率
	预测	实况	评定	预测	实况	评定	
昭通	30.0	22.5	正确	-1.7	55	错误	31/38
沾益	-36.3	-8.6	正确	22.6	27.9	正确	27/38
昆明	-25.0	62.9	错误	22.9	82.7	正确	26/38
玉溪	-1.8	29.0	错误	4.1	107.8	正确	36/38
楚雄	0.1	14.7	正确	14.3	77.6	正确	30/38
大理	2.7	-8.2	错误	0.3	34.6	正确	30/38
丽江	-50.9	49.4	错误	12.9	44.9	正确	34/38
中甸	-11.6	35.1	错误	-5.2	25.8	错误	35/38
泸水	-25.6	-23.7	正确	1.7	15.9	正确	33/38
保山	43.5	-32.9	错误	-23.1	18.7	错误	33/38
潞西	-41.8	-12.8	正确	25.1	60.5	正确	35/38
临沧	-28.6	1.5	错误	17.1	21.2	正确	34/38
思茅	21.8	12.8	正确	-9.9	19.5	错误	30/38
景洪	22.3	26.3	正确	-2.6	-14.3	正确	33/38
蒙自	2.5	7.4	正确	3.1	28.3	正确	32/38
文山	26.9	8.3	正确	-14.7	-6.9	正确	29/38

3 典型相关因子物理意义分析

表3为前10个特征根的方差贡献和累计方差贡献。从表中可看出,前10个特征根

表3 前10个特征根的方差贡献和累计方差贡献(%)

序号	1	2	3	4	5	6	7	8	9	10
方差贡献	91	88	83	79	73	63	61	58	40	31
累计方差贡献	12.7	25.0	36.6	47.7	57.9	66.7	75.3	83.4	89.0	93.3

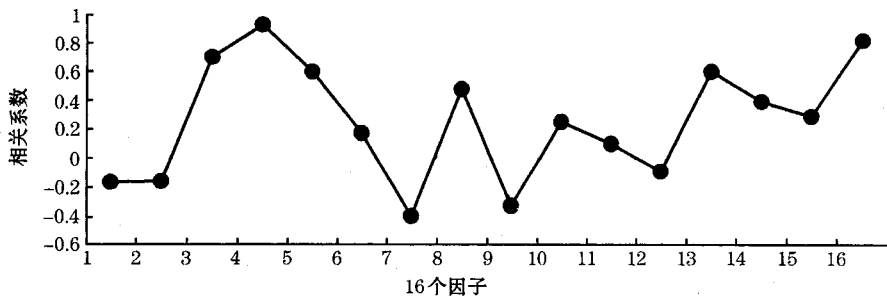


图1 云南汛期雨量与前期特征量的典型相关

4 讨论与小结

(1)本文的分析应用表明,用典型相关方法来作预报,主要考虑了整个自变量场的典型特征分布形式对整个因变量场典型特征分布形式的预报作用,通过典型相关场的分析,其预报物理意义比较清楚,本文分析中也证实了这一点。

(2)当用典型相关模型来作预报时,取前10个典型相关作为预报因子已能很好地反

的累计方差贡献已占总体方差的93.3%。对云南16个地州站汛期雨量(6~8月雨量)预报模拟检验表明,取前10个典型特征变量因子作预报模型,已具有较好的预测能力。

典型相关系数反映了两组变量之间的相关。图1为云南汛期降水与前期因子的典型相关。从图中可看出,影响云南汛期雨量的前期因子是2月贝加尔湖地区8点距平和、3月、4月阿留申10点距平和,与前期2月太阳黑子相对数呈反相关。可以说,前期特征量对汛期降水最显著的是2月贝加尔湖地区8点距平和、4月阿留申10点距平和(显著性水平为0.01),其次是4月贝加尔湖地区8点距平和、3月阿留申10点距平和、2月太阳黑子相对数(显著性水平是0.05)。因此,采用典型相关分析方法作预报,取少数典型相关因子来代替全部变量,可以排除随机因素的影响,使选取的典型因子物理意义更加清晰。

映外界因子对因变量场的作用。从对低纬高地区降水的预报表明,模型有一定的预报能力。有关更广泛的应用,有待今后作进一步的应用研究。

(3)由于雨量受地形等因素影响极大,特别是在低纬高原地区更显得突出。因此,在作雨量预报时,最好能综合考虑外界因子场的共同作用。

(下转封二)

参考文献

1 施能. 气象科研与预报中的多元分析方法. 北京: 气象出版社, 1995, 89~91.

- 2 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, 9~90.
- 3 魏凤英. 现代气候统计诊断预测技术. 北京: 气象出版社, 1999, 159~168.

Application of Canonical Correlation Method to Forecasting Rainfall during Flood Season in the Plateau of Low Latitude

Zhang Wancheng Zheng Jianmeng

(Honghe Meteorological Office, Mengzi 661100)

Xie Ming'en

(Yunan Meteorological Observatory)

Abstract

By using canonical correlation method, the forecasting experiment of the rainfall during the flood season in Yunan Province is performed. The mean circulation field of 500hPa, its height anomalies in some key regions and sunspot number from January to April during 1960—1997 are selected as the common influence factors, and the canonical correlation models, which include the whole factors and the first ten factors, are built and used to predict the rainfall during the flood season. The results show that the model with the first ten canonical variables is better than that with the whole factors for 1998 and 1999, respectively. The accuracy of prediction in 2000 is 10/16.

Key Words: canonical correlation rainfall of flood season forecasting experiment