第19卷 第2期 大 气 科 学 Vol. 19, No. 2
1995 年 3 月 SCIENTIA ATMOSPHERICA SINICA March 1995

用典型相关分析作副高的统计动力预报模式可预报性研究

黄嘉佑

(北京大学地球物理系，北京 100871)

提 要

本文提出一个关于西太平洋副热带高压（简称副高）的统计动力预报模式，利用它和典型相关分析方法对冬、春和夏季逐月副高预报的可行性进行研究。结果表明，模式的可预报性依赖于预报量场和因子场所提取的分量数，模式的差分形式及预报落后步长。对逐月和不同步长所作的可预报性分析发现步长为 1 个月有较高的可预报性，不同月份可预报性有所不同，一般夏季较冬季和春季要差。虽然如此，用该模式作夏季副高预报还是具有一定的可能性。在独立样本中所作的预报试验表明，月际预报符号相关系数一般均接近或超过 0.60。

关键词：副高；统计动力预报；典型相关分析；可预报性。

一、前 言

副高是我国夏季降水的重要天气系统之一，它的位置和强弱的变化对我国夏季雨带分布有明显的影响。不少研究指出夏季 500hPa 西北太平洋副高的位置和强度变化，与太平洋地区的海平面气压、东亚环流形势和极地环流有密切关系[1]，对我国夏季降水分布有举足轻重的作用[2]。对副高位置及强度的准确预报直接关系我国夏季降水分布形势预测的准确性。目前对副高的位置及强度的描述常用的指数分别有西伸脊点、北界、面积和强度等。它们能反映副高整个天气系统的变化特征，常为我国气象工作者作为研究和预报的对象。但是，过去的研究仅单独分析其中某一指数的变化规律或者分析它与其他气象要素的关系。我们知道上述 4 个指数是有机联系的，它们的综合才能反映副高的变化特征。因此本文将着重研究副高，即由上述 4 个指数所构成的整体的统计动力预报模式可预报性。

二、副高的统计动力预报模式

由 Hasselmann[3]提出的随机气候模式就是从气候系统内不同时间尺度过程的相互作用角度出发建立起来的一种理论模式。他把大气长时期缓慢的气候状态变化和短期的天气变化统一考虑在一个预报方程中，若把气候状态的变化记为 y，短期天气变化作为

1992 年 7 月 15 日收到，10 月 15 日收到修改稿。

随机外力 f，建立关于预报 y 的随机动力气候模式：

$$\mathrm{d}y/\mathrm{d}t = -\lambda y + f, \tag{1}$$

其中，$\lambda(>0)$为气候系统的反馈系数。如果把气候状态的变量 y 推广为描述整个系统的向量 y，则上式可写为

$$\mathrm{d}y/\mathrm{d}t = -Cy + f, \tag{2}$$

其中，C 为反馈系数矩阵，f 为随机外力向量。对副高气候系统可看成为遵从这样规律的气候系统。由于副高由上述 4 个指数所描述，即它包含 4 个分量，分别记为 y_1、y_2、y_3 和 y_4。

若令 $b_{ij} = -c_{ij}(i,j=1,4)$，并进一步把微分方程改为差分方程则有

$$\begin{cases} \dfrac{y_1(t+\Delta t)-y_1(t)}{\Delta t} = b_{11}y_1(t)+b_{12}y_2(t)+b_{13}y_3(t)+b_{14}y_4(t)+f_1, \\ \dfrac{y_2(t+\Delta t)-y_2(t)}{\Delta t} = b_{21}y_1(t)+b_{22}y_2(t)+b_{23}y_3(t)+b_{24}y_4(t)+f_2, \\ \dfrac{y_3(t+\Delta t)-y_3(t)}{\Delta t} = b_{31}y_1(t)+b_{32}y_2(t)+b_{33}y_3(t)+b_{34}y_4(t)+f_3, \\ \dfrac{y_4(t+\Delta t)-y_4(t)}{\Delta t} = b_{41}y_1(t)+b_{42}y_2(t)+b_{43}y_3(t)+b_{44}y_4(t)+f_4, \end{cases} \tag{3}$$

式中 Δt 为差分步长。记上式左边副高差分分量组成为预报量向量 y，副高某月份 4 个指数组合为因子向量 x，系数组合成系数矩阵 B，则有预报矩阵方程

$$y = Bx + f. \tag{4}$$

若作适当变化，模式(3)也可写作

$$\begin{cases} y_1(t+\Delta t) = (1+b_{11}\Delta t)y_1(t)+b_{12}\Delta t y_2(t)+b_{13}\Delta t y_3(t)+b_{14}\Delta t y_4(t)+f_1, \\ y_2(t+\Delta t) = (1+b_{21}\Delta t)y_1(t)+b_{22}\Delta t y_2(t)+b_{23}\Delta t y_3(t)+b_{24}\Delta t y_4(t)+f_2, \\ y_3(t+\Delta t) = (1+b_{31}\Delta t)y_1(t)+b_{32}\Delta t y_2(t)+b_{33}\Delta t y_3(t)+b_{34}\Delta t y_4(t)+f_3, \\ y_4(t+\Delta t) = (1+b_{41}\Delta t)y_1(t)+b_{42}\Delta t y_2(t)+b_{43}\Delta t y_3(t)+b_{44}\Delta t y_4(t)+f_4. \end{cases} \tag{5}$$

把上式左边下 Δt 月副高指数作为预报量向量记为 y，右边系数列表记为矩阵 B，当月副高指数资料向量记为 x。则同样有形如模式(4)的统计预报模式。动力模式(2)中的反馈矩阵的估计可以通过实际资料对式(4)的矩阵 B 来实现。为讨论方便起见，无论预报量还是因子变量均作标准化处理。本文将主要探讨上述两种转化为统计模式的动力模式的可预报性。

三、资料与方法

在上述统计动力模式基础上，选取择某月月平均 4 个副高指数资料，构成统计动力模式(3)或式(5)中的预报量矩阵 $Y(q\times n)$及因子矩阵 $X(p\times n)$。其中 q 为预报量变量个数，p 为因子个数，并设 $q<p$，n 为样本容量。对它们均可作主分量分析[4]，即

$$Y = A_y F_y, \qquad X = A_x F_x, \tag{6}$$

式中 $A_y(q\times k_y)$、$F_y(k_y\times n)$和 $A_x(p\times k_x)$、$F_x(k_y\times n)$分别为关于预报场和因子场的

因子荷载及标准化主分量（即主因子）。由于主分量能代表场的主要特征，因此研究两个场的主分量之间的关系可反映场的相关关系。进一步对各个场的主分量作线性组合生成新的变量，由它们构成相应的新变量矩阵，表示为

$$V = PF_y, \qquad U = QF_x, \tag{7}$$

其中$V(k_y \times n)$由预报量构成，$U(k_x \times n)$由因子变量构成的新变量矩阵。若各个场的新变量为标准化变量且相互独立，则称此种新变量为典型因子，其相关系数称为典型相关系数。

由典型相关系数构成的对角阵可表示为

$$R = (1/n)VU' = P[(1/n)F_y F_x']Q'. \tag{8}$$

典型因子与原变量的关系为

$$Y = DV, \qquad X = CU, \tag{9}$$

$C(p \times k_x)$和$D(q \times k_y)$为典型因子的荷载阵，它们可用下式求出：

$$D = A_y P', \qquad C = A_x Q', \tag{10}$$

为进一步求回归方程(4)式的系数矩阵，可导出

$$B = DRC'(CC')^{-1}. \tag{11}$$

由于对预报场和因子场提取的主分量数是可以选择的，因此用这种方法作的典型相关与传统典型相关分析有所不同，后者是考查两个变量场全部的协方差结构，从中提取它们的典型相关，当前者取两个变量场所有的主分量时，两者结果是一致的。但当提取分量数少于变量数时，结果会有差别。由不同分量数所作的典型相关分析可定出预报量向量与因子向量关系的最佳线性组合的系数矩阵，并利用它们可作出预报效果的估计。

利用回归方程作预报值估计，把它们与实测值进行比较作为可预报性估计。我们选取4个统计量描述其可预报性。其一为依赖样本中的残差方差

$$Q_1 = [\sum\sum(y_{ij} - \hat{y}_{ij})^2]/(q \times n), \tag{12}$$

其中y_{ij}为第i个预报量第j个样品值。其二用刀切法计算依赖样本中去掉第i个样品（独立样品）中的残差方差[4]，即

$$Q_2 = \frac{[\sum\sum(y_{ij} - \hat{y}_{ij})^2]/(1 - h_{ii})}{(q \times n)}, \tag{13}$$

其中h_{ii}为帽子矩阵中第i行第i列元素。在典型相关中表示为

$$h_{ii} = x(i)'(nCC')^{-1}x(i), \tag{14}$$

其中$x(i)$为第i个因子样品向量。为度量预报值与实测值距平符号的一致性，用距平符号相关系数RS及预报与实测距平相关列联表的检验值KF作为第三和第四个统计量来描述可预报性，它们表示为

$$RS = \frac{N_{11} + N_{22}}{n}, \tag{15}$$

$$KF = \frac{n(N_{11} \times N_{22} - N_{12} \times N_{21})}{(N_{11} + N_{12})(N_{11} + N_{21})(N_{22} + N_{12})(N_{22} + N_{21})}. \tag{16}$$

式中N_{11}和N_{22}分别为预报与实测标准化正和负距平符号相同的次数，N_{12}和N_{21}分别为距平符号正与负和负与正相反的次数。为比较不同样本的可预报性，本文以1951—1987年1—8月为依赖样本（简记为*YL*），以1988—1990年相应各月为独立样本（简记为*DL*），在不同样本中计算上述统计量。

四、两种差分模式的比较

利用当月副高指数为因子，下一个月与当月的副高指数的差分作预报量（模式(3)，简记为M1）和以下月指数为预报量（模式(5)，简记为M2）作预报并进行逐月(1—7月）试验，由于典型相关预报依赖于所取的原变量场主分量个数的选取。表1给出两个模式在$k_y = k_x = k = 2$、3和4时最大的典型相关系数的比较。

从表中可见，一般而言，模式所取的分量数越多，所提取的最大典型相关系数越大。总的比较结果是模式M2比模式M1所提取的典型相关系数大。但是是否取的分量数越多，模式可预报性越大呢？表2给出1—7月各月可预报性4个统计量的平均值的比较。

表1 两种差分模式的最大典型相关系数比较

月份	1	2	3	4	5	6	7
M12	0.384	0.579	0.570	0.526	0.496	0.749	0.646
M22	0.668	0.789	0.813	0.732	0.714	0.738	0.443
M13	0.712	0.795	0.725	0.737	0.869	0.754	0.722
M23	0.726	0.802	0.826	0.775	0.717	0.840	0.606
M14	0.762	0.832	0.761	0.808	0.908	0.779	0.750
M24	0.834	0.837	0.835	0.786	0.742	0.859	0.633

(*)模式最后的数字表示所选分量数

表2 1—7月各月可预报性4个指标的平均值的比较（依赖样本）

	M1			M2		
k	2	3	4	2	3	4
Q_1	0.859	0.854	0.894	0.675	0.659	0.688
Q_2	0.999	0.990	1.047	0.778	0.750	0.786
RS	0.631	0.600	0.599	0.688	0.707	0.706
KF	2.727	1.840	1.446	6.304	8.075	8.396

从表中可见，对模式M1而言，分量数取$k=2$有较好的可预报性，它有最大的符号相关系数和较小的拟合残差方差和预报残差方差，*KF*值也较大。对模式M2而言，同样理由，分量数取$k=3$有较好的可预报性。可见并不是分量数取得越多越好。当选取分量数$k=4$时相当于多预报量的多元回归模式，这种模式却没有最好的预报效果。模式间比较而言，模式M2比M1有较高的可预报性。值得指出的是，在单个独立样品的预报中，模式M2有最好的可预报性（见表中Q_2值的比较），而且*KF*值大大超过5%显著性水平。

为进一步检验模式在多样品独立样本中的可预报性，利用在逐月依赖样本中由模式所建立的回归方程作独立样本（1988—1990 年）的预报。表 3 给出类似的比较。

表 3　1—7 月各月可预报性统计量的平均值的比较（独立样本）

	M1			M2		
k	2	3	4	2	3	4
Q_1	1.044	1.157	1.120	0.875	0.840	0.869
RS	0.583	0.512	0.500	0.703	0.726	0.655

检验表明，在独立样本中，两个模式的可预报性比依赖样本均有不同程度的下降。但两个模式比较而言，比较 M1 模式，模式 M2 仍有较好的可预报性，而且模式 M2 除有较高的符号相关系数外还有小于 1.0 的残差方差。因为所检验的预报量均为标准化变量，其方差为 1.0，若残差方差大于它，则表明误差太大。比较进一步表明在 M2 模式中分量数取 $k=3$ 仍有最好的效果。

上述试验表明，无论在依赖样本还是独立样本，取分量数为 3 的 M2 模式均有较好的可预报性。因此在检验模式逐月的可预报性仅给出该模式在依赖样本和独立样本中符号相关系数和残差方差在各月的比较（见图 1）。

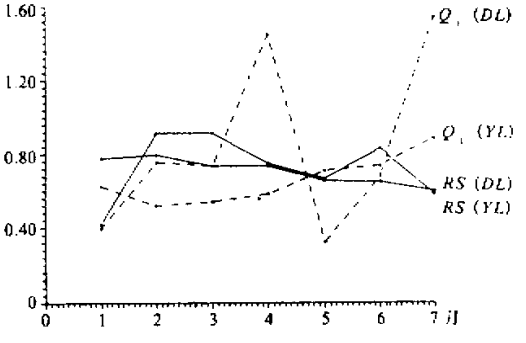

图 1　各月符号相关系数（实线）和残差方差（虚线）的比较

从图可见，在依赖样本中冬季和春季（1—4 月）有较好的可预报性，在初始月为 5、6 和 7 月时可预报性有所下降，但符号相关仍能超过 0.60。在独立样本中可预报性表现不太稳定，2 和 3 月及 6 月有较高的符号相关系数和较小的残差方差，但在强烈变化的月份，如 4—5 及 7—8 月则效果较差。

五、不同步长差分模式的比较

从模式 (5) 可见，当取不同步长时会有不同的预报模式。上一节的讨论实际是取步长为 1 个月的情况。由于试验结果表明，取分量数为 3 的 M2 模式有较好的可预报性。因此这里仅比较该模式在不同步长下的试验情况。步长不同，可预报试验月数不同，步长为一个月时可试验月数为 7，步长为 2 时有 5 个试验月数，步长为 4 时仅有 3 个试验月数（即夏季 6—8 月）。表 4 给出在依赖样本及独立样本中可预报性各统计量所有月份平均值的比较（因为在独立样本中无刀切法，故无Q_2值）。

从表中可见，随步长增加，可预报性减少。在依赖样本中步长为 1 个月有最好的可预报性，在独立样本中表现不太稳定，步长为 1 个月时有最高的符号相关系数，但步长为 4 个月时则有最小的残差方差，总的表现似乎仍是步长为 1 个月时较好。表 5 给出起

表4 分量数为3的M2模式不同步长的可预报性

步长（月）	1		2		3		4	
样本	*YL*	*DL*	*YL*	*DL*	*YL*	*DL*	*YL*	*DL*
Q_1	0.659	0.840	0.736	1.121	0.796	0.987	0.865	0.803
RS	0.707	0.726	0.669	0.667	0.643	0.616	0.627	0.584

表5 不同步长的最大典型相关系数的比较

步长（月）		1	2	3	4
起始月	1	0.612	0.640	0.612	0.533
	2	0.802	0.768	0.801	0.642
	3	0.826	0.816	0.545	0.468
	4	0.775	0.639	0.688	0.483

始月为1—4月时不同步长的最大典型相关系数的比较。从表中可见，在每个起始月中步长为1个月时差不多均有较高的典型相关系数，然后随步长增大典型相关系数逐步下降，这一现象以3月为起始月表现最为明显。

六、预报试验

由于步长为1个月的分量数为3的M2模式有较好的可预报性，我们使用该模式作夏季副高预报试验。其方法除起始场外其余各月均用模式作预报，并用预报场作为下月的起始场。在逐月预报中使用不同月份的典型相关预报模式。表6给出以不同月份为起始月（即因子场）所作的逐月副高场预报可预报性统计量平均值的比较。从表中可见，以2月为起始月所作的预报有较为好的效果（在依赖样本中有最低的残差方差、最高的符号相关系数和最大的*KF*值，在独立样本中有最高的符号相关系数和最大的*KF*值）。

表6 逐月副高场预报可预报性统计量平均值的比较

起始月份	1		2		3		4	
样本	*YL*	*DL*	*YL*	*DL*	*YL*	*DL*	*YL*	*DL*
Q_1	0.861	0.920	0.796	0.986	0.805	0.923	0.823	1.237
RS	0.655	0.631	0.664	0.722	0.643	0.683	0.647	0.625
KF	4.774	0.390	5.758	0.777	4.574	0.576	3.682	0.273

尽管用典型相关回归作不同起始月的逐月预报基本上是利用副高场的持续性作的，但是它还不同于纯持续性预报，如果纯持续性预报是把起始场的值不变地作为预报场的值。以此法作的预报其残差方差平均值均大于1.0，可见用典型相关回归作的逐月预报比纯持续性预报要好，且具有一定的可预报性。

七、结论与讨论

本文根据Hasselmann的随机气候模式基础上提出一个关于西太平洋副热带高压

(简称副高) 的统计动力预报模式，动力模式通过不同的差分形式可转变为统计回归模式 (M1 和 M2)。利用典型相关分析方法对它作夏季 (6—8 月) 副高预报的可行性进行研究。结果表明，模式的可预报性依赖于预报量场和因子场所提取的分量数，当分量数取 3 时有较好的可预报性。检验表明，在独立样本中，两个模式的可预报性比依赖样本均有不同程度的下降。但两个模式比较而言，模式 M2 有较好的可预报性。用 M2 模式对逐月和不同步长所作的可预报性分析发现步长为 1 个月有较高的可预报性，不同月份可预报性有所不同，一般夏季较冬季和春季要差。虽然如此，用该模式作夏季副高预报还是具有一定的可能性。在独立样本中所作的预报试验表明，符号相关系数一般均接近或超过 0.60。由于所使用的预报模式是利用副高的持续性作的，对夏季的某些月份 (如 6—7 月) 预报效果并不十分理想，看来预报的改进还有赖于动力模式本身的改进，例如在模式中考虑海温的影响，或建立海气耦合随机模式。这方面的研究还有待进一步的探索。

参 考 文 献

[1] 廖荣孙，赵振国，1990，东亚阻塞形势与西太平洋副高的关系及其对我国的影响，长期天气预报论文集，气象出版社。

[2] 陈兴芳，杨义文，1990，北半球绕极环流的不对称性及其与西太平洋副高的关系，长期天气预报论文集，气象出版社。

[3] Hasselmann, K., 1976, Stochastic climate models, *Tellus*, **28**, 473–485, 1976.

[4] 黄嘉佑，1990，气象统计分析与预报方法，气象出版社，387pp。

A Study of Predictability of Statistical–Dynamic Model for Subhigh by Using Canonical Correlation Analysis

Huang Jiayou

(Department of Geophysics, Peking University, Beijing 100871)

Abstract

A simple statistical–dynamic model for forecasting the subtropic high pressure (subhigh) is presented in this paper. The predictability for its forecasting in winter, spring and summer is studied by the model using canonical correlation anaiysis. The results show that the predictability depends on the number of the principal components, which are extracted from the predictand and predictor fields using principal component analysis, the difference form and forecasting step of lag months. It is found, from the experiments in different step of lag months, that the highest predictability presents with the step of lag of one month. The predictability is different in various seasons. In general, the predictability in winter and spring are better than summer. Nevertheless, the experimental results show that it is possible for subhigh forecasting. The anomalous sign correlation coeffients between observation and forecasting can reach or surpass 0.60 for most months.

Key words: subhigh; statistical–dynamic model; canonical correlation analysis; predictability.