

典型相关分析(CCA)对我国冬季 气温的短期气候预测试验*

毛恒青 李小泉

(国家气象中心, 北京 100081)

提 要

利用典型相关分析(CCA)方法建立统计气候预测模型, 对我国冬季气温进行了预测试验, 采用历史资料独立样本检验的方法, 对预报技巧给出合理的评定. 结果表明, 使用 CCA 方法对我国冬季气温进行短期气候预测, 有一定的预报技巧, 对于特定地区和特定时期优选的因子场组合, 可以取得较为满意的预报效果. 大部分地区的季平均预报时效在 2 个季以内时, 最佳预报相关系数在 0.5 以上. 季平均的预报水平明显高于月平均的预报.

海温场是所有因子场中最好的预报因子. 不仅单独海温场的预报效果较好, 而且与其它因子场组合后的预报水平还可以得到进一步提高.

关键词: 典型相关分析(CCA) 预报试验 相关系数

引 言

典型相关分析(CCA)是近年来发展起来的一种新的统计方法. 它是双重的主分量分析^[1], 是在经验正交函数(EOF)分析的基础上发展起来的一种多变量统计技术. 在交叉相关极大的条件约束下, 找出两个场之间最高相关的典型分布型的联系, 以解析最多的预报量方差, 而典型因子相互之间又都是独立的. 它的优越之处在于利用了整个场的信息, 客观定义了因子场与预报量场高相关的类型. 利用主要典型分布型作预报方程的回归, 提取预报场与因子场之间的线性相关信息中的主要部分, 能排除一些随机因素的影响, 可以取得较好的预报效果. 并且 CCA 过程中的本征值是典型相关系数的变形, 反映出典型分布型之间相关性的大小, 而本征值的总和则可以反映出模式的预报水平.

Graham^{[1][2]}、Barnett^[3]、Barnston^{[4][5]}等先后用 CCA 建立模式, 应用在短期气候预测中; 美国国家气候预报中心(CPC/NCEP)已将以 CCA 为基础的统计预报模式应用于长期预报业务^[6].

本文利用 CCA 的方法, 结合 EOF 展开和物理因子分析, 建立统计气候预测模型, 进行我国短期气候预测试验. 并研究 CCA 的性能及各因子预报性能的气候特征.

* 中国气象局“八五”重点科研项目《长期天气预报理论与方法的研究》课题资助。
1996-07-05 收到, 1996-09-02 收到修改稿。

1 方法原理及试验概况

1.1 原理

典型因子分析是分别对两组变量进行线性组合,在典型相关极大的条件约束下,找出最佳的线性组合,而这种反映两组变量最佳线性关系的新变量称作典型因子.设因子场为 X_1, X_2, \dots, X_p , 共 p 个空间点, 预报场为 Y_1, Y_2, \dots, Y_q , 共 q 个空间点, 样本容量为 n , 并分别进行距平化. 资料阵记为:

$$X = (X_1, X_2, \dots, X_p)' \quad (1)$$

$$Y = (Y_1, Y_2, \dots, Y_q)' \quad (2)$$

由因子场与预报场分别进行线性组合, 构成两组新的变量:

$$U = C'X \quad (3)$$

$$V = D'Y \quad (4)$$

其中 U, V 称作典型因子. 要求它们有以下特性:

(1) 标准化变量, 即平均值为零, 方差为 1:

$$\frac{1}{n}U U' = \frac{1}{n}(C'X)(C'X)' = C'(\frac{1}{n}XX')C = C'S_{11}C = I \quad (5)$$

$$\frac{1}{n}V V' = \frac{1}{n}(D'Y)(D'Y)' = D'(\frac{1}{n}YY')D = D'S_{22}D = I \quad (6)$$

I 为单位矩阵.

(2) 它们之间的协方差即相关系数在两组变量所有线性组合而成的新变量中是逐次取极大的:

$$R = \frac{1}{n}U V' = \frac{1}{n}(C'X)(D'Y)' = C'S_{12}D \rightarrow \text{极大} \quad (7)$$

在式(5)、(6)和(7)的条件下求二次型的条件极值, 经过变换可以导出以下特征值及特征向量方程组的矩阵表达形式:

$$(S_{22}^{-1}S_{21}S_{11}^{-1}S_{12} - \lambda)D = 0 \quad (8)$$

$$(S_{11}^{-1}S_{12}S_{22}^{-1}S_{21} - \lambda)C = 0 \quad (9)$$

求解式(8)和(9)得出特征值 $\lambda_i (i=1, 2, \dots, q)$ 及对应的特征向量构成的矩阵 C, D . 其中式(5)~(9)中的 $S_{11}, S_{12}, S_{22}, S_{21}$ 分别为 X, Y 的协方差阵和交叉协方差阵, 典型相关系数 $r_i (i=1, 2, \dots, q)$ 与特征值之间有关系式:

$$r_i = \sqrt{\lambda_i} \quad (10)$$

这样求得的典型因子首先是标准化变量, 同时又具有相互独立性及逐次极大相关性的特点.

利用典型因子 U, V 进行回归, 令 Λ 是特征值为对角线元素、其余元素均为零的矩阵, 预报方程的矩阵形式为:

$$\hat{Y} = S_{22}D \sqrt{\Lambda}C'X \quad (11)$$

1.2 预报场及预报因子

气候预测主要是大范围的趋势性预测, 本试验选取我国冬季气温作为预报对象, 其中气温场用 8 个区的区域平均气温等级表示. 它们分别为: 东北区①、华北区②、长江中下游区③、华南区④、西南区⑤、西北区⑥、新疆区⑦和全国的平均⑧.

用国家气候中心业务上应用的太平洋 10 个关键区的海温场作为预报因子. 它们分别为: 亲潮区、黑潮区、西风漂流区、秘鲁洋流区、Niño1+2 区、Niño3 区、Niño3* 区、Niño4 区、北大洋锋区和南大洋锋区.

大气环流的长期变化是大气、海洋、极冰等相互作用的结果, 也是气候异常与变化的反映. 同样, 大气环流的变化及异常, 也直接影响着气候的变化. 通过大量的统计分析, 选用北半球极涡场的面积和强度指数作为因子场之一.

1.3 模式的建立

在进行 CCA 过程之前, 首先将资料进行距平化和标准化. 对标准化后的因子场和预报场, 分别进行 EOF 展开, 展开后的空间函数称作典型分布型, 时间函数称之为典型主分量. 根据特征值的大小, 对主分量进行截断. 一方面压缩资料中不必要的“噪音”部分, 缩小维数, 减少计算量. 同时避免 CCA 过程中可能会出现逆矩阵的蜕化现象.

分别将截断后的典型主分量作为新因子场和预报场, 进行 CCA 的计算, 并拟合预报方程. 再利用典型分布型, 将计算所得的预报方程还原为标准化变量方程(运算过程图略).

1.4 试验概况

为对比分析, 将各要素场进行 6 种不同的因子场组合, 分别为气温、海温、极涡、气温和海温、气温和极涡以及综合(气温、海温和极涡)因子场. 对 6 种因子场组合, 分别取其前期的连续因子场和挑选因子场进行预报试验.

预报场为 1952~1993 年每年的冬季, 因子场为相应各冬季前期 7 个季的各要素场. 样本大小为 42 年. 预报时效为 0~5 个季, 所用的因子场的数目最大为 4 个季. 冬季各月的预报, 因子场分别取相应于各月前期 11 个月的各要素场. 样本数为 43 年.

对预报结果, 采用历史资料独立样本检验(Cross-Validation)的方法进行评估. 即从试验资料序列的第一个样本开始, 每次轮流留出一个样本, 用余下的样本建立预报方程, 并对留出的样本作预报检验. 这样依次进行, 直到全部样本都作为独立的预报检验完毕. 这种方法虽然工作量大, 但充分利用了现有的历史资料序列, 能够对预报技巧给出一个较为真实的无偏估计. 我们计算独立样本检验的预报值与实况的相关系数, 以平均相关系数值作为评价结果好坏的标准.

2 预报试验

2.1 季平均预报

2.1.1 不同连续因子场的预报结果 图 1a、1b、1c、1d 分别是各种长度的连续气温、海温、极涡和综合因子场的平均预报相关系数随预报时效的变化.

* Niño3: 5°S~5°N, 90°~150°W; Niño3*: 0~10°S, 85°W~180°.

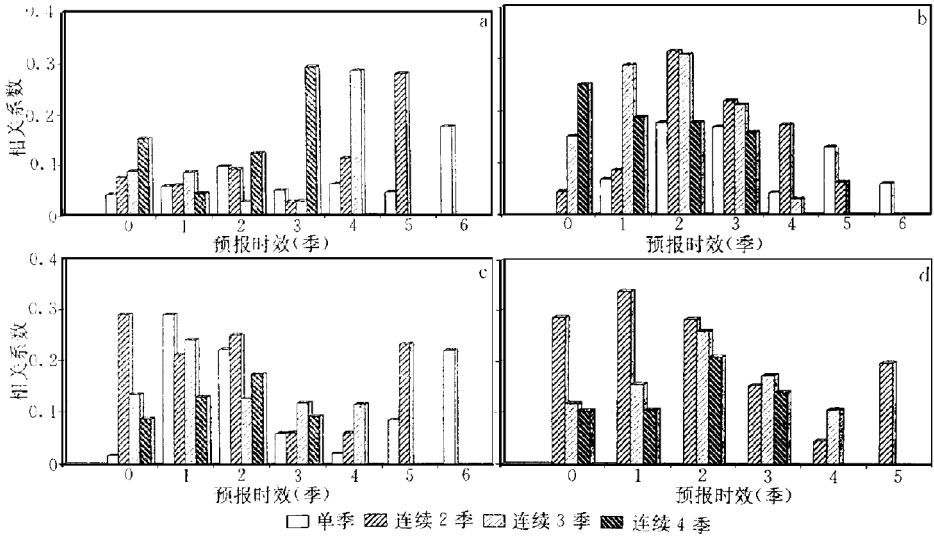


图1 各种长度连续因子场(a)气温 (b)海温 (c)极涡 (d)综合对冬季平均气温的预报相关系数随时效的变化

Fig.1 Correlation coefficient between forecasts of various consecutive predictor fields (a) temperature (b) SST (c) polar vortex (d) synthetic and observed mean temperature in winter with different lead time

由图 1a 可知：单季气温因子场以时效为 6 季(即前春)时预报技巧最高，其它连续因子场则均以包括该季(前春季)要素场组成的因子场预报效果最好。而近期的气温因子场预报技巧并不高，预报时效在 3 季以上时预报效果较好，即前一年的气温场对冬季气温有较高的预报技巧。反映出气温场本身随时间的演变变化规律。海温因子场时效为 2 季时预报结果最好(见图 1b)，其次是时效为 1、3 季时，由于单季因子场在时效为 2、3 季时预报相关系数最大，从而使得包括这两个季的连续因子场效果最好，即所用因子场为半年前左右的海温场时预报效果最好，与海洋和大气之间存在着半年左右的韵律规律相

表 1 两次 El Niño 过程中海温因子场的预报效果

Table 1 Forecast effects using SST predictor field during two El Niño processes

年	时效(季)	1 区	2 区	3 区	4 区	5 区	6 区	7 区	8 区
1969~1970	实况	3.63	3.27	3.20	3.37	3.43	3.30	2.23	3.23
	0	5.55*	4.68*	4.23*	4.48*	3.66*	5.15*	5.58	4.70*
	1	3.42*	3.91*	4.78*	5.91*	4.69*	4.43*	3.52	4.36*
	2	2.96	3.29*	3.20*	4.19*	3.79*	2.97	1.93*	3.22*
1987~1988	实况	2.00	2.37	3.20	3.07	2.97	2.40	2.37	2.63
	0	1.85*	2.03*	2.75	2.86	2.61*	2.48*	2.54*	2.40*
	1	2.69*	1.39*	1.35	1.40	2.14*	1.71*	3.22	1.92*
	2	2.07*	2.05*	2.58	3.14*	3.56	2.51*	1.96*	2.54*

* 表示预报与实况距平符号一致。

一致。在大多数情况下，El Niño 过程中的海温因子场预报结果与实况比较接近，且距平同号率也较高，检查近 40 年来 9 次中等以上 El Niño 年份的预报中，有 7 次具有相同距平符号的区在 5 个以上。表 1 是分别用 1968 年 9 月~1969 年 11 月和 1986 年 9 月~1987 年 11 月两次 El Niño 过程时连续 3 个季的海温因子场，对 1969~1970 年和 1987~1988 年冬季的预报结果。可见大部分区的距平符号与实况相同，有些区的预报值与实况非常接近，反映出海温场对于气温预报的重要性。极涡因子场的整体预报水平在短时效(0~2 季)时较高(见图 1c)，这可能反映出大气的记忆时间较短，因而主要是近期的极涡因子场对气温影响较大。而单季极涡因子场预报水平则起伏很大。综合因子场的预报效果比较稳定(见图 1d)，整体预报水平较高的时效为 0~2 季。其中以连续 2 季因子场预报效果最好，大多数预报时效的相关系数都在较高的水平，在预报时效为 1 季时相关系数最大达 0.34。总体来讲综合因子场预报效果最好也最稳定，预报水平普遍高于其它因子场。

由图 2 可以看出，由于综合因子考虑的因素较多，综合因子总体预报水平比较稳定，且相关系数保持在较高的水平。但许多情况下，综合因子却不一定比同时效下效果最好的单要素因子的相关系数值大。这可能因为在综合考虑多种因子作用的同时，也增加了一定随机误差的缘故。

2.1.2 不同区的比较及其最佳预报因子场 从上述可知，不同要素场预报效果较好的时段是不同的，因此对各要素场进行优选组合，优选后的因子场预报效果明显提高。同时，全国各区的预报结果之间相差很大，不同的区域对因子场的敏感性及其预报效果也有很大差别。大部分地区提前两个季时的预报与实况的相关系数在 0.5 以上(表略)，相关系数最大值在 0.7 左右，其中 1 区和 7 区最好，5 区最差。而多数情况下的最佳因子场为组合因子，只有个别情况下是单要素因子场。

2.1.3 预报值与实况的对比 为更直观的了解预报效果，将历史资料独立样本检验的预报值与实况进行对比。如图 3(a)和 3(b)所示。可见预报结果比较理想，预报趋势在大多数时期与实况吻合得很好。

各区最佳因子场对 1995~1996 年冬季气温等级的预报结果(表 2)，绝大部分地区的预报与实况距平符号相同，有些区的预报值与实况非常接近，预报效果比较理想。

2.2 冬季各月预报结果

对冬季各月进行预报试验的结果表明，各月预报效果明显不如季平均的预报，尤其

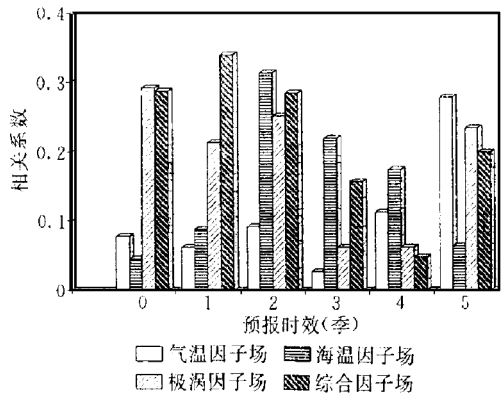


图 2 连续 2 季各单要素因子场与综合因子场对冬季平均气温的预报相关系数随时效变化

Fig. 2 Correlation coefficient between forecasts using consecutive single element and synthetic predictor fields of two seasons and observed values in winter with different lead time

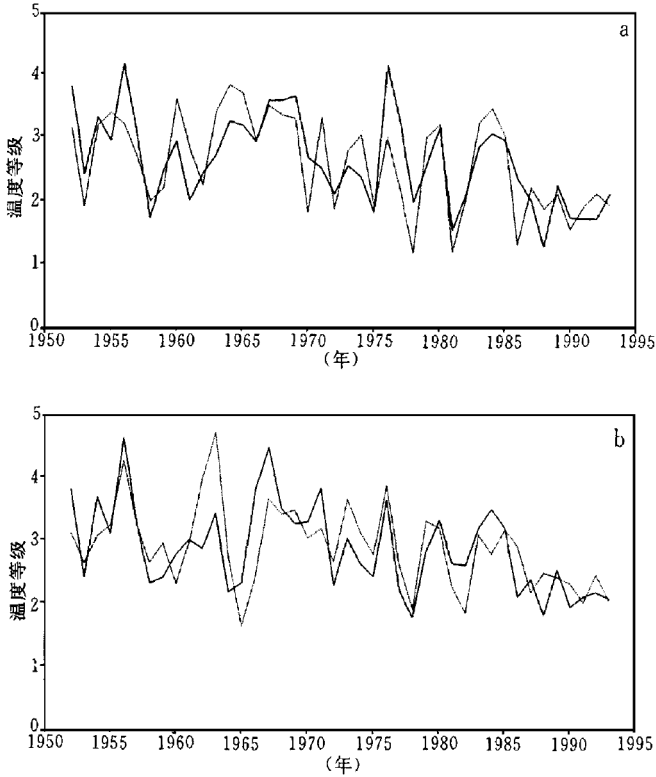


图3 最佳预报结果(虚线)与实况(实线)的对比(a) 1区预报时效为2季 (b)2区预报时效为1季

Fig. 3 Contrast between best forecasts (dashed) and observed values (solid) with time (a) lead time of two season forecast for region 1 (b) lead time of one season forecast for region 2

表2 1995~1996年冬季气温等级的预报结果与实况对比

Table 2 Contrast between forecasts and observed values of winter temperature grades from 1995 to 1996

分区	实况	提前0季	提前1季	提前2季	提前3季
第1区	1.60	1.30*	1.30*	1.30*	1.79*
第2区	2.43	3.65	3.65	1.92*	3.39
第3区	3.10	4.79*	4.79*	4.03*	2.39
第4区	3.60	3.44*	3.44*	3.44*	3.75*
第5区	3.37	3.27*	3.27*	0.00	2.74
第6区	3.10	3.63*	3.63*	3.54*	3.54*
第7区	2.80	1.42*	1.42*	1.42*	1.42*
第8区	2.83	3.49	3.49	1.21*	2.77*

* 表示预报与实况距平符号一致。

是连续因子场的预报效果随预报时效起伏很大, 规律性较差. 这也许是由于月尺度的预

报因子场稳定性较差, 随机性误差较大引起的。

各月的优选因子场预报效果都有不同程度的提高, 其中 12 月份的最佳, 预报效果好于季平均。而 1、2 月份仍然不如季平均的预报效果好。

3 结 论

(1) 用典型相关分析的方法, 选择良好的物理因子场, 对我国基本气候要素气温场的短期气候预测试验表明, 有一定的预报技巧, 特别是对于特定的地区选取特定的因子组合, 可以取得较满意的效果。

(2) 季平均预报结果总体上好于月平均的预报, 预报水平(相关系数)稳定, 且用连续因子场作预报规律性明显。在连续因子场的季平均预报中, 单要素因子场以海温较好, 预报水平随预报时效稳定, 多要素因子场以综合因子最好。挑选的因子场预报效果显著改善, 大多数地区的最佳预报独立样本检验相关系数在 0.5 以上。

(3) 以 CCA 为基础的统计预报模型, 不仅可以取得较好的预报效果, 同时也揭示出不同预报因子其预报性能的气候特性。结果表明, 海洋、大气环流以及前期气候要素场本身对气候变化影响的主要特征为: ① 海洋-大气之间半年左右的相互作用, ② 大气环流对气候变化的短期影响效应, ③ 气候要素自身的演变规律。同时, 也反映出气候的变化及其异常是多种物理因子共同作用的结果, 在气候预测中必需考虑多种物理因子的共同影响, 才能取得好的预报效果。

(4) 海温场是所有因子场中最好的预报因子。不仅单独海温场的预报效果较好, 而且与其它因子场组合后的预报水平可以得到进一步的提高。这也许是由于对月、季时间尺度的气候变化和异常来说, 海温场的作用是十分重要的缘故。

参 考 文 献

- 1 Graham N E, Michaelsen J and Barnett T P. An investigation of the El Niño South Oscillation cycle with statistical models 1 predictor field characteristics. *J. Geophys. res.*, 1987c, **92**: 14251~14270.
- 2 Graham N E, Michaelsen J and Barnett T P. An investigation of the El Niño South Oscillation cycle with statistical models. 2. Model results. *J. Geophys. Res.*, 1987c, **92**: 14271~14289.
- 3 Barnett T P and Preisendorfer R. Origins and levels of monthly and seasonal forecast skill for United States surface air temperatures determined by canonical correlation analysis. *Mon. Wea. Rev.*, 1987, **115**: 1825~1850.
- 4 Barnston A G and Ropelewski C F. Prediction of ENSO episodes using canonical correlation analysis. *J. Climate*, 1992, **5**: 1316~1345.
- 5 Barnston. A G. Linear statistical short-term climate predictive skill in the Northern Hemisphere. *J. Climate*, 1994, **7**: 1513~1564.
- 6 李小泉. 美国长期天气预报业务的新发展——发布气候展望. *气象*, 1995, **21**(9): 49~52.

SHORT-TERM CLIMATE FORECAST EXPERIMENT OF WINTER TEMPERATURE IN CHINA USING CANONICAL CORRELATION ANALYSIS

Mao Hengqing Li Xiaoquan

(*National Meteorological Center, Beijing 100081*)

Abstract

By using the multivariate linear statistical climate forecast model developed from canonical correlation analysis (CCA), a forecast experiment for winter temperature in China is carried out. Forecast skill is assessed using the test method of historical data independent samples.

The results show that there is specified statistical predictive skill for short-term climatic prediction of winter temperature in China using CCA method. In most areas the best correlation coefficient between forecasts and observed values at lead times of 0~2 seasons is more than 0.5. The mean seasonal prediction is more effective than mean monthly one in general.

The sea surface temperature is the most effective predictor field. The SST as predictor field can get higher skill score than other single one, and SST combined with other predictors would get much higher skill score.

Key words: Canonical Correlation Analysis (CCA) Forecast experiment Correlation coefficient