

3 个月平均气温距平的 CCA 预报方法

余金波, 吴洪宝

(南京气象学院大气科学系, 南京 210044)

摘要: 用根据 CCA 方法设计的一个统计预报模式对我国 3 个月平均气温距平进行预报试验, 并用交叉检验方法进行估计。结果表明: 预报技巧随提前时间增长而减小得较少; 用全球表面温度作预报因子有较高的预报技巧; 7、8、9 月 3 个月较易预报, 而 10、11、12 月 3 个月较难预报。

关键词: 典型相关分析, 交叉检验, 温度距平

中图分类号: P456.8 **文献标识码:** A **文章编号:** 1000-2022(2001)02-0171-07

典型相关分析(Canonical Correlation Analysis, 简称 CCA) 是多元统计中的一种基本方法, 它用于研究两组随机变量(向量)之间的相关关系, 在地质、生化、医学、气象等许多实际问题中都得到了应用。国外对 CCA 方法的研究和应用已取得一定的成果, 业务预报中也已开始使用。国内应用 CCA 的例子还不多。毛恒青等^[1]、江志红等^[2]分别用 CCA 做了我国冬季气温短期气候预测试验和 ENSO 海区海温季节预报试验, 结果表明 CCA 方法具有一定的预报能力。

1 资料

原始月平均资料有: (1) 1958-01 ~ 1998-12 全球 500 hPa、700 hPa 高度场(H_{500} 、 H_{700}), 网格距 $2.5 \times 2.5^\circ$; 来源于 NCEP/NCAR 再分析资料^[3]; (2) 1948-01 ~ 1998-12 全球表面温度资料(T_s), 网格距 $2.5 \times 2.5^\circ$; 来源于 NCEP/NCAR 再分析资料。 T_s 在海洋上为海表面温度, 在陆地上为地面温度, 处理后可得 $87.5 \text{ S} \sim 87.5 \text{ N}$ 准全球海表面温度(T_{SS})、中国区域外的准全球陆地表面温度(T_{NCLS}), 以及中国区域内格距为 $5 \times 2.5^\circ$ 的 75 个格点表面温度(T_{CS}); (3) 1948-01 ~ 1995-12 北半球 $20 \sim 80 \text{ N}$ 的海平面气压资料(p_{SL}); (4) 1956-01 ~ 1998-12 中国 160 个基准测站的月平均气温资料。除 T_{CS} 和 160 站气温资料外, 其他资料都处理成 10×5 网格。另外, 用 T_{SS} 、 T_{NCLS} 和 T_{CS} 重新构成全球表面温度因子 T_{GS} 。

T_{CS} 有比 160 站资料时间序列长、空间覆盖均匀的优点, 故本文用 T_{CS} 代表中国气温作为预报对象, 使用前对 T_{CS} 和 160 站气温资料进行了相关分析。全国按文献[4]定义的 7 个年平均气温分区, 各区内挑选靠近测站且与测站数相等的格点, 分别求 T_{CS} 与 160 站气温资料各区各月 1956 ~ 1998 年区域平均值的相关系数(序列长度为 43), 列于表 1。各区对应的范围分别是:

收稿日期: 2000-09-29; 修订日期: 2001-01-04

基金项目: 国家重点基础研究发展规划项目 G1998040901-3 资助

作者简介: 余金波, 男, 1975 年 12 月生, 硕士, 助理工程师, 现在浙江省气象台工作

1 区: 华北、东北; 2 区: 华东; 3 区: 华南; 4 区: 四川盆地、云贵高原东部; 5 区: 滇西南; 6 区: 青藏高原; 7 区: 西北南部。相关系数 t 检验在自由度为 41, 信度为 0.001 下的相关系数临界值为 0.495, 信度为 0.05 时, 相关系数临界值为 0.301。考虑 5 区和 6 区地形的复杂性, 由表 1 可以认为 T_{CS} 近似地代表了中国气温。

表 1 7 个分区各月 T_{CS} 与测站气温的相关系数

Table 1 Correlation coefficient between T_{CS} and station temperature of seven sections in China in each month

月份	相关系数							平均值
	1 区	2 区	3 区	4 区	5 区	6 区	7 区	
1 月	0.873	0.860	0.854	0.791	0.591	0.523	0.720	0.744
2 月	0.959	0.874	0.863	0.859	0.336	0.462	0.809	0.737
3 月	0.936	0.852	0.737	0.827	0.383	0.381	0.754	0.696
4 月	0.883	0.844	0.856	0.875	0.541	0.076	0.658	0.676
5 月	0.892	0.727	0.638	0.473	0.527	0.243	0.577	0.583
6 月	0.780	0.832	0.615	0.523	0.562	0.616	0.495	0.632
7 月	0.624	0.805	0.512	0.483	0.714	0.660	0.700	0.643
8 月	0.785	0.824	0.611	0.589	0.476	0.513	0.620	0.631
9 月	0.907	0.851	0.715	0.720	0.621	0.730	0.693	0.748
10 月	0.867	0.888	0.739	0.606	0.702	0.453	0.627	0.697
11 月	0.927	0.910	0.890	0.829	0.581	0.206	0.788	0.733
12 月	0.955	0.919	0.911	0.891	0.578	0.323	0.826	0.772

2 CCA 预报模式

CCA 方法的基本思想是分别在两组随机变量内做线性组合, 构成各自有代表性的综合变量, 使得两组间成对的综合变量之间的相关达到最大、次大、……。这种综合变量称为典型相关变量。然后用典型相关变量之间的相关关系代替原来两组变量之间的相关关系进行研究。在 Barnett 等^[5]最早应用 CCA 做美国地面月平均气温预报的工作中, 先对预报因子场和预报对象场各自做 EOF 分析, 截取前几个 EOF 的时间系数(主成分, 简称 PC), 然后把因子和对象场各自的前几个 PC 作为 CCA 对象, 组成典型相关变量。详细的模式算法可参考文献[5]。

由 CCA 的结果得到 Q 个模态, 每个模态包括相应的典型相关系数 μ , 预报因子、预报对象各自的典型相关变量 U, V , 典型图 g, h 。如果将 CCA 方法看成类似于 EOF 的分解, 则 U, V 是相应的时间系数, g, h 是相应的空间型。 g 图中的“活动中心”(绝对值大值区)表示该区能提供较好的预报能力, h 图中“活动中心”表示该区有较好的可预报性。 g 实际上是预报因子与典型相关变量 U 的相关系数场, h 则是预报对象与 V 的相关系数场。最后采用因子场的前几个典型相关变量作为建立预报方程的因子, 模式预报方程为

$$\hat{T}(x, t) = \sum_{j=1}^Q U_j(t) \mu_j h_j(x), \quad x = 1, 2, \dots, q, \quad Q \leq Q_0.$$

式中 Q 一般取 1 或 2^[5]。

模式预报技巧的估计采用被称为“one-year out cross-validation”^[6]的交叉检验方法。具体做法是, 每次模式预报方程由所有可用资料中去掉第 v 年的资料建立, 然后用保留的第 v 年的因子资料作为因子观测值进行预报, 而保留的第 v 年的预报对象资料作为实况。重复以上过

T_{NCL5} 、 T_{SS} 、 T_{CS} 、 T_{GS} 为因子时, 预报技巧随提前时间增长也略为降低, 提前时间达 13 个月时, 降低得也不多; 以 H_{500} 、 H_{700} 、 p_{SL} 为因子时, 预报技巧随提前时间增长反而增高。

3.2 预报技巧的地理分布

预报技巧的地理分布随提前时间的增长没有太大的变化, 中心位置一般不变, 中心最大值略降低, 范围略缩小。分析以 T_{GS} 为预报因子, 提前时间为 1 个月时, 各个目标季预报技巧的地理分布(图略)可以看出, 预报技巧在不同目标季的地理分布差别较大, 这说明同一个预报因子对不同时间不同区域的预报起的作用相差较大。用 CCA 做预报时, 应考虑预报区在预报时段内的可预报性。

3.3 预报技巧来源

JAS 的平均预报技巧最高, 预报技巧在全国的分布范围也最广。交叉检验预报时, 每次拿掉一个样本都可以得到相应的预报因子和预报对象典型图及典型相关变量。这里分析拿掉预报因子和预报对象最后一个样本得到的预报因子和预报对象典型图及典型相关变量。为方便各种提前时间下交叉检验的计算, 并使表 2 有较好的可比较性, 所有试验的样本数取 48, 做 JAS 预报时因子的总体样本为 1949-07 ~ 1997-06。

图 2、图 3a、图 4 分别是 CCA 第一模态的预报因子典型图、预报对象典型图和典型相关变量 U_1 、 V_1 。该模态的典型相关系数为 0.93。因子提供的预报信息对中国起预报作用的区域如图 3a 所示, 主要在西北、中部和东北东南部。该模态提供的预报信息的物理含义, 由图 2 和图 4 可以看出是温度的长期变率。图 4 中 U_1 的变化特点是: 50 年代到 70 年代中期有线性下降趋势, 70 年代中期到 90 年代较稳定。这样, 图 2 上正值区提供的预报信息是该区 50 年代到 70 年代中期的线性降温趋势, 负值区则相反。

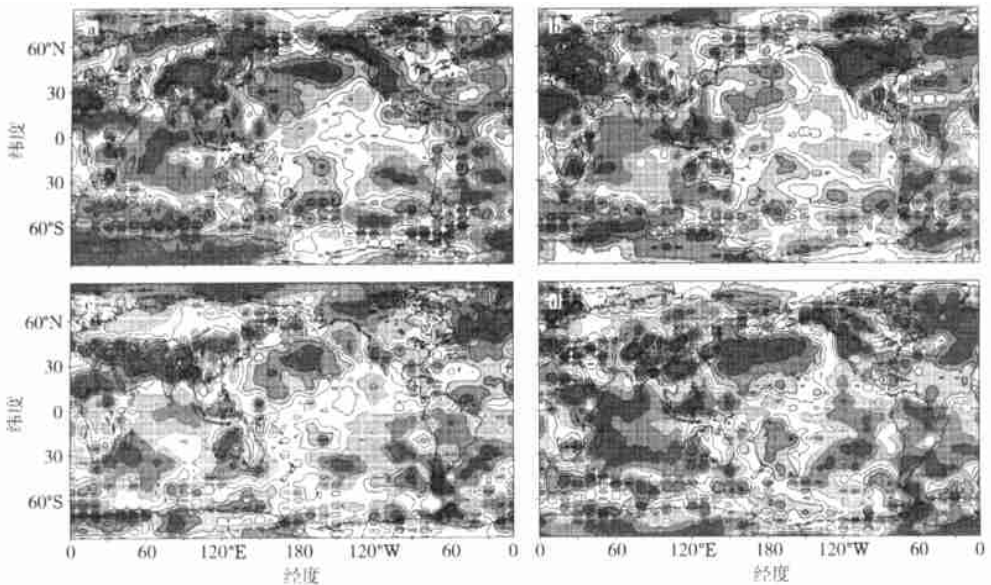


图 2 CCA 第 1 模态的因子典型图

a、b、c、d 分别对应 JAS、OND、JFM 和 AMJ; 等值线间隔 0.2, 负值用虚线表示

Fig. 2 CCA loadings for mode 1

Panels a, b, c, d show loading for JAS, OND, JFM and AMJ, respectively

Contour interval is 0.2, negative contours are dashed

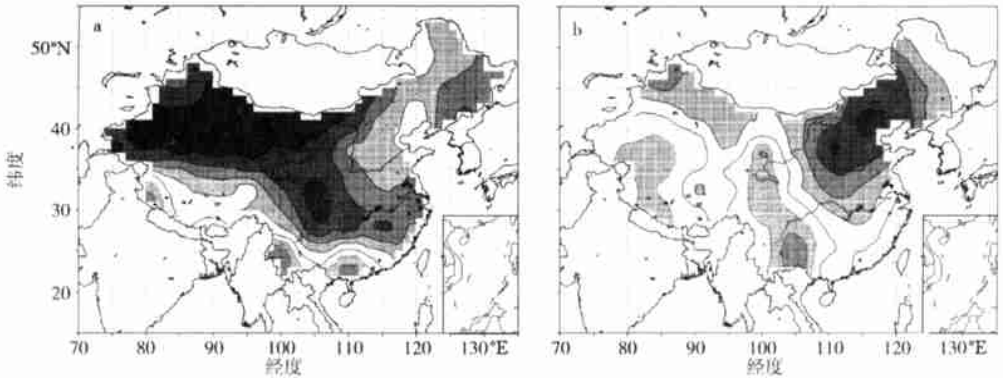


图 3 CCA 预报对象典型图
 a. 第 1 模态; b. 第 2 模态
 等值线间隔 0.2, 负值用虚线表示

Fig. 3 The principal predictand loading, panel a for mode 1 and panel b for mode 2
 Contour interval is 0.2, negative contours are dashed

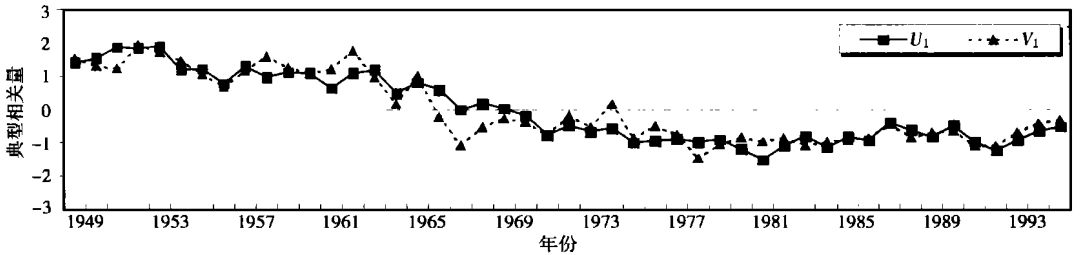


图 4 CCA 第 1 模态的典型相关变量 U_1 、 V_1
 横坐标取为因子第一个月(7月)所在年

Fig. 4 The canonical component time series for mode 1
 Year is defined by the year of first month (July) of predictor

由图 2 可以看出, 全球陆地和海洋的大部分区域都有一定的预报能力。其中, 在海洋上预报能力较强的是印度洋(特别是中部和北部海区)、北大西洋中部和北部、南大西洋中部、北太平洋北部和南太平洋的东部。陆地上各大洲都能提供预报信息, 但在不同的时间里, 提供信息的能力有所不同。

图 5、图 3b、图 6 分别是 CCA 第二模态的预报因子典型图、预报对象典型图和典型相关变量 U_2 、 V_2 。该模态的典型相关系数为 0.80。图 6 上符号“W”(“C”)表示预报因子第 1 组 3 个月(JAS)有连续 2 个或 3 个月处在文献[8]中表 2 的厄尔尼诺(拉尼娜)事件中。由图可以看出, 这一模态分析出的厄尔尼诺和拉尼娜事件都基本吻合。结合图 5 可以得知, 这一模态主要描述的是一次冷 ENSO 事件由 JAS 开始消亡的过程。该模态提供的预报信息主要对我国华北和东北西南部起预报作用(图 3b)。

4 结 论

(1) 用 CCA 方法预报我国 3 个月平均气温距平是一种可行的方法, 在预报试验中取得了较好的预报技巧。预报技巧随提前时间的增长一般减少不大。

(2) 本文试验的诸预报因子中, 预报效果最好的因子是全球表面温度。预报得最好的是 7、

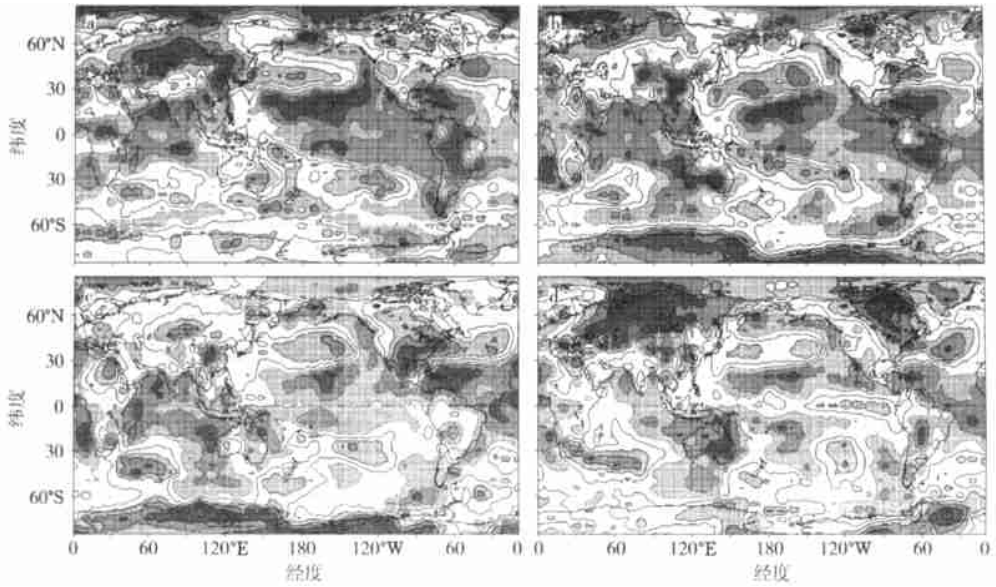


图5 CCA 第2模态的因子典型图

a、b、c、d 分别对应 JAS、OND、JFM 和 AMJ；等值线间隔 0.2，负值用虚线表示

Fig. 5 CCA loadings for mode 2

Panels a, b, c, d show loading for JAS, OND, JFM and AMJ, respectively

Contour interval is 0.2, negative contours are dashed

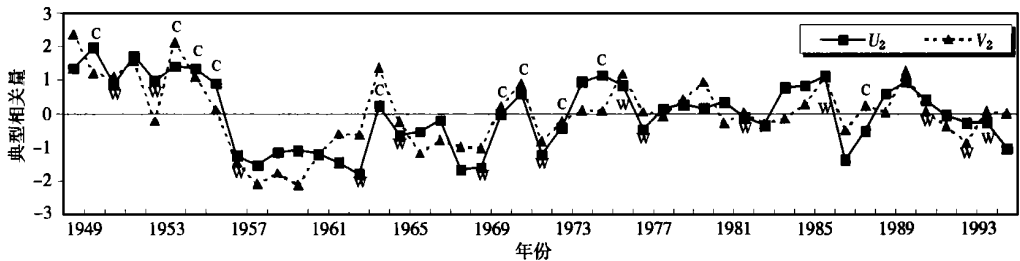


图6 CCA 第2模态的典型相关变量 U_2 、 V_2

横坐标取为因子第一个月(7月)所在年

Fig. 6 The canonical component time series for mode 2

Year is defined by the year of first month (July) of predictor

8、9月，而10、11、12月较难预报。

(3) 我国气温的长期变率与全球表面温度的长期变率关系较好，尤其是印度洋中部、北部海区。从50年代到70年代中期，印度洋、热带太平洋、南大西洋有增暖趋势，而同时南北太平洋海盆和北大西洋海盆有降温趋势。

(4) ENSO 与我国3个月平均气温距平有一定联系，在CCA预报方法中对我国气温预报能起一定的作用。

致谢：全球表面温度资料由NCEP/NCAR的汪学良博士和朱跃建博士整理并提供，谨致谢！

参考文献:

- [1] 毛恒青, 李小泉. 典型相关分析(CCA)对我国冬季气温的短期气候预测试验[J]. 应用气象学报, 1997, 8(4): 385 ~ 391
- [2] 江志红, 丁裕国, 周琴芳. 用于 ENSO 预测的一种广义典型混合回归模式及其预报试验[J]. 热带气象学报, 1999, 15(4): 322 ~ 329
- [3] KALNAY E, KANAMITSU M, KISTLER R, et al. The NCEP/NCAR 40-year reanalysis project [J]. Bulletin of American Meteorological Society, 1996, 77(3): 437 ~ 471
- [4] 辛若桂, 屠其璞. 我国气温长期变化的区域特征[J]. 南京气象学院学报, 1997, 20(1): 54 ~ 63
- [5] BARNETT T P, PREISENDORFER R. Origins and levels of monthly and seasonal forecast skill for United States surface air temperature determined by Canonical correlation analysis[J]. Mon Wea Rev, 1987, 115(9): 1825 ~ 1847
- [6] BARNSTON A G, HE Yuxiang. Skill of canonical correlation analysis forecasts of 3-month mean surface climate in Hawaii and Alaska[J]. J Climate, 1996, 9(10): 2579 ~ 2605
- [7] BARNSTON A G. Linear statistical short-term climate predictive skill in the Northern Hemisphere[J]. J Climate, 1994, 7(10): 1513 ~ 1564
- [8] KEVIN E T. The definition of El Nino[J]. Bulletin of American Meteorological Society, 1997, 78(12): 2771 ~ 2777

CCA forecast scheme of 3-month mean temperature anomaly

YU Jin-bo, WU Hong-bao

(Department of Atmospheric Sciences, NIM, Nanjing 210044)

Abstract: A statistical model is CCA-designed to forecast 3-month mean temperature anomaly in China, which is estimated by using cross-verification scheme, indicating that the skill decreases slowly with the increased leading time intervals; higher skills are found for quasi-global surface temperature as a predictor; it's easy to predict JAS temperature and hard to deal with OND analog. Some meaningful results are obtained from the forecast skill analysis.

Key words: canonical correlation analysis, cross-verification, temperature anomaly