

简化典型相关预测模型面雨量中期预报试验^①

毛恒青 王秀文 晁淑懿 胡 博

(国家气象中心, 北京 100081)

提 要

以5天平均的500hPa高度场格点高度值为预报因子,三峡地区4个区的候面雨量为预报对象,建立简化的典型相关预报模型,并采用历史资料独立样本检验的方法进行预报试验。对预报效果的分析表明:简化的典型相关预报模型对三峡地区面雨量的中期预报有一定的技巧,预报效果基本上能达到中期天气预报业务的要求,对预报员有一定的参考价值。

关键词: 典型相关分析 面雨量 中期预报

引 言

目前数值天气预报产品已相当丰富,如何将反映环流形势的数值天气预报产品变为天气要素的预报,是中期天气预报业务中关心的重要问题。现有的方法大都是用数值预报场与天气要素建立回归方程以作出要素预报,但这里的最大困难在于对预报因子的选取。首先,数值预报场格点资料很多,仅选取某几个格点值作为预报因子,很难作出对天气要素的准确预报;其次在众多格点中要找到好的因子十分困难,且工作量大;另外,回归分析对因子数量有很大限制,不可能有很多的因子入选方程。典型相关预测模型正好可以解决这一困难。

典型相关分析是一种多变量统计技术。它的优点是可以提取场与场之间的相关信息,而这些信息又是因子场与预报场之间相关性最好的。近年来典型相关预测模型在气候预测中有较多的应用^[1~3]。本文将文献[1]提出的简化典型相关模型(简称BP-CCA)应用于三峡地区面雨量中期预报中。一方面检

验典型相关预测模型在面雨量中期预报中的应用效果,为长江三峡二期工程提供面雨量预报服务;另一方面通过统计预测模型与数值预报产品的结合,增强中期预报业务中对数值预报产品的统计解释应用能力,尝试对数值预报产品解释应用的客观化和定量化。

1 预报模型的建立和预报试验

1.1 典型相关分析原理简介

典型相关分析是分别对因子场和预报场进行线性组合生成新的变量,这种新变量称为典型因子。而这种线性组合是在因子场和预报场的典型因子间相关性极大的条件下完成的,因此它是所有线性组合中最佳的组合,典型因子携带了两个场之间主要的相关性信息。设因子场和预报场资料阵为:

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

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

其中 j, l 分别为因子场与预报场的空间点数,资料样本为 n 。典型因子分别为:

$$U_{k \times n} = C'_{j \times k} \cdot X_{j \times n} \quad (3)$$

$$V_{k \times n} = D'_{l \times k} \cdot Y_{l \times n} \quad (4)$$

^① “长江三峡二期工程气象保障服务”课题资助。

其中 k 为典型因子对的个数, C 、 D 为典型因子的组合系数。在典型因子 U 、 V 相关性极大的条件约束下, 对典型因子的求解转化为特征值问题。用典型因子反推即可得出预报方程。

1.2 预报因子及试验方案

对三峡地区降水量与 500hPa 高度场进行了相关性分析。由相关系数分布图(略)知道, 在 $20^{\circ}\sim 60^{\circ}\text{N}$ 、 $30^{\circ}\sim 150^{\circ}\text{E}$ 范围内, 500hPa 的高度场与三峡地区降水量相关性较大。因此选取该范围内的 ECMWF 500hPa 高度场候平均格点高度值作为因子场, 以 4 个区区域平均(共 28 个站)和总平均的候降水总量为预报对象, 建立典型相关模型进行预报试验, 实际应用时将数值预报产品代入模型即可。资料为 1971~1997 年 3~10 月逐候的因子场和预报场, 每月的资料序列长度为 162。

用历史资料独立样本检验的方法进行预报试验, 即从第一候开始, 每次剔除 1 候, 用余下的 161 候资料建立方程, 并对剔除的 1 候作预报, 这样对每月进行 162 次建立方程和预报的过程, 最终可形成与实况长度相同的预报序列。

1.3 试验概况

在进行典型相关过程之前, 首先对因子场和预报场进行标准化和 EOF 展开, 根据特征值的大小, 去掉尾部特征值较小的主分量, 然后用剩余的主分量进行典型相关分析, 建立预测模型。对主分量的截断原则是当某一个主分量的数值小于前一个主分量的 1% 时, 就去掉后面的所有主分量。因此在历史资料的独立样本试验中, 对因子场每次截断的长度并不相同, 一般为 80~90 个(共有 161 个)主分量之间。由于预报场仅有 5 个预报量, 因此预报场保持 5 个主分量而不截断。这样最后在作典型相关分析时只有 5 对典型因子进入预报方程。

2 结果分析

2.1 预报距平同号率

分别将各区降水量预报值和实况与相应的气候平均值比较, 计算距平值并统计预报和实况正负距平符号相同的频率, 简称距平同号率。结果如表 1 所示: 各区的距平同号率大都在 60% 左右。各月比较同号率相差也不大, 变化稳定, 8 月相对较差, 其中 3 区 8 月份的距平同号率小于 50%。

平均是指 4 个区预报值平均与实况平均的距平同号率。比较可知: 平均的距平同号率最高, 且较为稳定。最大值为 62.96%, 出现在 3 月份。最小值也有 53.09%, 出现在 10 月。一般而言, 3 月的预报效果最好, 8 月和 10 月较差。

表 1 预报与实况的距平同号率/%

月	3	4	5	6	7	8	9	10
1 区	62.35	59.26	59.26	58.02	61.11	55.56	54.94	50.62
2 区	61.11	60.49	59.88	52.47	58.02	59.26	59.88	51.85
3 区	56.17	54.32	58.64	56.79	58.64	46.91	58.02	53.09
4 区	67.28	62.96	53.09	61.11	55.56	52.47	62.96	58.02
平均	62.96	58.15	55.93	57.53	58.27	53.33	58.02	53.09

2.2 相关系数

分别计算了各区预报值与实况的相关系数, 并对相关系数进行显著性检验。在显著水平 $\alpha = 0.05$ 、自由度为 $n - 2 = 160$ 时, t 检验统计量的临界值为 $t_{\alpha} = 1.96$, 按此计算显著相关的相关系数临界值为 $r_{\alpha} = 0.153$, 即相关系数大于 0.153 时认为相关性显著。

各区相关系数的分布如图 1 所示。1 区有 5 个月通过了显著性检验, 3~5 月预报相关系数较大, 为 0.24~0.27, 而 6、8、10 各月较差, 其相关系数不到 0.1。2 区预报效果较好, 除 10 月份外均在 0.153 以上, 相关性显著, 3 月份相关系数最大达 0.4 以上。3 区有 6 个月通过了显著性检验, 其中 3~5 月及 9 月的预报相关系数在 0.2 以上, 而 8、10 月份较差没有通过显著性检验。4 区的预报效果最好, 有 3 个月的相关系数大于 0.3, 仅 8 月份没有通过给定信度的显著性检验。平均相

关系数较为稳定,有6个月通过了显著性检验,3月最好,8月最差。

各月比较而言,除8、10两个月各仅有一个区的相关系数通过显著性检验,其余6个月基本上都通过了给定信度的显著性检验。3~6月相关系数较高,预报效果好,7~10月

相关系数较小,预报效果差,特别是8月最差。分析认为,夏季月份天气变化剧烈,局地性降水过程较多,有些降水过程如低层切变降水在500hPa高度层并没有反映,仅用500hPa一层作为因子预报效果不够理想,如果加入800hPa温度场也许会有所提高。

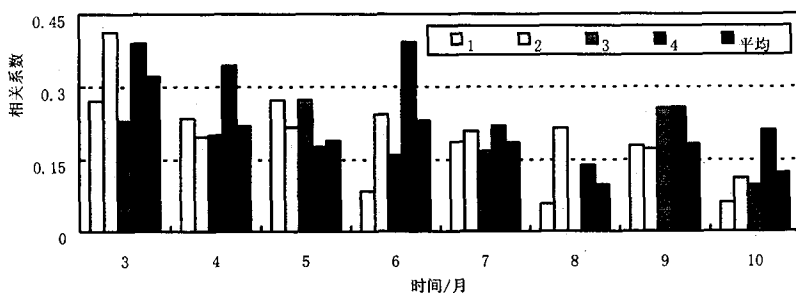


图1 各区预报相关系数

2.3 量级比较

为了比较预报与实况之间量级上的差别,分别将逐候降水量的预报值和实况按大小划分成5个级别,分别为:小于10mm、10~24.9mm、25~49.9mm、50~99.9mm、大于等于100mm。然后分别统计预报与实况同一级别、相差1级、相差2级、相差3级、相差4级的频率。

图2是预报值与实况同级(a)、相差1级(b)和两者之和(c)的频率,由图可知:

(1) 同级率3月份最高,达53.7%~77.7%,其余月份大都在40%左右,夏季7、8月频率最小。相差1级的频率大都在30%~50%之间。两者之和在70%以上,3、4月份大于90%,5、10月份在80%~90%之间,6~9月为70%~80%,仍以8月最差。

(2) 比较而言,以全区平均的预报值最为接近实况,同级率和相差一级的频率和大都在90%以上,其中3月份的级差全部在1级之内。即全区预报值的平均绝大多数与实况同级或相差1级,表明该方法对区域性降水的预报效果较好。

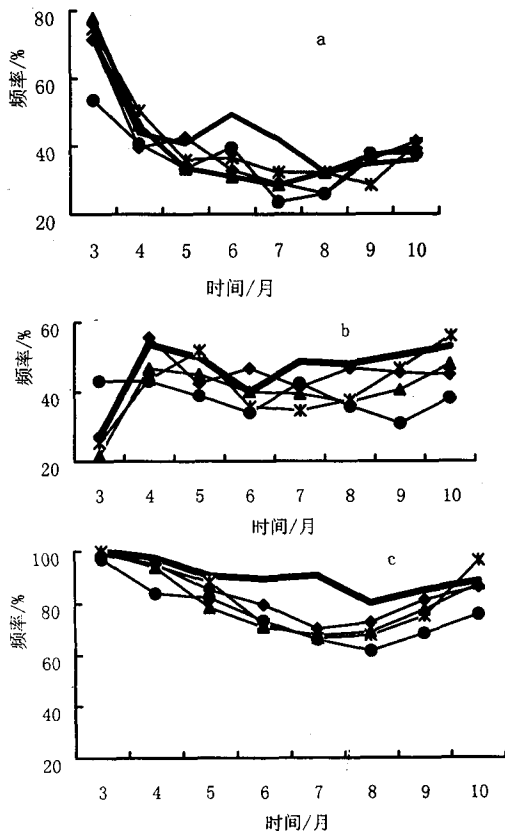


图2 预报值与实况同级(a)、相差1级(b)及两者之和(c)的频率

3 讨论和小结

典型相关是一种多变量统计技术,它的优越之处在于可以提取整个场的有用信息,并且可以同时用几层资料作为因子场,增大了因子场包含的信息量;最重要的是它有效地解决了回归分析时寻找因子的困难。但作为统计方法本身也有其缺点,就是不可能作出十分详细的预报,如对单站逐日降水量的预报等。适合对区域性降水趋势的预报,如区域性候或旬降水总量的预报等,这点也正是中期天气预报业务的重要任务之一。

对预报试验结果的分析认为:用典型相关预测模型,对三峡地区候降水量预报效果较好,预报同号率大都达到60%左右;从量级上比较绝大多数情况下预报值与实况同级或相差1级,特别是对全区平均的预报效果最好;在大多数情况下预报与实况的相关性

显著。预报技巧基本上能达到中期天气预报业务的要求,对预报员有一定的参考价值,应用时将数值预报产品代入相应的预报模型即可。

以上仅作了预报试验,实际预报效果如何,还有待于业务预报检验。

参考文献

- 1 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.
- 2 毛恒青,李小泉.典型相关分析对我国冬季气温的短期气候预测试验. *应用气象学报*, 1997, 8(4): 385~392.
- 3 毛恒青,李小泉.我国夏季降水与前期太平洋海温场关系的典型相关分析. *南京气象学院学报*, 1998, 21(1): 130~137.

Medium-range Forecast of Areal Precipitation in Sanxia Areas by Using the BP Canonical Correlation Analysis Model

Mao Hengqing Wang Xiuwen Chao Shuyi Hu Bo

(National Meteorological Center, Beijing 100081)

Abstract

The 5-day average of the 500hPa height field as predictors and the pentad regional precipitation of 4 areas in sanxia, China as predictands were used to build the BP Canonical Correlation Analysis (BP-CCA) model. The forecast skills of BP-CCA for the precipitation of 4 areas were estimated by using a cross-validation technique. The forecast test shows that the predictive skills of BP-CCA for areal precipitation forecast in sanxia areas are good enough for the operational medium-range weather forecast and the results are valuable for reference to forecasters.

Key Words: canonical correlation analysis areal precipitation medium-range forecast