

长江中游夏季降水场预报技术研究*

张礼平

(武汉中心气象台, 武汉 430074)

提 要 经 SVD 分析, 截取足够多的预报场和因子场时间系数, 使其相互关系代表两场的大尺度联系, 预报场时间系数与其奇异向量线性组合估计场能反映原场主要特征. 借助最优化技术, 选择合理的系数, 建立预测公式, 由因子场时间系数预测预报场时间系数, 同时订正预报场时间系数 $a_1 a_2 a_N$ 本身的误差和反演过程中分析误差造成的场格点趋势预测的误差. 最后将预测的预报场时间系数和对应奇异向量反演为整个场的预报. 预报过程重点考虑可预报的大尺度变化, 滤去不可预报的小扰动, 依据两场主要耦合关系, 预测预报场未来的主要变化.

关键词 最优化技术 奇异值分解 场估计

分类号 P457.6

由于季风活动的显著年际变化, 长江中游是我国气象灾害的高发区, 其中, 旱涝是最重要的气象灾害. 20 世纪 80 年代以来, 大范围、持续性的大旱大涝灾害频繁发生, 给国民经济发展造成了严重影响. 及时、准确地预测这些天气、气候异常的趋势、程度及分布, 对各级政府部署、领导防灾减灾, 将损失降到最低, 有着极其重要的意义.

20 世纪下半叶, 天气和气候分析基本对象已由单站要素时间序列逐渐转变为区域、半球或全球要素场时间序列. 单站要素分析、预测没有考虑周围站点的相互联系, 往往导致预测结果水平分布出现不合理的不连续, 与天气学原理相悖, 同时也远不能满足要求各级政府决策需求. 场上的点与周围的点有着密切联系, 因而单点、或数个点的值难以使用, 也不应单独、或分块使用, 而应从整个场上提取有用的预报信息. 场的分析方法, 考虑的是因子场对预报场主要的影响关系, 重点考虑大尺度变化, 滤去小扰动, 力图从因子场的大尺度变化对预报场的变化进行客观、总体的描述, 这种思路是与天气学原理相一致. 因此, 场的分析、预测必然最终成为气候分析和预报的主流.

自 1976 年 Prohaska 提出将自然正交函数分解 (EOF) 技术应用于分析两气象场间的关系以来^[1], 奇异值分解 (SVD) 方法在气象中应用逐渐增多. 1992 年 Bretherton 等比较了已有的提取两场主要耦合信号的方法后认为: SVD 操作简单, 易于解释说明其意义, 有广泛的应用前景^[2]. 丁裕国等也从理论上证明了 SVD 在场诊断分析中的普适性^[3].

目前 SVD 的应用, 主要在两气象场相互关系的诊断分析^[4-5], 提取前 N 个能解释两场协方差大部的模态和对应同场相关图, 根据因子场实况距平分布与同场相关图分布相似推测

*2003-07-10 收稿; 2003-11-28 收修改稿. 张礼平, 男, 1956 年生, 高级工程师, email: zhangliping_wh@yahoo.com.cn.

预报对象场未来的距平分布^[6]. 考虑到 SVD 提取的因子场时间系数与预报对象场大部格点 (或测站) 随时间相互变化存在某种规律, 作者提出了构造因子场前 N 时间系数线性函数, 借助最优化技术, 在统计时段内能拟合预报场每个格点正负距平趋势次数最多意义下确定待定系数的方法, 实现了由因子场对预报场的客观预报^[7], 但这种方法仅给出了预报场正负距平趋势的水平分布, 并没有给出量值, 因此, 这仅是一个定性预报方法.

本文将根据 SVD 原理及场的估计方法, 借助最优化技术, 导出场对场定量预报方法.

1 基本原理

SVD 原理

设有 x 、 y 两个场, 称为左场和右场, 分别包括 p 和 q 个空间格点, 有 n 次经过方差标准化处理的观测值, 用矩阵表示为:

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ x_{p1} & x_{p2} & \cdots & x_{pn} \end{pmatrix} \quad Y = \begin{pmatrix} y_{11} & y_{12} & \cdots & y_{1n} \\ y_{21} & y_{22} & \cdots & y_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ y_{q1} & y_{q2} & \cdots & y_{qn} \end{pmatrix}$$

设左场 x 为预报对象场, 右场 y 为因子场. 由 SVD, 可找到两个正交线性变换矩阵 L 和 R , 使得^[2]

$$\text{cov}(L^T X, R^T Y) = \text{Max}$$

其中, $L^T X = A$, 称为左场时间系数矩阵, $R^T Y = B$, 称为右场时间系数矩阵. L 和 R 的第 k 列向量分别称为第 k 左、右奇异向量.

$$L = \begin{pmatrix} l_{11} & l_{12} & \cdots & l_{1p} \\ l_{21} & l_{22} & \cdots & l_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ l_{p1} & l_{p2} & \cdots & l_{pp} \end{pmatrix} \quad R = \begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1q} \\ r_{21} & r_{22} & \cdots & r_{2q} \\ \cdots & \cdots & \cdots & \cdots \\ r_{q1} & r_{q2} & \cdots & r_{qq} \end{pmatrix}$$

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{p1} & a_{p2} & \cdots & a_{pn} \end{pmatrix} \quad B = \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ b_{q1} & b_{q2} & \cdots & b_{qn} \end{pmatrix}$$

SVD 同时对左、右场的 p 和 q 个变量实施不同的变换, 得到的新变量——时间系数具有这样的优良性质: 左场 (或右场) 某一时间系数仅与对应的右场 (或左场) 时间系数协方差不为 0, 其余均为 0, 也即左场 (或右场) 时间系数仅与右场 (或左场) 对应时间系数相关系数不为 0, 其余均为 0. 时间系数是按其成对的协方差大小排列的, 且大协方差一般集中在前 N ($N \leq \min(p, q)$) 对上, 余下的协方差较小, 因此, 可选取前 N 对时间系数,

这前 N 对时间系数的相互关系就可在很大程度上代表两场的相互关系, 分析左、右两场 p 个变量与 q 个变量随时间相互变化的众多关系变为分析这 N 对时间系数随时间相互变化的简单关系, 不仅使问题大大简化, 且突出了重点.

1.2 场的估计、逼近与预测

众所周知, EOF 能将数据场分解为不随时间变化的空间函数 (特征向量) 以及只依赖时间变化的时间系数 (主分量), 利用方差集中在前 N 个主要分量的特征, 用前 N 空间函数和时间系数的线性组合构成对原场的估计, 略去原场中方差较小分量, 保留较大分量, 由于正交变换不改变场的总方差, 因此估计场保留了原场大部方差, 反映了原场的主要特征. 若 N 取得足够大, 越接近原场格点数 p , 越能以高精度逼近原场, 特别地, 若 $N=p$, 则估计场完全等同于原场.

既然 EOF 为 SVD 的特例, 它们对原场都是实施的正交线性变换, 且在 SVD 中, 考虑到右场为因子场, 可由简单关系实现右场时间系数对左场时间系数预测, 由此自然想到: 能否由右场足够多的时间系数, 预测左场时间系数, 这些左场时间系数和左奇异向量线性组合构成对原场的估计, 从而实现右场对左场的定量预测?

设前 N 对时间系数累积协方差占两场协方差大部, 若时间系数间存在线性关系, 由

$$a_k = c_{0k} + c_{1k} b_k \tag{1}$$

($k=1, 2, \dots, N$, c_{0k} 、 c_{1k} 为待定系数, a_k 、 b_k 分别为第 k 左场和右场时间系数) 实现对 a_k 的预测. 另一方面, 由 $L^T X=A$, 注意到 L 为正交矩阵, 则有 $X=LA$, 即 $N=p$ 时

$$\begin{pmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \dots & \dots & \dots & \dots \\ x_{p1} & x_{p2} & \dots & x_{pn} \end{pmatrix} = \begin{pmatrix} l_{11} & l_{12} & \dots & l_{1N} \\ l_{21} & l_{22} & \dots & l_{2N} \\ \dots & \dots & \dots & \dots \\ l_{p1} & l_{p2} & \dots & l_{pN} \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{N1} & a_{N2} & \dots & a_{Nn} \end{pmatrix} \tag{2}$$

当 $N < p$ 时, 上式中 “=” 应改写为 “ \cong ”.

若外推 (预测) 1 个样本, 则预报公式为:

$$\begin{pmatrix} x_{1,n+1} \\ x_{2,n+1} \\ \dots \\ x_{p,n+1} \end{pmatrix} = \begin{pmatrix} l_{11} & l_{12} & \dots & l_{1N} \\ l_{21} & l_{22} & \dots & l_{2N} \\ \dots & \dots & \dots & \dots \\ l_{p1} & l_{p2} & \dots & l_{pN} \end{pmatrix} \begin{pmatrix} a_{1,n+1} \\ a_{2,n+1} \\ \dots \\ a_{N,n+1} \end{pmatrix} \tag{3}$$

将(1)预测得到的 $a_{1,n+1}$ 、 $a_{2,n+1}$ 、 \dots 、 $a_{N,n+1}$ 代入 (3) 式, 即可得到左场预测值 $x_{1,n+1}$ 、 $x_{2,n+1}$ 、 \dots 、 $x_{p,n+1}$.

1.3 SVD 实现场预测的条件及可行性

由上所述, 当满足以下条件时, 上述预报过程是可实现的:

- (3) 前 N 对时间系数可解释两场大部协方差, 这 N 对时间系数相互关系足以反映两场相互变化的主要信息.
- (2) 右场时间系数可预测左场时间系数.
- (3) 前 N 左场奇异向量 (空间函数) 和时间系数线性组合构成的估计场总方差占原场

总方差大部, 使得估计场可反映原场主要特征, 保证预测过程有意义.

一般来说, 气象场中空间格点间关系较密切、均匀, 由 SVD 过程的主要目标——极大化协方差, 条件 (1) 通常是能得到满足的.

由于场可视为空间函数和时间系数的线性组合, 空间函数被认为是固定不变的, 因而时间系数的预测精度直接影响整个场的预测精度. 与典型相关分析 (CCA) 不同, SVD 是极大化协方差过程, SVD 得到的两场时间系数相关关系往往不如 CCA, 回归预测一般难以满足精度要求. 另一方面, 即使有了高精度时间系数预测值, 由于前 N 时间系数和对应空间函数线性组合反演的场与原场存在的分析误差, 也将影响场的预测精度.

既然两场时间系数的协方差较大, 因而可能存在某种关系. 考虑到线性函数是最简单的函数形式, 构造时间系数预测公式为:

$$\begin{aligned} a_1 &= c_{01} + c_{11}b_1 \\ a_2 &= c_{02} + c_{12}b_2 \\ &\dots\dots\dots \\ a_N &= c_{0N} + c_{1N}b_N \end{aligned} \quad (4)$$

式中 $k=1, 2, \dots, N, c_{0k}, c_{1k}, \dots, c_{nk}$ 为待定系数, a_k, b_k 分别为第 k 左场和右场时间系数. 显然, 预测时间系数的目的是为了预测整个场, 而趋势预报正确的站点数是衡量场预测精度的重要指标. 为此, 依据下面原则确定 c_{0k}, c_{1k} 统计年段内, 由 (4) 确定的 a_1, a_2, \dots, a_N 与对应空间函数反演的场平均每年能拟合原场的正、负距平趋势的格点数 N_0 最多. 显然, N_0 是 a_1, a_2, \dots, a_N 的函数, 而 a_1, a_2, \dots, a_N 又是 $c_{01}, c_{11}, \dots, c_{0N}, c_{1N}$ 的函数. $c_{01}, c_{11}, \dots, c_{0N}, c_{1N}$ 取不同的一组值时, a_1, a_2, \dots, a_N 一般是不同的, N_0 也随之不同, 也即 N_0 为 $c_{01}, c_{11}, \dots, c_{0N}, c_{1N}$ 的函数. 要找到这样的 $c_{01}, c_{11}, \dots, c_{0N}, c_{1N}$ 将其代入 (4) 式后得到 a_1, a_2, \dots, a_N 其反演的场平均每年能拟合原场的正、负距平趋势的格点数 N_0 最多, 也即 $N_0(c_{01}, c_{11}, \dots, c_{0N}, c_{1N}) \rightarrow \text{Max}$. 这就是数学上的“无约束最优化问题”, 其数学模型如下:

$$\text{opti } N_0(c_{01}, c_{11}, \dots, c_{0N}, c_{1N})$$

其中 opti 表示最优值.

借助最优化技术, 满足要求的 $c_{01}, c_{11}, \dots, c_{0N}, c_{1N}$ 是可以找到的. 所谓最优化技术就是从所有可能方案中选择最合理的一种以达到最优目标 (使目标函数最大或最小) 的数学方法, 一般地由最优化技术确定的 $c_{01}, c_{11}, \dots, c_{0N}, c_{1N}$ 得到的反演的场不可能每一年拟合原场的正、负距平趋势的格点数都多, 但是这是在实数范围内取的所有 $c_{01}, c_{11}, \dots, c_{0N}, c_{1N}$ 中, 其反演的场平均每年能拟合原场的正、负距平趋势的格点数 N_0 最多的一组, 因而选取这组 $c_{01}, c_{11}, \dots, c_{0N}, c_{1N}$ 确定 (4) 是合理的. 现代的最优化技术, 一般都可在计算机上得到实现. 由最优化技术确定了 $c_{01}, c_{11}, \dots, c_{0N}, c_{1N}$ 后, 实际预报时, 将 b_1, b_2, \dots, b_N 的实况值代入 (4), 计算得到 a_1, a_2, \dots, a_N 与对应空间函数线性组合即可得到预报场. 如此条件 (2) 得到满足.

尽管这样确定的 (4) 不能较好地拟合 a_1, a_2, \dots, a_N 本身, 但由此反演的场可较好地拟合原场众多格点的正、负距平趋势, 同时订正了 a_1, a_2, \dots, a_N 本身的预报误差和反演过程的分

析误差造成的场格点趋势预测的误差.

设变量 x_l 和 y_l 的协方差为

$$\text{cov}(x_l, y_l) = \frac{1}{n} \sum_{i=1}^n (x_{li} - \bar{x}_l)(y_{li} - \bar{y}_l) \tag{5}$$

其中, n 为观测次数, x_l 和 y_l 分别为 x_l 和 y_l 的样本平均值, x_{li}, y_{li} ($i=1,2, \dots, n$) 为变量 x_l 和 y_l 的第 i 次观测值.

由 (5) 式和实践经验, 有两种情形将导致协方差大: (1) 两变量的方差大; (2) 两变量距平同号频次多. 尽管 SVD 过程对左场实施的线性变换不会像 EOF 那样时间系数具有极大化方差, 较快收敛速度, 但极大化协方差过程仍有可能使得新变量——时间系数方差较大, 即我们要取比 EOF 多的时间系数才能完成场的估计, 毕竟为满足条件 (3) 提供了可能.

2 实例

根据经验, 1 月北半球 500hPa 度场、海平面气压场、北太平洋海温场与汛期降水场关系较好, 用这 3 场组成因子场 (右场), 预测长江中游湖北省、湖南省、江西省境内 47 代表站 6-8 月降水场 (左场). 这里 $p=47, q=576+576+286=1438$, 分析样本 1960-1997 年, 1998 年用于试验和检验.

图 1、图 2 分别给出了 1998 年 6-8 月降水场实况和预测. 大范围的偏涝区大致吻合, 只是预测偏涝的程度较实况偏小.

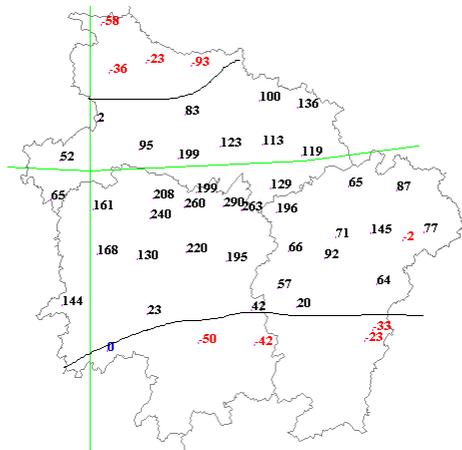


图 1 1998 年 6-8 月降水场实况 (距平%)
 Fig.1 The fields of observed precipitation during the summer 1998 (Anomaly percent)

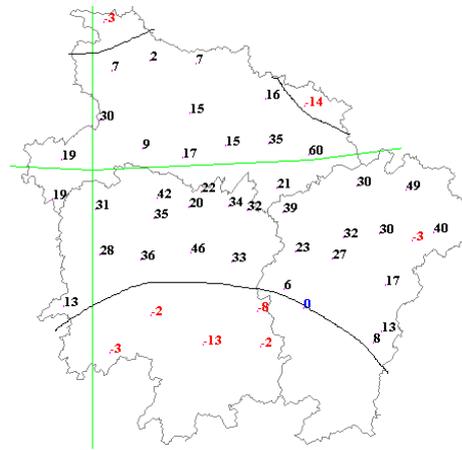


图 2 1998 年 6-8 月降水场预测 (距平%)
 Fig.2 Forecasts of fields of the precipitation during the summer 1998 (Anomaly percent)

3 结语

气象场上的资料难以逐点使用, 也不应该逐点使用, 逐点使用没有考虑相邻格点固有的物理联系, 不具备连续性, 与天气学原理相悖. 预报场与因子场 SVD 分析, 将分析预报场与

因子场逐点与逐点随时间相互变化的众多关系变为分析前 N 对时间系数随时间相互变化的简单关系, 不仅使问题大大简化, 且突出了重点. 借助最优化技术, 选择合理的系数, 建立预测公式, 由因子场时间系数预测预报场时间系数, 同时订正预报场时间系数 a_1, a_2, \dots, a_N 本身的预报误差和反演过程的分析误差造成的场格点趋势预测的误差. 截取足够多的时间系数, 保证左场时间系数与左奇异向量 (空间函数) 线性组合构成的估计场能反映原场主要特征. 最后将预测的左场时间系数与左奇异向量 (空间函数) 线性组合构成场的预测, 由此实现场对场的定量预测. 这种方法, 考虑的是场对场的主要耦合关系, 重点考虑可预报的大尺度变化, 滤去不可预报的小扰动, 对场的主要变化进行的预测, 其思路与天气学原理一致.

参 考 文 献

- 1 Prohaska J. A technique for analysis the linear relationship between two meteorological fields. *Mon Wea Rew*, 1976, 104: 1345-1353
- 2 Bretherton C S, et al. An intercomparison of methods for finding coupled patterns in climate data. *Journal of Climate*, 1992, 5(6):541-560
- 3 丁裕国, 江志红. SVD 方法在气象场诊断分析中的普适性. *气象学报*, 1996, 54 (3) :365-372
- 4 Wallace J M, et al. Singular value decomposition of wintertime sea surface temperature and 500-mb height anomalies. *Journal of Climate*, 1992, 5(6):561-576
- 5 江志红, 丁裕国. 我国夏半年降水距平与北太平洋海温异常的奇异值分解法分析. *热带气象学报*, 1995, 11 (2) :133-141
- 6 谢炯光, 秦冰冰, 王静渊. 奇异值分解方法在季降水预测中的应用. *气象学报*, 1997, 55 (1) :117-123
- 7 张礼平等. 奇异值分解方法及其在短期气候预测中的应用. *热带气象学报*, 2002, 11 (3) :133-141

Research in Forecast Technology of the Fields of Precipitation during the summer over the Middle Reaches of the Yangtze River

ZHANG Liping

(Wuhan Central Meteorological Observatory, Wuhan 430074, P.R.China)

Abstract

With SVD (singular value decomposition) enough the time coefficients of predicted fields and factor fields are cut so that their relationships represent ones between two fields and estimate of linear combination of the time coefficients and the vector of predicted fields indicate the original. With optimization technique the whole field is predicted with estimating the time coefficients of predicted field according to the coefficients of factor fields and linear combination of the coefficients and the vectors of predicted field. Predictable large scales are analyzed mainly and unpredictable disturbance are canceled. Large-scale changes of predicted fields are predicted with chief coupled relationships between two fields. Forecast of the precipitation fields during the summer 1998 year have been simulated and large areas of singular climates were predicted exactly.

Keywords: Optimization technique, singular value decomposition, estimate of field