原观测场时间序列两个统计场的相似性讨论

王盘兴 李 刚 王建新 姚文青

(南京气象学院, 南京, 210044)

摘 要

一些研究认为, 原观测场时间序列的第一特征向量场与时间平均场表现出强相似。经研究得出, 原观测场时间序列的这两个统计场间存在本质差异。它们的相似性依赖于原观测场时间序列在相空间中的结构。时间平均场(气候场)与时间偏差(异常)场集的相对强弱在二者相似与否中起到重要作用。

关键词: 原观测场时间序列, 时间平均场, 第一特征向量场, 相似性。

1 引 言

一些教科书和论文认为, 对气象学中原观测场时间序列 $\boldsymbol{F}$(以下简称场集 $\boldsymbol{F}$)作经验正交函数(简称 EOF)分析时, 其第一特征向量场(简记为 $\boldsymbol{X}_1$)与时间平均场(简记为 $\overline{\boldsymbol{F}}$)有最大的相似[1]; 或者, $\boldsymbol{X}_1$ 代表了平均场 $\overline{\boldsymbol{F}}$ 的主要信息[2]。这里, "原观测场"是相对于"距平场"或"标准化距平场"而言的。例如, 700 hPa 图上的高度、温度和风场及24 h 变高场都是原观测场。这类场的特征是时间域上的非中心化。

在以短期气候预测为目的的分析中, EOF 分析对象通常不是原观测场时间序列。但在另外一些场合, 原观测场时间序列可能成为 EOF 分析的对象; 例如, 文献[3]曾对1979年5～7月逐日热带印度洋、西太平洋区域的100 hPa, 500 hPa, 1000 hPa 层风场时间序列作了EOF 分析, 此时, 确实出现了第一特征向量场与时间平均场的很大相似。

问题是, 作为原观测场时间序列两个不同统计场的 $\boldsymbol{X}_1$ 与 $\overline{\boldsymbol{F}}$ 存在强相似是否必然?能否给出 $\boldsymbol{X}_1$ 与 $\overline{\boldsymbol{F}}$ 不相似的实例?能否对这种反例出现的条件作较一般的分析?这些问题的讨论, 将引导对气象场集在相空间中的结构差异有一个较深入的认识。

2 $\boldsymbol{X}_1$ 与 $\overline{\boldsymbol{F}}$ 的差异和它们的相似性度量

一个包含 m 个时刻的原观测场集 $\boldsymbol{F}$ 可以记为

$$\boldsymbol{F} = \{\boldsymbol{F}_t; t = 1, 2, \cdots, m\} \tag{1}$$

其中, 元素 $\boldsymbol{F}_t$ 为 t 时刻的原观测场, 它定义在 n 个格(站)点上, 记为

初稿时间: 1996年4月18日; 修改稿时间: 1997年12月9日。
资助课题: 国家自然科学基金项目49375242。

$$\boldsymbol{F}_t = (f_{t1} f_{t2} \cdots f_{tn}) \tag{2}$$

时间平均场按定义为

$$\overline{\boldsymbol{F}} = \frac{\sum_{t=1}^{m} \boldsymbol{F}_t}{m} = (\overline{f}_1 \overline{f}_2 \cdots \overline{f}_n) \tag{3}$$

第 s 格(站)点上的分量

$$\overline{f}_s = \frac{\sum_{t=1}^{m} f_{ts}}{m}, \qquad s = 1, 2, \cdots, n \tag{4}$$

而按EOF 分析一般步骤, 可求得 $\boldsymbol{F}$ 的第一特征向量

$$\boldsymbol{X}_1 = (x_{11} x_{12} \cdots x_{1n}) \tag{5}$$

这里, $\boldsymbol{X}_1$ 是最大特征值($_1$) 对应的特征向量。

下面, 根据文献[4] 的思路, 在相空间 E^n 中给出 $\boldsymbol{X}_1$ 与 $\overline{\boldsymbol{F}}$ 的差异和它们的相似性度量。

因为 $\boldsymbol{F}_t$ 可看作相空间 E^n 中一个向量(或点) , 故场集 $\boldsymbol{F}$ 可看作 E^n 中的一个包含 m 个向量(或 m 个点) 的集合。文献[4] 从相空间中向量的几何关系出发, 证明了 $\boldsymbol{X}_1$ 是 $\boldsymbol{F}$ 的全部元素在其上投影平方和达极大值的方向, 因此, 它不是完全确定的向量(允许有非零常数倍之差) 。而 $\overline{\boldsymbol{F}}$ 是起自原点、指向点集 $\boldsymbol{F}$ 质心的一个确定的向量。因此, $\boldsymbol{X}_1$ 与 $\overline{\boldsymbol{F}}$ 是两个存在本质差异的向量; 它们的相似与否应取决于 $\boldsymbol{F}$ 在 相空间中的结构。

气象学中度量两个场(例如这里的 $\boldsymbol{X}_1, \overline{\boldsymbol{F}}$) 的相似程度可用几何意义明确的相似系数

$$a_1 = (\boldsymbol{X}_1^*, \overline{\boldsymbol{F}}^*) / (\boldsymbol{X}_1^* \quad \overline{\boldsymbol{F}}^*) \tag{6}$$

度量; 因为 $\boldsymbol{X}_1$ 允许有非 0 常数(含负常数) 倍之差, 故总可以通过改变 $\boldsymbol{X}_1$ 方向使 $a_1 \geq 0$ 。或直接用 $\boldsymbol{X}_1^*, \overline{\boldsymbol{F}}^*$ 在相空间中的夹角

$$_1 = \arccos \{ (\boldsymbol{X}_1^*, \overline{\boldsymbol{F}}^*) / (\boldsymbol{X}_1^* \quad \overline{\boldsymbol{F}}^*) \} \tag{7}$$

度量。式(6) , (7) 中的 $\boldsymbol{X}_1^*, \overline{\boldsymbol{F}}^*$ 是 $\boldsymbol{X}_1, \overline{\boldsymbol{F}}$ 的空间(指地理范围) 中心化场, 定义为

$$\boldsymbol{X}_1^* = \boldsymbol{X}_1 - [\boldsymbol{X}_1], \quad \overline{\boldsymbol{F}}^* = \overline{\boldsymbol{F}} - [\overline{\boldsymbol{F}}] \tag{8}$$

式(6) ～(8) 中出现的算符意义是:

* : 空间(指地理范围) 中心化场;

(,) : 两向量的内积;

: 标量的绝对值;

: 向量的模;

[ ] : 空间(指地理范围) 平均场。

式(6) , (7) 给出的相似性度量是确定的, 当 a_1 从 1 趋于 0, $_1$ 从 0 趋于 / 2时, 描述了场 $\boldsymbol{X}_1$ 与 $\overline{\boldsymbol{F}}$ 从完全相似到完全不相似(即正交) 的状态。

3　两个实际样本的 $\boldsymbol{X}_1, \overline{\boldsymbol{F}}$ 相似性分析

从美国环境预报中心(NCEP)全球格点网客观分析资料中摘取1991年5月1日—7月31日共92 d的逐日12: 00时700 hPa格点网D上的高度场集(称$\boldsymbol{H}$场集)及24 h变高场集(称 $\boldsymbol{H}$场集)。D位于20～40 ℕ, 100～130 ℇ, 格距为 × = 5 °× 5 °。因此, 两个样本的时、空参数均为m = 92, n = 35。注意到D的位置, 它们给出了1991年由春入夏东亚700 hPa逐日环流($\boldsymbol{H}$)及其日际变化($\boldsymbol{H}$)。显然, 场集$\boldsymbol{H}$、$\boldsymbol{H}$的平均场$\overline{\boldsymbol{H}}$, $\overline{\boldsymbol{H}}$均为非零场、且均有明确的天气学意义。特别是$\overline{\boldsymbol{H}}$给出了此期间平均的日际环流型变化。按文献[4]的步骤求得了它们的第一特征向量$\boldsymbol{X}_1$(方差拟合率 = 0.999963), $\boldsymbol{X}_1$(方差拟合率 = 0.298189), 它们也均有明确的天气学意义。

图1给出了样本$\boldsymbol{H}$, $\boldsymbol{H}$的两个统计量的比较。显然, $\boldsymbol{X}_1$与$\overline{\boldsymbol{H}}$十分相似, $\boldsymbol{X}_1$与$\overline{\boldsymbol{H}}$则明显不相似。相似参数计算结果为, 场集$\boldsymbol{H}$的$a_1$ = 0.9999994, $_1$ = 6.2557 × 10^{-2}度; 场集 $\boldsymbol{H}$的a_1 = 0.1636332, $_1$ = 80.58度。可见, 场集 $\boldsymbol{H}$是一个第一特征向量场与时间平均场不相似的实例。

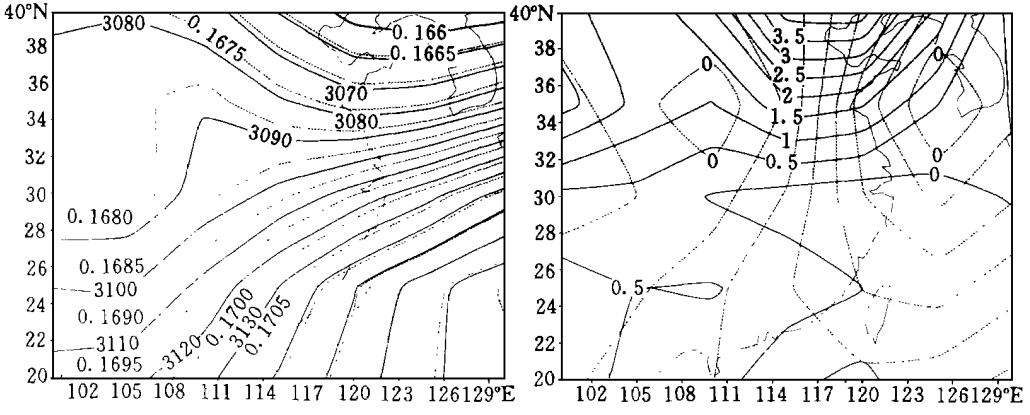


图1 第一特征向量场(虚线)与时间平均场(实线)的比较
(a. 1991年5月1日—7月31日, 12: 00时700 hPa高度场集$\boldsymbol{H}$, 实线等值线间隔10, 虚线等值线间隔0.0005, 单位是gpm。b. 24h变高场集 $\boldsymbol{H}$, 实线等值线间隔0.5, 虚线等值线间隔0.05, 单位是gpm)

4 $\boldsymbol{X}_1, \overline{\boldsymbol{F}}$相似程度与$\boldsymbol{F}$相空间结构的关系

由前面的实例, $\boldsymbol{X}_1$不相似于$\boldsymbol{F}$的场集$\boldsymbol{F}$是存在的。现将在相空间中分析$\boldsymbol{X}_1, \boldsymbol{F}$相似或可能不相似的$\boldsymbol{F}$在结构上有何差异; 分析的最终结果将给出一个判据。

对于场集$\boldsymbol{F}$及其第一特征向量场$\boldsymbol{X}_1$与平均场$\overline{\boldsymbol{F}}$, 容易给出它们在$E^n$中的示意图(图2a)。图中, 中心位于$\overline{\boldsymbol{F}}$矢端的虚线圆给出了点集$\boldsymbol{F}$对应的距平(即异常)场集$\boldsymbol{F}' = \{\boldsymbol{F}_t'; t = 1, 2, \cdots, m\}$的主要散布范围, 其特征半径为

$$R = \{\sum_{t=1}^{m} \boldsymbol{F}_t'^{\,2}/m\}^{\text{瀡}} \tag{9}$$

式中, $\boldsymbol{F}_t'$是t时刻距平场

$$\boldsymbol{F}_t' = \boldsymbol{F}_t - \bar{\boldsymbol{F}} \qquad t = 1, 2, \cdots, m \tag{10}$$

由 R 的定义知, 它是 $\boldsymbol{F}$ 异常平均强度的度量, 它在后面的相似判据中是一个重要的量。

根据式(6)、(7), $\boldsymbol{X}_1$ 与 $\bar{\boldsymbol{F}}$ 的相似程度直接取决于 $\boldsymbol{X}_1^*$ 与 $\bar{\boldsymbol{F}}^*$ 的夹角。而由式(8), 它们是 $\boldsymbol{X}_1, \boldsymbol{F}$ 与其空间平均场$[\boldsymbol{X}_1]$、$[\boldsymbol{F}]$ 之差, 因此需要在 E^n 中将向量 $\boldsymbol{X}_1(\boldsymbol{F})$ 分解为$[\boldsymbol{X}_1]$ + $\boldsymbol{X}_1^*$ ($[\bar{\boldsymbol{F}}]$ + $\bar{\boldsymbol{F}}^*$) 。因为所有空间平均场都是均匀场, 故它们对应的向量均与 E^n 中的如下直线 $\boldsymbol{Y}$ 共线

$$y_1 = y_2 = \cdots = y_n \tag{11}$$

$\boldsymbol{Y}$ 直线经过坐标原点、与所有坐标轴交角均为 arccos (1/ $\sqrt{n}$) 。空间平均场 $[\bar{\boldsymbol{F}}]$ 是 $\boldsymbol{Y}$ 上矢端位于

$$y_s = \sum_{s=1}^{n} \overline{f_s}/n \qquad s = 1, 2, \cdots, n \tag{12}$$

的向量; 因 $\boldsymbol{X}_1$ 的不确定性, $[\boldsymbol{X}_1]$ 可以是 $\boldsymbol{Y}$ 上的任一向量, 为分解方便, 取$[\boldsymbol{X}_1]$ = $[\bar{\boldsymbol{F}}]$ 。

用几何作图方法, 可直观显示 $\boldsymbol{X}_1, \bar{\boldsymbol{F}}$ 的分解: 过 $\bar{\boldsymbol{F}}$ 矢端 A 点作垂直于 $\boldsymbol{Y}$ 的平面 p (当 $n \geq 4$ 时, p 为超平面), 其方程为

$$\sum_{s=1}^{n} p_s = \sum_{s=1}^{n} \overline{f_s}$$

且与所有坐标轴交点均为 $\sum_{s=1}^{n} \overline{f_s}$, 与直线 $\boldsymbol{Y}$ 交于 P 点, 与直线 $\boldsymbol{X}_1$ 交于 B 点。向量 OB 是这里选用的第一特征向量场 $\boldsymbol{X}_1$(它通常不归一化), OA 是 $\bar{\boldsymbol{F}}$, OP 既是$[\bar{\boldsymbol{F}}]$, 又是$[\boldsymbol{X}_1]$; 而位于 p 面上的 PA, PB 分别是 $\bar{\boldsymbol{F}}^*$, $\boldsymbol{X}_1^*$, $\angle APB$ 是直接用于度量 $\boldsymbol{X}_1, \bar{\boldsymbol{F}}$ 相似程度的 $_1$。结果如图 2b 所示。

由图2揭示的几何关系, 可以根据原观测场集在 E^n 中的结构分析它们的 $\bar{\boldsymbol{F}}, \boldsymbol{X}_1$ 的相似性。例如, 最常见的原观测场集在 E^n 中的结构特点是, 它们的矢端点集相对集中地分布在远离原点的一个小区域内(例如 $\bar{\boldsymbol{F}} >> R$), 并且该区域离 $\boldsymbol{Y}$ 较远(可写为 $\bar{\boldsymbol{F}}^* >> R$) 。由图2b, 这类原观测场集的 $\bar{\boldsymbol{F}}, \boldsymbol{X}_1$ 会非常接近(指 $AB << \bar{\boldsymbol{F}}^*$), 此时 $_1$ 只能取小值, 故 $\bar{\boldsymbol{F}}, \boldsymbol{X}_1$ 必相似。假如原观测场集相应的点集充分靠近直线 $\boldsymbol{Y}$(对应 $\bar{\boldsymbol{F}} << R$) , 由图2b, $_1$ 就可能增大, $\bar{\boldsymbol{F}}, \boldsymbol{X}_1$ 就可能不相似。由此可以定义一个参数

$$K = \bar{\boldsymbol{F}}^* / R \tag{13}$$

并得到如下定性判据: 当 K 较大时, $\bar{\boldsymbol{F}}, \boldsymbol{X}_1$ 必相似; 而当 $K \to 0$ 时, $\bar{\boldsymbol{F}}, \boldsymbol{X}_1$ 可能不相似。表1 汇总了场集 $\boldsymbol{H}$, $\boldsymbol{H}$ 的相应参数。

表1　场集 $\boldsymbol{H}$, $\boldsymbol{H}$ 的参数比较

$\boldsymbol{F}$	$\boldsymbol{F}^*$	R	K	$\bar{\boldsymbol{F}}, \boldsymbol{X}_1$ 相似性
$\boldsymbol{H}$	219.57	141.44	1.552	相似
$\boldsymbol{H}$	6.74	116.88	0.058	不相似

由于 $\boldsymbol{F}^*$ 给出了气候场强度, R 给出了异常场平均强度, 故 $\bar{\boldsymbol{F}}, \boldsymbol{X}_1$ 的相似与否与场集

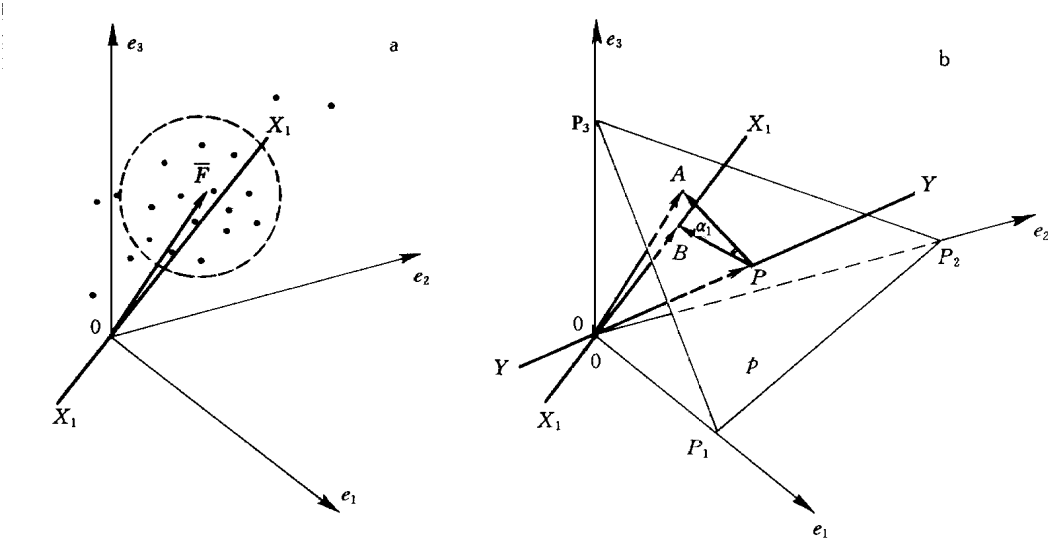


图2 $\boldsymbol{F}$ 在 E^n 中的结构及 $\boldsymbol{X}_1$, $\overline{\boldsymbol{F}}$ 分解示意图

(a. $\boldsymbol{F}$ 与 $\boldsymbol{X}_1$, $\boldsymbol{F}$ 关系。点集给出 $\boldsymbol{F}$, 直线给出 $\boldsymbol{X}_1$ 可能位置, 箭头给出 $\boldsymbol{F}$, 虚线圆(半径 R) 表示 $\boldsymbol{F}$ 点集在 $\boldsymbol{F}$ 矢端周围的散布范围; b. $\boldsymbol{X}_1$, $\boldsymbol{F}$ 的分解。箭头 OA 即a中 $\overline{\boldsymbol{F}}$, OB 是讨论选取的 $\boldsymbol{X}_1$; 直线 $\boldsymbol{Y}$ 是空间平均场向量集合, 其上虚线箭头 OP 既是 $[\boldsymbol{F}]$, 也是 $[\boldsymbol{X}_1]$; 平面 p 过 A 并垂直于 $\boldsymbol{Y}$, P, B 分别是 p 与 $\boldsymbol{Y}$, $\boldsymbol{X}_1$ 的交点, 故 PA, PB 是 $\overline{\boldsymbol{F}}^*$, $\boldsymbol{X}_1^*$, $\angle APB$ 是 $_1$)

的物理本质存在一定联系。在 $\overline{\boldsymbol{F}}^* > R$ 时, 原观测场集有一个稳定的气候场, 异常不足以过分改变气候状态, 最常见的原观测场集正好属于此类, 于是, 经常能见到 $\overline{\boldsymbol{F}}$, $\boldsymbol{X}_1$ 相似的分析结果。但也确实存在不属此类的原观测场集(如本文 $\boldsymbol{H}$), 它有可能使 $\overline{\boldsymbol{F}}$, $\boldsymbol{X}_1$ 不相似。

5 结 论

综上所述, 原观测场时间序列的时间平均场(气候场) 和第一特征向量场是两个存在本质差异的统计量。它们的相似性与场集在相空间中的结构有关, 时间平均(气候) 场与时间偏差(异常) 场集的相对强弱在决定二者相似与否中起重要作用, 当气候场强于异常场时, 二者相似; 而当异常场明显于气候场时, 二者可能不相似。

参考文献

1 苏炳凯. 大气科学中的统计诊断与预测. 南京: 南京大学出版社, 1988. 155

2 丁裕国, 施能. 气象场经验正交函数不同展开方案收敛性问题的探讨. 大气科学, 1992, 16(4): 436～443

3 王盘兴. 气象向量场的自然正交展开方法及其应用. 南京气象学院学报, 1981, 4(1): 37～48

4 王盘兴. 经验正交函数分析方法一些基本问题的讨论. 见: 长期天气预报理论和方法研究课题组编. 八五 长期天气预报理论和方法的研究, 北京: 气象出版社, 1996. 9～13

DISCUSSTON ABOUT SIMILARITY OF TWO STATISTICAL FIELDS OF THE TIME SERIES OF ORIGINAL OBSERVATIONS FIELDS

Wang Panxing Li Gang Wang Jianxin Yao Wenqing

(*Nanjing Institute of Meteorology*, *Nanjing*, 210044)

Abstract

Some studies holded that the first eigenvector field of the time series of original observations fields usually shows a striking similarity to the time-averaged field. However, this paper pointed out that there are intrinsic differences between the two statistical fields of the time series of original observations fields. Their similarity depends on the structure in the phase space of the time series of original observations fields. The relative intensity of the time-averaged(climatic) field and temporal anomalies field set plays an important role in their similarity.

Key words: Time series of original observations fields, Time-averaged fields, First eigenvector field, Similarity.