

EOF/PCA 诊断气象变量场问题的新探讨

丁裕国¹ 梁建茵¹ 刘吉峰²

¹ 中国气象局广州热带海洋气象研究所, 广州 510080

² 南京信息工程大学, 南京 210044

摘 要 进一步论证了经验正交函数/主分量分析(EOF/PCA)在气象变量场诊断中的物理内涵,证明基于 EOF/PCA 的 R 型和 Q 型展开,可描述为气象变量场主要振荡型分解和主要空间分布型分解两种方案。前者表明,气象变量场的准周期振荡可分解为各主分量的周期振荡,它们各自等价于不同网格点(或站点)以其载荷为权重的迭加周期振荡,因此,气象变量场准周期振荡可视为来自不同周期源(网格点或站点)的准周期振荡逐层叠加的结果;后者表明,气象变量场的水平空间分布可视为各种主要空间分布型的叠加,而 Q 型展开才是对各种主要空间分布型的正交分解。由此深化了 EOF/PCA 气象变量场诊断的物理内涵。

关键词 经验正交函数 主分量分析 主要振荡型分解 水平空间分布型分解

文章编号 1006-9895(2005)02-0307-07

中图分类号 P468

文献标识码 A

New Research on Diagnoses of Meteorological Variable Fields Using EOF/PCA

Ding Yu-Guo¹, Liang Jian-Yin¹⁾, and Liu Ji-Feng²

¹ Guangzhou Institute of Tropical and Marine Meteorology, China Meteorological Administration, Guangzhou 510080

² Nanjing University of Information Science and Technology, Nanjing 210044

Abstract The physics meaning of Empirical Orthogonal Function/Principal Component Analysis (EOF/PCA) is further verified for diagnoses of meteorological variable fields. EOF/PCA may shows the orthogonal resolution of principal oscillation patterns or the orthogonal decomposition of primary horizontal space distribution pattern. The former (i. e. R pattern EOF expansion) shows that the quasi-oscillation of meteorological variable fields may resolve many different oscillations which possesses different oscillation periods by means of each principal component, and the latter (i. e. Q pattern EOF expansion) shows that the horizontal space distribution of meteorological variable fields may decompose different primary space patterns. Thus, EOFs describe the complex quasi-oscillation of meteorological variable fields which is a composition of different periodic oscillations from different origins of the periods corresponding to different loading over the meteorological variable field by using each principal component.

Key words Empirical Orthogonal Function(EOF), Principal Component Analysis(PCA), resolution of mean oscillation pattern, decomposition of horizontal space distribution pattern

1 引言

经验正交函数(EOF)或主分量分析(PCA)及其推广形式,如因子分析、典型相关分析或奇异值分解、奇异谱分析、奇异交叉谱分析等等,一般都可归结为时空

场降维技术方法。多年来,这类方法一直十分活跃于大气科学及地球科学各领域,成为各学科不可缺少的重要数值分析工具。笔者曾对 EOF/PCA 及其研究成果加以综述,指出这类方法之所以能广泛应用,其本质的原因是它们具有适应各种学科领域的普适性。在数

收稿日期 2003-08-26 收到,2004-06-09 收到修定稿

资助项目 中国气象局“南海季风及其与太平洋海温相互关系的研究”项目和九五攻关(96-908-04-02)课题

作者简介 丁裕国,男,1941 年出生,教授,广州热带海洋气象研究所客座研究员,长期从事统计气候学和气候变化诊断与预测的研究,近年从事陆面过程统计—动力参数化的研究。E-mail: dygnim@jlonline.com

学上,它们具有一定的较为完整的理论框架;在物理上,它能简洁而巧妙地揭示出不同学科领域内各自的物理内涵;在计算方法和描述方式上,它又能适应于其他数学的、动力的、统计学的技巧^[1,2],然而,以往 EOF/PCA 用于气象变量场序列的诊断分析,主要是从具有时间和空间分量的时空场如何降维的观点出发加以理解应用。例如,Lorenz^[3]最早将 PCA 用于气象要素场时,仅从降低场变量的维数的观点表明用前 8 个分量已描述了原场总方差的 91%,从而说明用前 8 个分量即可代替原场 64 个变量。又如,对全球和半球气压场的分析,对大气环流场遥相关的研究,等等^[4~7],以及国内学者的研究工作^[8~10],几乎都强调了降低场变量维数的观点。近年来,我国老一辈气候学家么枕生教授对 EOF 理论和应用有所发展,提出了“载荷相关模式”并将其应用于气候分类区划,更加客观地解释了 EOF/PCA 在气象要素场分析中的作用^[11]。

从时间域来看,R 型的 EOF/PCA 实际上是一种提取气象变量场主要振荡信号型的分析方法;从空间域来看,Q 型的 EOF/PCA 可解释为主要空间分布型的分解。而这两者又可相互转化、相互依存。本文将从理论上对 EOF/PCA 展开的物理内涵作进一步探讨,以便更加准确地解释其诊断意义和作用。

2 EOF/PCA 气象变量场诊断新解

2.1 气象变量场主要振荡型分解

假定有气象变量场序列 $X = (x_1, \dots, x_h, \dots, x_p)$, 对其作 EOF 展开,写成矩阵式,即有

$$X = LY. \quad (1)$$

按惯例,对(1)式的解释是: $L = (l_1, \dots, l_p)$ 为其典型场(即特征向量,代表各个空间分布型),它们仅为空间坐标的函数; $Y = (y_1, \dots, y_p)$ 为其相应的时间权重系数(即主分量,代表各典型场随时间变化分量),它们仅为时间坐标的函数。因此,一般认为 EOF 就是将气象场 X 分解为仅与时间及空间有关的两部分的一种数据处理方法。换言之,气象变量场序列可经 EOF 展开而表示为各按一定(时间)权重加权的不同典型场之线性组合。从理论上说,EOF 本身就是多元统计分析中的主分量分析(PCA)在气象场序列中的应用,其数学模型可表述如下:

设有多元正态变量 $X = (x_1, x_2, \dots, x_p)$, 它可理解为网格点(或站点)资料序列,也可理解为任何一组具有各种学科意义的多元正态变量。本文特指气象变量场(网格点或站点)序列。以下为叙述方便,仍称为变量。假定已对其作了中心化预处理,它们等价于具

有 $E(X) = (\mu_1, \mu_2, \dots, \mu_p) = 0$ 的多元正态变量,简记为 $X \sim N(0, \Sigma)$ 。显然, $X = (x_1, x_2, \dots, x_p)$ 的协方差矩阵为

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{bmatrix}, \quad (2)$$

上式中,矩阵元素 σ_{ij} 表示第 i 变量与第 j 变量的(总体)协方差;如果存在某种线性组合,

$$Y = L'X, \quad (3)$$

并使其具有极大化方差,则总可以寻求这样一组系数向量 $L = (l_1, l_2, \dots, l_p)$, 在满足条件 $L'L = LL' = I$ 时,就有 $Y = L'X$ 的极大方差

$$\text{Var}(Y) = \text{Var}(L'X) = \text{diag}(\lambda_1, \dots, \lambda_p) = \max. \quad (4)$$

事实上,要找到这一线性组合,等价于求解矩阵方程

$$(\Sigma - \lambda I)L = 0 \quad (5)$$

的特征值和特征向量。由此可求得原变量 $X = (x_1, x_2, \dots, x_p)$ 的新组合变量 $Y = (y_1, \dots, y_p)$, 其各分量称为原变量的主分量[如(3)式]。不难证明,方程(5)的相应特征值 λ 就是主分量方差,即

$$\text{Var}(Y) = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_p). \quad (6)$$

根据矩阵正交变换理论^[8,9,11],上述各主分量为统计上互不相关的分量,而相应的系数向量(即特征向量) $L = (l_1, l_2, \dots, l_p)$ 则为解析正交向量,故可将其称为主分量正交变换。由此可见,以(1)式为代表的经验正交函数(EOF)展开,只不过是主分量正交变换的逆形式,两者并无本质区别。事实上,EOF/PCA 分析气象变量场序列时,对任一主分量而言,有

$$y_{jt} = \sum_{h=1}^p l_{hj} x_{ht} \quad (j = 1, \dots, p; h = 1, \dots, p; t = 1, \dots, n), \quad (7)$$

而对各网格点(站点)原序列,则有

$$x_{ht} = \sum_{j=1}^p l_{hj} y_{jt} \quad (h = 1, \dots, p; j = 1, \dots, p; t = 1, \dots, n). \quad (8)$$

为了进一步论证其在气象变量场诊断中的物理内涵,假定各原序列为具有某些谐振分量的正态序列,并从最简情形出发,考虑主分量的正交性,设各个主分量仅为某一中心频率的振荡分量^[9],则有简谐振振荡序

列,

$$x_{ht} = \sum_{k=-m}^m C_{hk} e^{i\omega_k t} + \varepsilon_{ht}, \quad h = 1 \dots p, \quad (9)$$

$$y_{jt} = G_j e^{i\omega_j t}, \quad j = 1 \dots p. \quad (10)$$

上式中, ε_{ht} 为服从 $N(0, 1)$ 的正态标准化白噪声序列, 若将 (9) (10) 式代入 (7) 式, 就有下列表达式

$$G_j e^{i\omega_j t} = \sum_{h=1}^p l_{hj} \sum_{k=-m}^m C_{hk} e^{i\omega_k t} + \sum_{h=1}^p l_{hj} \varepsilon_{ht}. \quad (11)$$

为了推导方便, 假定各站谐波分量的差异仅在于振幅的不同, 由于主分量的统计正交性, 上式应为关于简谐波的恒等多项式。据恒等多项式原理, 必可有若干个

$$\omega_k^* = \omega_j, \quad k^* \in m, \quad (12)$$

k^* 为小于 m 的任意正整数。使得 (11) 式化为

$$G_j e^{i\omega_j t} = \left[\sum_{h=1}^p l_{jh} C_{hk^*} \right] e^{i\omega_k^* t}. \quad (13)$$

于是, 分别又有表达式

$$G_j = \left[\sum_{h=1}^p l_{jh} C_{hk^*} \right] \quad (14)$$

和

$$e^{i\omega_j t} = e^{i\omega_k^* t}. \quad (15)$$

(8)~(15) 式表明, 各网格点 (站点) 原序列如果隐含若干个谐振分量, 则其主分量必然为具有某一中心频率的振荡分量, 而这一中心频率正是从各原变量序列中提取的频率信号, 其振幅是以前主分量荷载 l_{jh} 为权重的各原变量序列同一频率谐振振幅的加权平均振幅。

若对 (13) 式两边同乘 $y_{jt}^* = G_j e^{i\omega_j^* t}$, 并取期望运算, 就有

$$G_j^2 = \lambda_j = \sum_{h=1}^p l_{hj} C_{hk^*} G_j, \quad (16)$$

(16) 式本质上代表了气象场变量中某一区域原变量序列谐振分量的平均情况。根据文献 [11], 在标准化变量条件下, 荷载值 l_{hj} 代表了主分量振型与原序列振型的相似程度 (以相关系数为度量)。换言之, EOF/PCA 实际上是从原变量场序列中逐一提取代表原变量场的各主要振荡信号型。即主分量可以解释为气象场序列振荡型的一种分解, 不同的分量序列对应着时间域上不同的振荡型, 而其空间函数数值大小则代表了相应振荡分量与哪些地理区域密切相关, 表明这种振荡型主要来源于哪些区域 (格点或站点), 或各站点原序列对该振荡型 (主分量) 的贡献。在此基础上的分型区划实际代表了气象变量场序列的各种振荡源的空间分布区域, 而不是对气象要素场数值在水平地理空间上分布的分型区划 (后文专门对此作出定义)。有鉴于

此, 以往一般认为 EOF 是“将气象场 X 分解为仅与时间及空间有关的两部分”, 或认为气象变量场序列经 EOF 展开后表示成“各按一定 (时间) 权重加权的不同典型场 (即特征向量) 之线性组合”等等提法, 仅仅是从时空场降维意义上来解释 EOF, 而并未从气象变量场在时频域上振荡的物理本质来理解, 从这个意义上说, 以往的提法并不全面。

将 (8)~(16) 式推广为一般情况, 假定各网格点 (站点) 原序列为隐含谐振分量的正态序列, 而其主分量亦为具有某种谐振分量的正态序列, 则其相应频率的幅谱可表示为

$$Y(\omega) = \int_{-\infty}^{\infty} y_{jt} e^{-i\omega t} dt, \quad (17)$$

$$X(\omega) = \int_{-\infty}^{\infty} x_{ht} e^{-i\omega t} dt, \quad (18)$$

根据傅立叶变换, 将以上两式的反变换式代入 (7) 式中, 经整理后即得

$$Y(\omega) = \sum_{j=1}^p l_{hj} X(\omega). \quad (19)$$

根据功率谱分析理论 [12], 各主分量方差可表示为相应的谱展式,

$$\begin{aligned} \lambda_i &= \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T y_j^2(t) dt = \\ &= \int_0^\infty G_{y_j}(f) df = \sum_{h=1}^p \int_0^\infty l_{hj}^2 G_h(f) df, \\ &j = 1 \dots p; h = 1 \dots p. \end{aligned} \quad (20)$$

因此, 对照 (7) 式和 (16) 式, 各主分量还可写成相应的功率谱分解式,

$$\int_0^\infty G_{y_j}(f) df = \sum_h \int_0^\infty l_{hj}^2 G_h(f) df, \quad (21)$$

或有

$$G_{y_j}(f) = \sum_h l_{hj}^2 G_h(f), \quad j = 1 \dots p; h = 1 \dots p \quad (22)$$

由此可见, 主分量功率谱是各原变量功率谱的加权线性和, 即原变量分解的线性组合谱, 其权重为各变量的荷载平方值。从原变量功率谱及其权重配置的结构中不难估计出控制某气象变量场的主要振荡周期信号。换言之, 从频率域来看, EOF/PCA 实质是对场的主要振荡型作正交分解, 从而将一个具有复杂振荡系统的场序列分解成若干个较为单一的振荡分量系统。另一方面, 从线性定常随机系统的观点来考察 EOF/PCA, 同样可以证明, 它们只不过是线性定常随机系统的一种特例。根据系统在时域上的输入输出关系 [13], 一般有

$$y_j(t) = \sum_{i=1}^p \sum_{\tau=-\infty}^{\infty} g_{ji}(\tau) x_i(t-\tau),$$

$$j = 1 \dots, q; q \leq p; \tau = 0, \pm 1, \pm 2, \dots, \pm \infty \quad (23)$$

式中 $x_i(t-\tau)$ 为具有时间后延的输入变量序列, $y_j(t)$ 为输出变量序列, 而 $g_{ji}(\tau)$ 则为对应于输入(输出)的频率响应函数。对照(7)式, 不难看出, 当 $\tau=0$ 时, 即输入变量序列对于输出变量序列在没有时间后延的条件下, 频率响应函数仅为常向量, 则(23)式退化为(7)式, 而相应的频率响应函数就是其荷载向量(即特征向量)。

值得一提的是, 这里所谓主要振荡型正交分解与由 Hasselmann^[13]在 20 世纪 90 年代提出的主振荡模态 (POP) 方法是不完全相同的两种概念。因为这里未加进原变量的时滞相关过程, 但至少有一点是共同的, 即都是对场的振荡型态作分解。由此, 从这个意义上也佐证了上述观点的正确性。

2.2 气象变量场主要空间分布型的分解

如上所述的 EOF/PCA 展开, 主分量可解释为气象场序列的振荡型分解, 不同的分量序列对应着时间域上彼此互不相关的振荡型, 而其空间函数数值大小(即荷载值)则代表了相应振荡分量与哪些地理区域密切相关, 在此基础上的分型区划仅代表了各种振荡型的空间分布区划, 而不是针对气象变量场水平地理空间分布的分型区划, 这一点在以往的许多应用中并未引起学者们的注意, 以致于易将两者混为一谈。为了区别这两种不同性质的区划, 我们将前述的 EOF/PCA 展开方案称之为 R 型展开, 而针对气象变量场水平地理空间分布的分型区划, 则表述如下:

假定有气象变量场序列 $X = (x_1, \dots, x_i, \dots, x_p)$, ($i = 1, \dots, p$), 对其作转置, 就有

$$X_q = X', \quad (24)$$

仿 R 型展开, 可写出

$$X_q(t) = L_q Y_q(s), \quad (25)$$

$$Y_q(s) = L_q' X_q(t). \quad (26)$$

这里, (26)式就可称之为 Q 型 EOF/PCA 展开, 其中 s 代表空间坐标, t 代表时间坐标; 主分量 $Y_q(s) = (y_1^{(q)}, \dots, y_p^{(q)})$ 的每一分量 $y_j^{(q)}$ 即代表原气象变量场的某种典型的水平空间分布型。而其特征向量 $L_q = (l_1, \dots, l_n)$ 的每一分量 l_j 则代表其相应的主分量在时间坐标上的荷载值。从理论上可以证明, 由于空间分布型(主分量)相互正交, (25)式代表了各个不同时刻

相应的水平空间分布图形的线性组合, 其权重是与时间坐标有关的特征向量 $L_q' = (l_1, \dots, l_n)$ 。显然, 对于各主分量 $Y_q(s) = (y_1^{(q)}, \dots, y_p^{(q)})$, 可用一组正交的空间函数来定量描述, 例如, 假定它们可用不规则格点切比雪夫正交多项式来表示^[14], 即有

$$Y_h^{(q)} \approx \hat{Y}_h(s_1, s_2) = \sum_0^{k_0} \sum_0^{s_0} A_{ks}(h) \varphi_k(\xi_{1i}) \psi_s(\xi_{2j}), \quad (27)$$

上式 $s = (s_1, s_2)$ 表示水平空间坐标; $\varphi_k(\xi_{1i}) \psi_s(\xi_{2j})$ 为经非线性变换后的二维切比雪夫正交多项式, 其中坐标 ξ_{1i}, ξ_{2j} 分别代表站点的二维序号。同理, 对于各原变量 $X_q(t) = (x_1^{(q)}, x_2^{(q)}, \dots, x_n^{(q)})$, 也可写出不规则格点切比雪夫正交多项式表达式, 为方便, 以下将代表 Q 型展开的上标^(q)一律省略。

$$X_t^{(q)} \approx \hat{X}_t(s_1, s_2) = \sum_0^{k_0} \sum_0^{s_0} A_{ks}(t) \varphi_k(\xi_{1i}) \psi_s(\xi_{2j}). \quad (28)$$

利用(25)、(26)式并结合(17)~(19)式, 可以证明

$$\sum_0^{s_{10}} \sum_0^{s_{20}} A_j(s_1, s_2) \varphi_{s_1}(\xi_{1i}) \psi_{s_2}(\xi_{2j}) = \sum_1^n l_{hj} \sum_0^{s_{10}} \sum_0^{s_{20}} A_h(s_1, s_2) \varphi_{s_1}(\xi_{1i}) \psi_{s_2}(\xi_{2j}). \quad (29)$$

同样, 根据恒等多项式原理, 对给定的坐标 $s_1 = s_1^*$, $s_2 = s_2^*$, 可以证明

$$A_j(s_1^*, s_2^*) = B_j(s_1^*, s_2^*), \quad (30)$$

其中,

$$B_j(s_1^*, s_2^*) = \sum_{h=1}^n l_{hj} A_h(s_1^*, s_2^*). \quad (31)$$

上式表明, 若主分量(场)被描述为一种正交函数(如切比雪夫正交多项式), 其相应的系数正是各原变量(场)正交函数(切比雪夫正交多项式)系数的加权线性和。根据文献[14], 切比雪夫正交多项式系数表明各正交多项式(即所对应的几何图形)对于原变量(场)总方差的方差贡献大小。由此可见, 基于切比雪夫正交多项式描述的主分量(场)正是原变量(场)各水平空间分布型方差贡献之线性组合, 其权重是各变量(场)的荷载值(特征向量)。从原变量场的组合分布型及其权重配置结构中不难估计控制某气象变量场的主要水平空间分布型(某种正交函数)。可以想象, 这与 R 型展开所得主要振荡型分解十分类似。

显然, (29)式表明, 空间主分量与原气象场各不同时刻相应的水平空间分布的相似系数, 实质上就是

与各空间主分量相对应的分布在时间坐标上的荷载值。换言之, EOF/PCA 的 Q 型分解, 等价于从各个不同时刻的水平空间分布场中, 提取典型的水平空间分布型, 而在时间坐标上的荷载值, 等价于它们和各个不同时刻空间分布场之间的相似系数。因此, Q 型分解意味着在逐个时次中, 提取有代表性的水平空间分布型, 作为其典型的水平空间分布型(空间型主分量), 而其在时间坐标上的荷载值正好表明各水平空间分布型来源于哪些时段。显然, 从水平空间分布型的意义上诊断气象变量场序列, 可借助于这种 Q 型分解, 将气象变量场序列看作各个不同的典型水平空间分布型的线性组合, 其权重就是它们在时间坐标上的荷载值。

2.3 时空分布的相互转换——EOF/PCA 展开的双重意义与物理解释

由于空间点数与时间点数并不相同, Q 型分解与 R 型分解所得主分量数并不一样, 因此, 某气象变量场序列究竟能分解成几种互不相关的水平空间分布型, 并不能从通常的 R 型 EOF/PCA 展开得到, 而只能借助于 Q 型 EOF/PCA 展开。但是, 理论上已经证明 R 型展开与 Q 型展开(因 AA' 和 $A'A$ 具有相同的非零特征根)的求解可以相互转换^[8,15]。即它们在数学上具有双重意义, 而在物理解释上则有不同的内涵。由于 R 型展开和 Q 型展开可以相互转换, 则还可证明, 对 Q 型展开来说, 非规一化的 R 型展开的特征向量正是 Q 型展开的主分量。

综上所述, EOF/PCA 展开具有双重物理意义, 对 R 型展开, 应看作气象场序列随时间变化的主要振荡型分解(如前所述), 对于 Q 型展开, 则应看作气象变量场随水平地理空间而变化的主要水平空间型分解。从理论上说, 前者的主分量本质上是后者的特征向量, 而后者的主分量本质上是前者的特征向量。

表 1 试验周期(T)与振幅(G)

Tabel 1 Test periods (T) and their amplitudes (G)									
格点序号 Grid No.	1	2	3	4	5	6	7	8	9
T	3.0000	3.0000	3.0000	7.0000	7.1000	14.9000	15.1000	15.2000	15.5000
G	4.0000	4.5000	5.0000	7.0000	8.0000	5.0000	9.0000	10.0000	11.0000

表 2 各主分量最大熵谱值及其周期(括号内的数值)

Tabel 2 The spectrum values from MESM for each principal component and its periods (the values in brackets)

PC 方差贡献(%) Variance contribution(%)	39.97	33.31	19.41
PC1	277.99(15)		
PC2	226.83(3.0)		
PC3	141.61(7.0)		

3 简单模拟试验

3.1 主要振荡型分解(R 型展开)

构造具有不同谐振分量(例如含 9 个网格点)的仿真变量场序列(序列长度 $N = 100$)。为简单起见, 假定随机生成的各格点仅含有一种谐振周期的场序列(如表 1)。由表 1 可见, 整个场序列实际上共有三种基本周期($T = 3, 7, 15$), 而各自的振幅有一定差异。对其作 R 型 EOF/PCA 得到各主分量(即所谓的 EOF 时间系数)及其荷载向量(即所谓的典型场)如表 2 和表 3 所示。由表 2 可见, 方差贡献最大的主分量 PC1 (39.97%), 其最大熵谱(MES)峰值在 $T = 15$, 而对应的显著的高荷载值恰好在格点 6, 7, 8, 9(分别为 0.48, 0.52, 0.52, 0.46), 主分量 PC2 (33.31%), 其 MES 峰值在 $T = 3$, 对应的显著的高荷载值恰好在格点 1, 2, 3(分别为 0.58, 0.58, 0.58), 主分量 PC3(19.41%), 其 MES 峰值为 $T = 7$, 对应的显著高荷载也正好在 4, 5 格点, 前三个主分量所占方差已达 92.68%, 恰好对应着三种基本周期。换言之, 它们分别代表该场序列有三种主要振荡型, 而这些主要振荡型分别来源于哪些站点或区域, 可从表 3 的各荷载场及其取值分布情况一目了然。例如, 第 1 主分量(第一振荡型)的高荷载格点序号为 6, 7, 8, 9, 而第 2 主分量(第二振荡型)的高荷载格点序为 1, 2, 3, 第 3 主分量(第三振荡型)的高荷载格点序号为 4, 5, 相当有规律, 这一试验结果正好与(22)式吻合。利用该式计算各主分量谱得到了完全一致的结果。

3.2 水平空间分布型分解(Q 型展开)

作 Q 型 EOF/PCA 所得各主分量正是各个典型水平空间分布型, 其荷载向量则对应于各时刻(逐日)。表 4 列出前 3 个主分量(相应的前 3 个空间分布型, 图

略), 它表明 Q 型分解就是在逐个时次(逐日)的水平空间分布场型中, 提取那些有代表性的水平空间分布场型, 作为其典型的水平空间分布型(即空间型主分量), 而其逐个时次上的荷载值正好表明各典型水平空间分布型来源于哪些时段。从表 4 可见, 其前 3 个主分量取值的格点分布大致与表 3 所列的 R 型荷载值分布趋势相似, 尤其是 PC1, PC2 与 R 型荷载值分布的高低值中心最为吻合。这说明 EOF/PCA 的双重物理

表 3 各荷载场及其取值分布

Table 3 The distributions of the loading fields and their values

格点序号 Grid No.	1	2	3	4	5	6	7	8	9
l_1	-0.04	-0.04	-0.04	-0.03	-0.02	0.48	0.52	0.52	0.46
l_2	0.58	0.58	0.58	-0.01	-0.01	0.03	0.03	0.03	0.03
l_3	0.01	0.01	0.01	0.71	0.71	0.03	0.02	0.02	0.00

表 4 Q 型 EOF 的前 3 个主分量及其取值分布

Table 4 The first three principal components and its values for Q-EOF

格点序号 Grid No.	1	2	3	4	5	6	7	8	9
PC1	-5.548	-5.613	-5.679	-6.93	-6.687	1.053	7.989	10.072	11.342
PC2	5.283	5.612	5.941	-7.12	-9.582	1.687	0.176	-0.447	-1.549
PC3	-1.06	-1.355	-1.65	0.064	-0.308	5.516	4.279	1.777	-7.262

意义及其解释是合理的。

4 实测气象变量场的诊断举例

4.1 R 型 EOF 展开举例

对均匀分布于我国区域 203 个测站的(2000 年)夏季逐日高温资料场序列(92 天)作 R 型 EOF 展开。结果表明,前 10 个主分量累积方差贡献已达 90%,滤去季节趋势,则前 10 个主分量累积方差贡献可达 50%。将各个主分量功率谱与各相应荷载场高值区的原变量平均序列或其第 1 主分量的功率谱对照,不难发现两者基本吻合。这就表明,各主分量振荡型正是以原变量场序列的某一子场序列振荡型为基础的,图 1 所绘制的是几个主分量与相应子场平均序列的谱图对照,由图可见,主分量谱峰基本上与相应子场序列的谱峰对应。可见,R 型 EOF 展开是气象场序列振荡型的一种分解,不同主分量序列对应着时间域上不同的(互不相关)振荡型,而其空间函数数值大小代表了相应振荡分量与哪些地理区域密切相关,表明这种振荡型源自哪些区域(格点或站点)。换言之,气象变量场整体的振荡可视为来自不同周期源的准周期振荡逐层叠加的结果,而某个站(网格)点的时间序列所形成的复杂周期振荡正是不同来源振荡的叠加。

4.2 Q 型 EOF 展开举例

对上述资料作 Q 型 EOF 展开,表明在逐个时次的水平空间分布场型中,提取那些有代表性的水平空间分布场型作为其典型水平空间分布型(即空间型主分量),而其时间坐标上的荷载值正好表明各典型水平空间分布型来源于哪些时段。表 5 则为对应的各时刻(日)分布型与这些典型水平空间分布的相似系数达到显著性的数量(日数)。

值得指出的是,主分量方差贡献,反映出某一主分量对于场序列总方差的贡献大小,它并不表明该主分量所代表的原变量数目的多少或区域的大小。就 R

型来说,主分量方差贡献反映出某一振荡型在总振荡中的贡献,而这一振荡型究竟与哪些区域有关及其区域大小与此并无直接关系,这一振荡型的强弱与否并不取决于原变量场的某一局部区域(荷载),而是取决

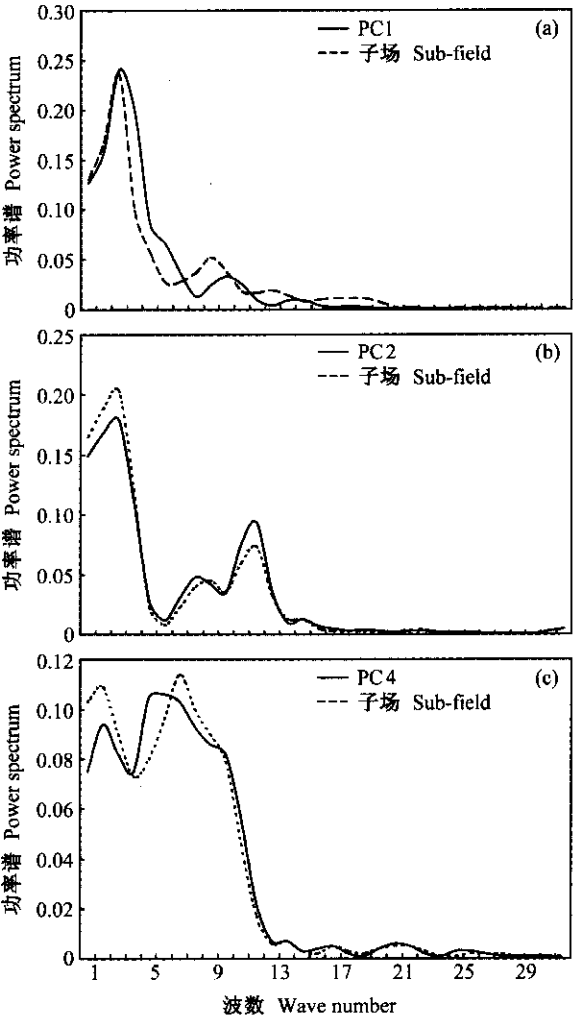


图 1 主分量谱与子场序列谱:(a) PC1;(b) PC2;(c) PC4
Fig. 1 The spectrums for the principal component and the sub-fields series:(a) PC1;(b) PC2;(c) PC4

表 5 逐日分布型与典型水平空间分布的相似系数达显著性 (信度 0.05) 的日数

Tabel 5 The number of days for the significant similarity coefficient between daily distribution patterns and typical horizontal space distribution (the degree of confidence is 0.05)

第 1 型	第 2 型	第 3 型	第 4 型	第 5 型
PC1 pattern	PC2 pattern	PC3 pattern	PC4 pattern	PC5 pattern
36	33	12	12	40

于分量所代表的某一振荡型的总能量(方差),即主分量时间序列的平均振幅。就 Q 型来说,主分量方差贡献反映出某一空间分布型在总空间分布中的贡献,空间分布型的强弱与否取决于该空间型(主分量) 的总能量即空间方差大小。而这一空间分布型究竟与哪些时段有关以及时段的长短也与方差贡献并无直接关系。因此,各型的相似时段多少并不与主分量次序一致。

5 结论

(1) 论证了 EOF/PCA 在气象变量场诊断中的物理内涵,基于 EOF/PCA 的 R 型展开是气象场主要振荡型分解,而对气象变量场序列的 Q 型展开才是对气象场序列的水平空间分布状况所作的主要空间分布型态的分解。其物理内涵包括: 1) 气象变量场序列整体的准周期振荡可视为是来自不同周期源的准周期振荡的叠加,其中某个站(网格) 点的时间序列所形成的复杂周期振荡正是不同来源振荡的叠加结果; 2) 由于 R 型和 Q 型展开可以相互转换,其结果在数学上具有双重意义,而在物理解释上则有不同的内涵; 3) 通常在实际应用中,首先应确定是 R 型展开还是 Q 型展开,而不能片面理解 EOF/PCA 展开结果,并在物理内涵上作出符合实际的相应解释。

(2) 数值试验表明,上述理论论证与实际结果相符。即 R 型展开,所得各主分量正是原变量场中某一子场序列的振荡型为主要振荡源(或主要振荡源来自某一地理区域) 的结果,这就奠定了可从原变量场序列(复杂的振动系统) 的整体来考察其主要振荡源的理论基础。类似的问题,同样适用于 Q 型展开。

(3) 对于以 EOF/PCA 为基础的其他推广形式或派生方法,根据上述理论,原则上都可深入探讨,我们将在另文叙述。

参考文献

[1] 丁裕国. EOF 在大气科学研究中的新进展. 气象科技, 1993, 3 : 10 ~ 19
Ding Yuguo. Advances of EOF in study of atmospheric science. Meteor. Science and technique (in Chinese), 1993, 3 : 10 ~ 19

[2] 周家斌,黄嘉佑. 近年来中国统计气象学的进展. 气象学报, 1997, 55(3): 297 ~ 305
Zhou Jiabin, Huang Jiayou. Advances in statistical meteorology in China in recent years. Acta Meteor. Sinica (in Chinese), 1997, 55(3): 297 ~ 305
[3] Lorenz E N. Empirical Orthogonal Functions and statistical weather prediction. Sci. Rept., No. 1, Statistical Forecasting Project, Mass. Inst. Tech., Dept. of Meteorology, Cambridge, Mass., 1956, 49 ~ 50
[4] Wallace J M, Gutzler D S. Teleconnections in the geopotential height field during the Northern Hemisphere winter. Mon. Wea. Rev., 1981, 109 : 784 ~ 812
[5] Trenberth K E, Paolio D A. Characteristic patterns of variability of sea level pressure in the Northern Hemisphere. Mon. Wea. Rev., 1981, 109 : 1169 ~ 1189
[6] Parnett T P, Preisendorfer R W. Multifield analog prediction of short term climatic fluctuations using a climatic state vector. J. Atmos. Sci., 1987, 35 : 1771 ~ 1787
[7] Preisendorfer R W. PCA in Meteor. and Oceanography. Elsevier Sci. Publishers, Amsterdam, Oxford, New York, 1988. 425pp
[8] 黄嘉佑. 气象统计分析与预报方法. 北京:气象出版社, 1990. 182 ~ 188
Huang Jiayou. Meteorological Statistical Methods in Analysis and Prediction (in Chinese). Beijing : China Meteorological Press, 1990. 182 ~ 188
[9] 么枕生,丁裕国. 气候统计. 北京:气象出版社, 1990. 624 ~ 629
Yao Zhensheng, Ding Yuguo. Climatic Statistics (in Chinese). Beijing : China Meteorological Press, 1990. 624 ~ 629
[10] 周家斌. 关于统计气象学的几个问题. 大气科学, 1993, 17(增刊): 1 ~ 6
Zhou Jiabin. On problems of statistical meteorology. Scientia Atmospherica Sinica (in Chinese), 1993, 17(Suppl.): 1 ~ 6
[11] 么枕生. 载荷相关模式用于气候分类与天气气候描述. 气候学研究——气候与环境. 北京:气象出版社, 1998. 1 ~ 9
Yao Zhensheng. A loading correlation model for climatic classification with climatic descriptions in synoptic climatology. Climatology Studies—Climate and Environment (in Chinese), No. 5. Beijing : China Meteorological Press, 1998. 1 ~ 9
[12] 丁裕国,江志红. 气象数据时间序列信号处理. 北京:气象出版社, 1998. 133 ~ 166
Ding Yuguo, Jiang Zhihong. Signal Process of the Time Series of Meteorological Data (in Chinese). Beijing : China Meteorological Press, 1998. 133 ~ 166
[13] Hasselmann K. PIPs and POPs : The reduction of complex dynamical systems using principal interaction and oscillation patterns. J. Geophys. Res., 1988, 93(D9): 11015 ~ 11021
[14] 周家斌. 车贝雪夫多项式及其在气象中的应用. 北京:气象出版社, 1990. 262 ~ 264
Zhou Jiabin. Chebyshev Polynomials and Its Applications in Meteorology (in Chinese). Beijing : China Meteorological Press, 1990. 262 ~ 264
[15] Murihead R J. Aspects of Multivariate Statistical Theory. John Wiley & Sons, Inc. 1982. 673pp