

# 非均匀站网 EOFs 展开的失真性及其修正\*

丁裕国 江志红

(南京气象学院, 南京, 210044)

## 摘 要

从理论上证明,非均匀站网 EOFs 展开有不同程度的失真。在同等条件下,均匀站网与非均匀站网 EOFs 展开结果有明显差异。提出了一种附加面积权重的修正方案用以弥补非均匀站网 EOFs 展开的失真现象。实例计算表明,中国气温场(160 站)经修正 EOFs 展开后,其气温变化主分量趋势与特征能更加客观地揭示中国地区增暖效应的局地差异。

**关键词:** 经验正交函数(EOFs),非均匀站网,面积权重,气温场变化。

## 1 引言

通常 EOFs 展开是在任意选站的非均匀分布站网区域上进行的。这种非均匀性对于 EOFs 展开结果是否有影响,理论上是否合理,值得推敲。对此,许多学者都曾作过有益的探讨,例如:Fortus (1975)<sup>[1]</sup>, Buell (1972, 1978)<sup>[2,3]</sup>, Karl 等 (1982)<sup>[4]</sup>, Preisendorfer (1988)<sup>[5]</sup>先后指出,选点的任意性,对 EOFs 结果有影响;通常对空间点离散的场作 EOFs 可视为对空间上连续的场进行正交分解的特例;非均匀网点必须内插成等面积网点后再作 EOFs 展开才能避免产生失真的特征向量场。文献[6]对北半球平均表面温度序列资料处理时,也考虑了站网分布的非均匀性订正问题。文献[7]提出了一种考虑面积因素的平均气温计算方案用以客观反映中国区域气温真实变化情况。

本文欲从理论上证明非均匀站网 EOFs 展开结果的失真性,比较均匀站网与非均匀站网 EOFs 展开的差异。由此出发,提出一种附加面积权重的 EOFs 修正方案,并以实例证明其可行性。

## 2 理论推导

设任一区域的气象场表为  $X(w, t)$ , 将其视为复杂的随机过程, 其中  $w$  代表样本空间的现实(它等价于事件, 如  $w = n$  则为第  $n$  次观测的事件或取值);  $t$  代表过程的载体参数, 例如时间或空间上的参数, 这里约定,  $t$  代表空间变量, 特别是  $t = (x, y)$  代表二维空间的坐标。一般, 对函数  $X(w, t)$  可作如下正交分解

$$X(w, t) = \sum_k Z_k(w) \varphi_k(t) \quad (1)$$

\* 1993年4月26日收到原稿,1993年9月16日收到修改稿。本文得到自然科学基金资助。

通常视函数  $\varphi_k(t)$  为非随机函数, 它依所选积分域  $(a, b)$  而定, 又称为特征函数, 它满足解析正交性, 即有

$$\int_a^b \varphi_k(t) \varphi_j(t) dt = \delta_{kj} \lambda_k \quad (2)$$

而  $Z_k(\omega)$  则为统计上正交或不相关, 它满足

$$E | Z_j(\omega) Z_k(\omega) | = \delta_{kj} \lambda_k \quad (3)$$

上两式中,  $\delta_{kj}$  为狄拉克函数,  $\lambda_k$  为随下标  $k$  变化的常数。对式(1), 系数  $Z_k(\omega)$  可反演为

$$Z_k(\omega) = \int_a^b X(\omega, t) \varphi_k(t) dt \quad (4)$$

上式表明, 函数  $Z_k(\omega)$  就是各现实  $X(\omega, t)$  在第  $k$  个特征函数  $\varphi_k(t)$  上的投影。根据随机函数理论, 定义相关函数

$$R(t, t') = E | X(\omega, t) X(\omega, t') | \quad (5)$$

对  $\varphi_k(t)$  构造积分函数

$$\int_a^b R(t, t') \varphi_k(t') dt' = \int_a^b E | X(\omega, t) X(\omega, t') | \varphi_k(t') dt' \quad (6)$$

在“均方意义”下, 借助于式(4), 可推得

$$\int_a^b R(t, t') \varphi_k(t') dt' = E | X(\omega, t) Z_k(\omega) | \quad (7)$$

利用式(1), 将式(7)改写为

$$\int_a^b R(t, t') \varphi_k(t') dt' = E \left| \sum_j Z_j(\omega) \varphi_j(t) Z_k(\omega) \right| \quad (8)$$

根据式(3)的正交性, 当  $k = j$  时,  $E | Z_k^2(\omega) | = \lambda_k$ , 而当  $k \neq j$  时,  $E | Z_j(\omega) Z_k(\omega) | = 0$ , 故有

$$\int_a^b R(t, t') \varphi_k(t') dt' = \lambda_k \varphi_k(t) \quad a \leq t \leq b \quad (9)$$

上式称为 Fredholm 积分方程(简记为 Fr 方程), 其中  $R(t, t')$  又称为半正定对称相关核, 而  $\varphi_k(t)$  称为它的特征函数,  $\lambda_k$  为其特征值。

根据积分方程理论<sup>[8]</sup>, 具有对称核的齐次 Fr 方程的特征函数为正交函数, 且它的特征值亦为实数。由于方程(9)恰好为齐次 Fr 方程, 并有式(2)和(3)成立, 所以, 从式(1)出发所获  $\varphi(t)$  正是式(9)所表示的 Fr 方程之解。通常这个函数并非唯一, 它依赖于相关核  $R(t, t')$  与  $t$  的域  $(a, b)$ , 例如, 在一维情况下, 它可为一个线性区间, 在二维情况, 它可为一个闭合区域, 本文讨论二维情况。显然, 若式(9)经离散化抽样后, 可有近似积分式, 一般说来, 方程(9)的积分近似可有各种形式, 式(10)仅是其一种近似表示。

$$\sum_{j=1}^p R(t^i, t^j) \varphi_k(t^j) \Delta t^j = \lambda_k \varphi_k(t^i) \quad (10)$$

式中  $\Delta t^j$  就表示位于  $t^j (= (x_j, y_j))$  的测站的面积元。广义地说, 气象场序列可视为时空域连续的随机过程  $X(\omega, t)$  的离散化取样结果, 一般说, 时间上的离散化都以等间隔观测值表示, 对于空间上的离散化, 式(9)就可化为式(10), 假定测点为均匀网格分布, 其面积

元  $\Delta t^j$  可视为单位常量。

众所周知,通常 EOFs 展开气象场是将包含有  $P$  个空间点(变量)的场序列资料作时空正交分解,即对资料阵  $X'_{p \times n}$  作正交分解,使任一空间点  $j$  和任一时间点  $i$  的观测值(或距平观测值)  $X_{ij}$  分解为  $P$  个时间函数  $T_{ik}$  和空间函数  $\psi_{jk}$  的线性组合

$$X_{ij} = \sum_k T_{ik} \psi_{jk} \quad i = 1, 2, \dots, n, \quad j = 1, 2, \dots, p \quad (11)$$

对于时间函数,其反演式为

$$T_{ik} = \sum_{j=1}^p x_{ij} \psi_{jk} \quad i = 1, 2, \dots, n, \quad j = 1, 2, \dots, p \quad (12)$$

而空间函数则可由式(13)求得

$$\sum_{j=1}^p R_{jl} \psi_{jk} = \lambda_k \psi_{lk} \quad k, l = 1, 2, \dots, p \quad (13)$$

式(13)中  $R_{jl}$  表示任一空间点  $j$  和  $l$  之间的协方差函数,对照式(1),(4),(9)不难看出,式(11),(12),(13)正是它们的特例。尽管 EOFs 展开的推导是基于矩阵理论,但它本质上是考虑时间上离散、空间上离散情形下气象场函数的时空正交分解。因此 EOFs 展开可视为式(1)~(9)所代表的理论模型的特例。换言之,按照式(9)或式(10),则式(13)应为在均匀站网假定下求解特征向量的特例。然而,通常 EOFs 推导,实际上只对时间域作了等距限制,而对空间点的选取并无任何均匀性限制(空间点选取是任意的),这样势必会产生失真性<sup>[5]</sup>。作者认为,弥补这一问题的办法,一是将非均匀站网转换成均匀站网,二是附加面积权重来消除非均匀站网的影响。但前者运算量大,特别是在站网严重不均情况下,处理困难;后者虽不如前者精确,但毕竟运算方便。因此,本文提出一种简单的面积加权修正 EOFs 展开方案。

设已知自然区域总面积为  $Q$ ,它由各子区域(如省、市、县)构成。若每个子区域面积为  $Q_r$  ( $r = 1, \dots, R$ ),一般地说,每个子区域选站数并不相等,设为  $m_r$ ,则全区域内总共有站点数  $m$ ,即  $m = \sum_{r=1}^R m_r$ ,对每个子区域,第  $i$  个测站的面积权重  $a_i$  ( $i = 1, \dots, m_r$ ) 可定义为

$$a_i = \frac{Q_r/Q}{m_r/m} = \frac{m_r Q_r}{m_r Q} \quad i = 1, \dots, m_r, \quad r = 1, \dots, R \quad (14)$$

上述定义实际上是假定了各个子区域内站点为均匀分布,当然这是一种近似。区域划分愈小,其误差愈小。在一定程度上,它对于修正站网分布上的非均匀性是有效的。不难证明,经面积加权修正后,气象场 EOFs 展开结果将有明显改变。

对某子区域记录  $x_{ij}$ , 面积加权后写为

$$y_{ij} = a_j x_{ij} \quad (15)$$

这里下标  $j$  对不同子区域其  $a_j$  不同,而同一子区域中  $a_j$  为相同。据此,若令对角阵

$$A = \text{diag}(a_1, a_2, \dots, a_p) \quad (16)$$

则有面积订正后的资料阵

$$Y = XA \quad (17)$$

对  $\bar{Y}$  阵施以 EOFs 展开, 可得

$$Y = \bar{T}\bar{\psi} \quad (18)$$

$$\bar{T} = Y\bar{\psi} \quad (19)$$

根据矩阵理论的同时对角化定理<sup>[9]</sup>, 可以证明, 它与由原资料阵  $X$  的 EOFs 展开结果(特征向量  $\psi$ , 时间函数  $T$ ) 具有一定的差异, 即  $Y$  与  $X$  的特征向量场实际上是相差某一比例系数向量, 它在很大程度上反映出因站点分布不均匀性所产生的效应, 已由相应的不同比例系数(或放大系数)得到弥补, 因而使特征向量场有一定的差异。此外, 由文献<sup>[10]</sup>借助矩阵扰动理论亦可证明上述事实。由此可见, 非均匀站网情况下, EOFs 展开结果必然失真。

### 3 实例计算与结果比较

以中国境内(160站)近40a(1951—1989年)年平均气温场为例。分别采用上述修正和不修正 EOFs 展开。修正后的 EOFs 第1主分量曲线(即时间函数)其长期趋势总振幅明显加大(见图1b), 而未修正的 EOFs 第1主分量虽有增暖趋势, 但幅度比修正后的 EOFs 小(见图1a), 尤其是1970年代中期以后增暖趋势比修正前更明显。这表明经面积加权后, 使得西北部面积权重增大, 而东南部面积权重减少(见附表), 实际上等价于对站网分布均匀性作了重新调整, 因而使西北部的增暖信息较充分地反映出来, 其结果与许多研究结论是吻合的<sup>[11-12]</sup>。例如, 研究表明, 中国近40a平均温度变化的倾向率分布不均匀, 其主要特点是: 东北、华北及广大西部地区有明显增暖趋势, 而35°N以南, 青藏高原东部的部分地区是逐渐变冷的。若采用 EOFs, 由于西北地区面积占中国大部分, 但站点密度小, 如不考虑面积因子则必然使广大的西北部增暖信息的比重下降, 而使东部(站点密度大)的比重加大, 但东部一部分地区有降温趋势, 因而使全国增暖总趋势不能如实反映。采用修正 EOFs 展开恰好可弥补这一缺陷, 使 EOFs 结果更为合理。图2b所绘修正后第1特征向量场的西北部地区荷载明显地大于修正前的 EOFs(图2a)正好与上述分析十分一致, 确为佐证。

除第1主分量有明显差异外, 经面积加权后, EOFs 第2, 3主分量的调整就更为显著(图略)。例如, 修正后 EOFs 第2主分量趋势由下降变为上升, 尤以1980年代最为显著。

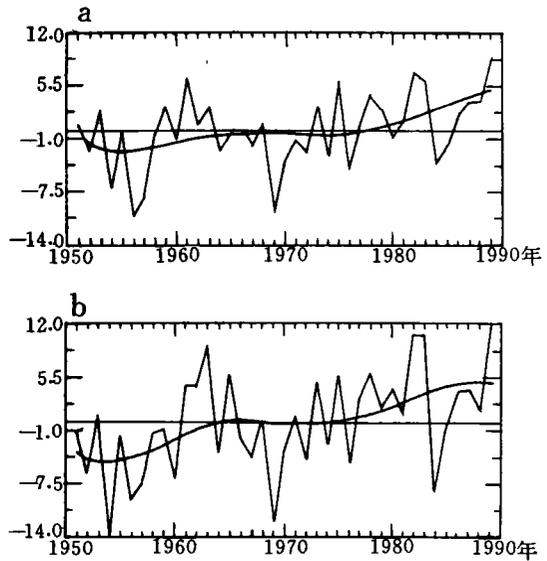


图1 面积修正前后 EOFs 第1主分量曲线  
(a 修正前, b 修正后)

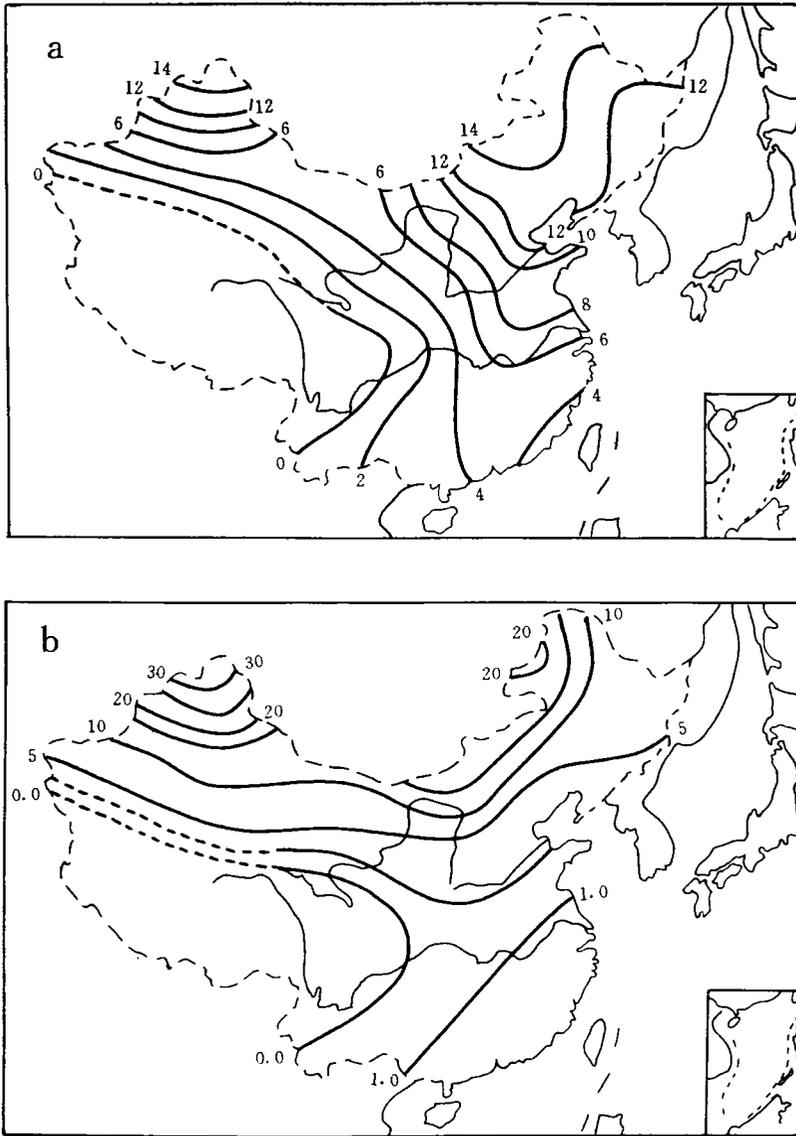


图2 EOFs 第1特征向量场(荷载场)  
(a 修正前, b 修正后)

这可能是由于减少了南方局部地区变冷趋势的负效应,而使西北部增暖信息的正效应重新显示的缘故。第3主分量也有类似的差异,且在1980年代最突出。这充分表明,近十多年来北半球增暖明显的趋势在中国西北部表现尤为显著,因而 EOFs 展开经面积加权修正后所揭示的气象场特征更为客观合理。

附表 中国各子区域面积(单位:10<sup>4</sup>km<sup>2</sup>)权重及站点数

子区域	面积	站数	面积权重	子区域	面积	站数	面积权重	子区域	面积	站数	面积权重
黑龙江	46.0	12	0.66	浙江	10.0	4	0.43	云南	38.0	8	0.82
吉林	18.0	3	1.03	江西	16.0	6	0.46	陕西	19.0	5	0.65
辽宁	15.0	5	0.52	湖北	18.0	5	0.62	甘肃	39.0	9	0.75
内蒙古	110.0	10	1.89	湖南	21.0	7	0.52	山西	15.0	3	0.86
河北*	21.78	6	0.62	福建	12.0	4	0.52	宁夏	6.6	2	0.57
河南	16.0	4	0.69	广东*	21.5	8	0.46	青海	72.0	3	4.13
山东	15.0	7	0.37	广西	23.0	6	0.66	新疆	160.0	12	2.29
江苏*	10.58	6	0.30	贵州	17.0	5	0.58	西藏	120.0	2	
安徽	13.0	5	0.45	四川	56.0	13	0.74	台湾	3.6	0	

\* 其中江苏子区域含上海市,河北子区域含北京、天津市,广东与海南合为广东子区域。

#### 4 小结

(1) 非均匀站网 EOFs 展开并不能客观地提取场的不同尺度时空信息特征;广义上说,通常 EOFs 仅是 Fredholm 积分方程求解的一个特例。

(2) 采用面积加权修正 EOFs 方案可大大改善站网的非均匀性,从而使 EOFs 展开结果更接近客观实际。例如对中国气温场增暖趋势的描述,采用修正 EOFs 能更加合理地揭示中国西北部地区对增暖总趋势的贡献。

#### 参考文献

- [1] Fortus M I. Statistically orthogonal functions for a random field specified in a finite region of plane. *Izv Acad Sci USSR Atmos Oceanic Phys.* 1975. 11:1107—1112.
- [2] Buell E C. Integral equation representation for factors analysis. *J Atmos Sci.* 1972. 28:1502—1505.
- [3] Buell E C. The number of significant proper functions of two-dimensional fields. *Appl Meteo.* 1978. 17:717—722.
- [4] Karl T R, Koscielny A J and Diaz H F. Potential error in the application principal component (eigenvector) analysis to geophysical data. *J Appl Meteor.* 1982. 21: 1183—1186.
- [5] Preisendorfer R W. PCA in meteorology and oceanography. Elsevier Sci. Publishers. Amsterdam B V. 1988.
- [6] Jones P D Wigley T M L and Kelly P M. Variations in surface air temperatures: Part 1. Northern Hemisphere. 1881—1980. *Mon Wea Rev.* 1982. 110(2): 59—70.
- [7] 祝昌汉. 我国气温变化诊断方法探讨. *应用气象学报.* 1992. 3(增刊): 112—118.
- [8] 《数学手册》编写组. 数学手册. 人民教育出版社. 1979. 750—761.
- [9] Rao C R. Linear statistical inference and its applications. John Wiley Sons, Inc. New York. London. Sydney. 1965. 36—37.
- [10] 丁裕国, 江志红. 气象场相关结构对 EOFs 展开稳定性的影响. *气象学报.* 1994. 4: 448—456.
- [11] 张先恭, 李小泉. 本世纪我国气温变化的某些特征. *气象学报.* 1982. 40(2): 198—208.
- [12] 林学椿, 于淑秋. 近四十年我国气候趋势. *气象.* 1990. 16(10): 16—23.

# THE LACK FIDELITY OF EOFs EXPANSION OVER HETEROGENEOUS NETWORK AND ITS REVISED SCHEME

Ding Yuguo Jiang Zhihong

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

## Abstract

Theoretically, it is demonstrated that current EOFs expansion results are not true to some extent over heterogeneous network. Under identical conditions, the obvious differences between homogeneous and heterogeneous network are researched for current EOFs expansion. A revised scheme using area weighting is developed, which can to a certain degree compensate for the lack fidelity of the EOFs expansion due to heterogeneous network. According to this technique, the fidelity of objective information from described meteorological fields is enhanced by using current EOFs method.

In a practical case, the regional differences of recent warming effect are more objectively revealed by using the revised EOFs technique to surface temperature field records (network from 160 stations) in response to the trends and features of China region.

**Key words:** Empirical Orthogonal Functions (EOFs), Heterogeneous network, Area weighting, Temperature change of China.