

# 經驗正交函數最原始的物理意義

林 和

國立台灣大學大氣科學系

(中華民國七十五年三月四日收件)

## 摘 要

任意一組由正基集合 (normal mode set) 構成的線性系統，經過初始值之隨機化 (Randomization)，如此收集的觀測資料，分析其經驗正交函數 EOF，當採樣數量足夠，EOF 與正基重合。

如果在觀察點上重疊雜訊或誤差，則 EOF 與正基亦產生偏差，雜訊若具統計均勻性 (homogeneity)，偏差不致太大，若雜訊為非均勻性，EOF 將失去代表意義。

本文試圖闡明 EOF 最基礎的物理解釋。

## 一、前 言

物理界觀察到的資料往往卷帙繁浩，如果將一組時空分佈的觀測值視為隨機變數 (random variable)，可經由線性轉換到另一組特殊的隨機變數，這組變數，按照次序排列，分別解釋原系統中差方 (variance) 的最大，次大，再次大……等變化。我們稱具備這種特徵的變數為經驗正交函數 (Empirical Orthogonal Function, 以下用 EOF)。因此，EOF 代表最濃縮的訊息。

自從 Obukhov (1947)，Lorenz (1956) 將 EOF 的概念引進大氣科學後，EOF 已被廣泛應用到分析大氣，海洋的結構，其後較重要的發展，包括將 EOF 推廣到時間序列 (Wallace and Dickinson, 1972)，當作預測因子 (Barnett and Hasselman, 1979)，以及有點採樣技術及統計穩定性等研究 (Buell, 1979, Proisendorfer et al. 1981)。

然而，從決定論 (deterministic) 的觀點來

看，EOF 真實的物理意義，迄今還停留在呼之欲出的階段。一般而言，我們觀測到的物理元子 (Physical entity)；即可沿時間座標予以追蹤，標識，或與其它元子區分者，泰半屬於週期性運動，間或有例外，如孤子 (Soliton) 等，往往處於線性—非線性間精細的平衡狀態，不易取得大量能量，(木星上的大紅點 big red spot 恰為反例)。

以大氣為例，假設地表光滑，沿軸對稱，按標準大氣分層 (Stratification)……等條件，忽略輻射，水汽效應，再將控制大氣的動力方程式線性化，結果將得到一 Hermitian 算子 (見 Daley 的 Review, 1981)，加上初始及邊界條件，數學上構成典型的特徵問題 (eigen problem)。

如此解出的特徵函數 (eigen function)，保證為正交集 (Orthogonal Set)，故稱為正基 (normal mode)，正基在自然界為週期性波動，雖然本身是線性解，但一般人相信，可以對應到大氣中確實觀測到的物理元子 (Daley, R, Williamson, 1985)。

大尺度現象中，週期性運動之所以突出，歸功於大氣裡具備如  $\beta$ -效應，垂直分層……等，相當有效的恢復機制 (recovering mechanism)，一旦能量集中在波型態的分量——即時，空尺度符合嚴謹的頻散 (dispersion) 律，以相對於鬆弛時空尺度關連之紊流型態——其後波與波之間，惟有靠共振 (resonance) 才能交換能量。共振的選擇律非常苛刻；所以正基能保持一定程度的週期性，( 頻率 frequency  $\omega \gg$  增幅 amplification  $\frac{1}{A} \frac{dA}{dt}$  )。

雖然我們熟知，如重力波與旋轉波涇渭分明，數值天氣預報更奉為圭臬。但實際收集到的氣象資料裡，仍舊含蘊著強烈的隨機性 (Stochasticity)，毫無疑問，線性系統省略的機制，如海面溫度，對流性降水，非線性紊流等，紛紛以複雜的隨機外力形式出現。我們處理這一類氣候資料，最簡化的方法是假設天氣在同一動力過程控制之下，但初始值為隨機狀態，故以套 (ensemble) 的集合處理，天氣學的合成圖 (Composite map) 即屬此類。

同屬線性系統，隨機的來源有二，一由初始值來，一為受到隨機外力 (係數 Parameter 或邊界條件亦可能為隨機變數) 影響。不論何種情形，我們以直覺判斷，EOF 和正基必定有牢不可分的關係；(當然，這是假設線性系統的正基存在的情況下。) 近年來，Preisendorfer (1979, et al 1981)，North (1984) 等人已證實由隨機外力驅動的振盪運動，經由校正後的 EOF 與正基合而為一。本文特着重在由初始值激發的線性隨機系統。我們由數值方法確定，EOF 與正基之間最基礎的物理解釋。

第二節將討論因正基性質及採樣數量導致 EOF 呈現的假相與真相。我們將以一項稱之為聖誕樹燈串的實驗做為例證。第三節說明正基外的雜訊 (noise) 將會影響到 EOF 代表正基的程度，雜訊分為均勻雜訊 (homogeneous noise) 與非均勻 (連站性) 雜訊 (inhomogeneous noise)，第四節討論 EOF 可能在資料同化上的應用。

## 二、EOF最原始的物理含義

這一節我們將從數值實驗顯示，以正算子 (normal operator, 見 Halmos 1958) 表示的線性系統當具有正基解，且正基的初始值取自某固定的機率密度分佈 (Probability density distribution) 的情況下，如採樣次數夠多的話，其統計套 (Statistical ensemble) 中的 EOF 即可一一對應到正基。

上述說明，機率密度分佈並不限於高斯 (Gaussian) 分佈，至少可推廣到白噪分佈 (white noise distribution)，或甚至更廣泛的情況。

這節我們舉一個例子說明以上觀點，放置一架照相機於聖誕樹前數公尺處，樹上懸掛了多條燈串 (i. e. normal modes)，各由不同的隨機調光器 (dimmer) 調節，若此照相機的析像力不足以分辨細節，經過多次快照 (i. e. EOF) 後，我們是否能從照片上確定每條燈串的位置？

首先，將照片上的聖誕樹，劃分成若干觀察點 (i. e. weather station) 每點的亮度  $I^j(x)$ ，係由經過該點週圍的燈串群線性重疊構成；

$$I^j(x) = \sum_{i=1}^M A_i^j \phi_i(x)$$

$x$  代表觀察點在樹上的地理位置， $j$  為第  $j$  次拍攝的照片 (i. e. realizations)， $A_i^j$  由隨機調光器在  $j$  照片上  $i$  燈串決定的光度， $\phi_i(x)$  為  $i$  燈串在聖誕樹上的分佈，為了增強這題目的適用性，我們假設  $\phi_i(x)$  可以為小數 (0.3 個燈泡靠近  $x$  觀察點) 或負數 (相對於背景光度的暗色燈泡)。

若燈串各自獨立 (Linearly independent)，且互相正交 (orthogonal) 放置， $\phi_i(x), \phi_j(x)$  符合下列條件

$$\sum_x \phi_i(x) \phi_j(x) dx = \delta_{ij}$$

則  $\{\phi_i, \phi_j, \dots\}$  為一組 orthonormal 的正基集合。從照片上觀測到的資料，計算 EOF 如下：

(1) 除去平均值

$$I^j(x) = I^j(x) - \langle I^j(x) \rangle$$

$\langle \rangle$  為套平均，即對所有照片的平均

$$\langle \psi \rangle = \frac{1}{N} \sum_{j=1}^N \psi_j$$

(2)建協方差矩陣

$$C = \begin{pmatrix} \langle I'(x_1) I'(x_1) \rangle & \langle I'(x_1) I'(x_2) \rangle & \dots \\ \langle I'(x_2) I'(x_1) \rangle & \langle I'(x_2) I'(x_2) \rangle & \dots \\ \vdots & \vdots & \ddots \\ & & \langle I'(x_M) I'(x_M) \rangle \end{pmatrix}$$

(3)分解 (decompose) C, 求得特徵值  $\lambda_i$  與特徵函數  $e_i$  (即 EOF) 與  $\phi_i$  比較。

數值實驗中考慮的三項因素為(1)初始值所對應的機率密度分佈(2)採樣的數量(3) $A_i^j$  的能譜, 即方差 (variance  $\sigma_i^2$ ) 之間隔 (discrimination)。我們使用 5 條燈串,  $\phi_i(x)$  由電腦分析一任意的 Hermitian 矩陣而來, 初始假設為高斯分佈。

$$A_i^j = \left[ -6 + \sum_{k=1}^{12} a^k \right] \sigma_i$$

$a^k$  從電腦的隨機數產生程式產生,  $a^k$  的或然率為  $P(a)=1$ , 在  $0 \leq x \leq 1$  之間;  $P(a)=0$ , 在  $[0, 1]$  以外空間。根據中限理論 (Central-limit theory), 上式將產生高斯變數。  $A_i^j$  的能譜設為  $\sigma_i = (i)^z$ ,  $Z=0, 1, 2, 3, 4$ 。採樣的次數分為  $(2)^n$ ,  $H=2, 3, \dots, 9$ 。

實驗的結果見圖一, 圖二, 圖三, 代表  $i=1, 3, 5$  的正基, 橫軸為採樣的次數, 縱軸為 EOF 減去正基的誤差

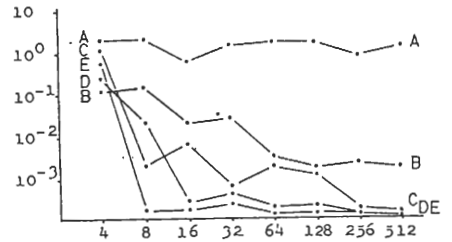
$$\text{Error} = \sum_{x=1}^{\text{全樹}} (e_i(x) - \phi_i(x))^2$$

其中 EOF 的符號視正基決定相同者, 因  $\|\phi_i\|=1$ , 故誤差同時為絕對與相對誤差值, 一般說來,  $\text{Error} \leq 10^{-2}$  則 EOF 與正基幾乎完全一致。

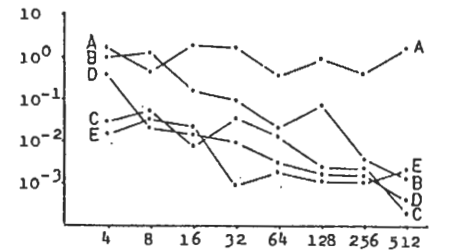
圖中線條 A, B, C, D, E 分別為  $Z=0, 1, 2, 3, 4$  的能譜,  $Z=0$  時為能量平均分配 (energy equi-partition) 的狀況; 此時 EOF 完全不具有解析的能力, 故無物理意義, (在此須提醒讀者, 非線性項最後使命, 即為迫使系統趨往能量平均分配的終極狀態, 尤其在缺乏摩擦力的情形下, (Onsager, 1949), 當能階平滑 ( $Z=1$ , line B)

時, EOF 須多次採樣, 方能和正基重合, 大致說來, 能階越陡峻, EOF 與正基合一所需的採樣也越少。

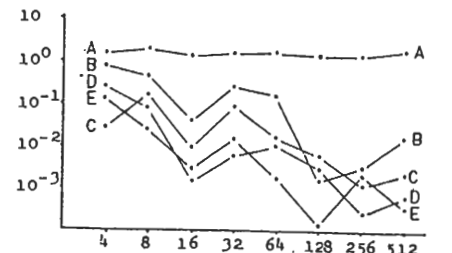
最主要的正基 (圖三,  $i=5$ ) 往往跳動的幅度也最大, 這一點與我們平常的觀念相違逆, 如果採樣次數相同時, 圖一顯示, 能量最小的 EOF ( $i=1$ ) 最準確, 很快達到正基的答案。而能量最大的 EOF ( $i=5$ ) 反而較參差。無論在任何情形之下, 128 張照片即足以精確的標明各條燈串的位置。我們同時嘗試不同的機率密度分佈, 或將各個正基分別採自不同的密度分佈, 所得結果與前述並無太大出入。



圖一 縱軸為 EOF, 正基間之誤差  $\sum_{x=1}^5 (e_i(x) - \phi_i(x))^2$ , 橫軸為採樣數目, A, B, C, D, E 線分別代表能譜  $\sigma_i = (i)^z$ ,  $Z=0, 1, 2, 3, 4$  的情況。此圖對應正基  $i=1$ , 即能量最小的分量。



圖二 同圖一,  $i=3$



圖三 同圖一,  $i=5$ , 能量最大的分量

### 三、觀測雜訊的影響

氣象觀測的不確定性，不止來自觀測過程的誤差，同時包括次網格內運動及物理的擾動。EOF 對雜訊 (noise) 的反應相當敏感，假設在  $x$  站的值加上雜訊  $\epsilon(x) * f$  :

$$I'(x) + \epsilon(x) * f$$

$f$  為雜訊強度，設為 0.1 或 1.0 訊號  $I'$  的強度。 $\epsilon(x)$  為另一系列隨機變數，雜訊依其特性分為均勻及非均勻兩種，均勻誤差的協方差矩陣為對角矩陣

$$H = \begin{vmatrix} 1 & 0 & 0 & \dots & \dots \\ 0 & 1 & 0 & \dots & \dots \\ 0 & 0 & 1 & \dots & \dots \\ 0 & \dots & \dots & \dots & \dots \\ 0 & \dots & \dots & \dots & 1 \end{vmatrix}$$

非均勻矩陣假設為

$$I = \begin{vmatrix} 1 & \frac{1}{2} & \frac{1}{2} & \dots & \dots \\ \frac{1}{2} & 1 & \frac{1}{2} & \dots & \dots \\ \frac{1}{2} & \dots & 1 & \dots & \dots \\ \vdots & \dots & \dots & \dots & 1 \end{vmatrix}$$

非均勻雜訊須經過轉換

$$\epsilon(x) = \sum_x R(x, x') a(x')$$

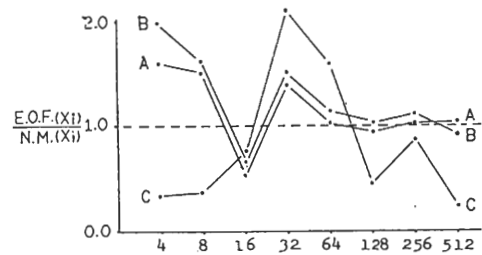
其中  $a(x')$  來自電腦隨機數產生方程式 (均勻性)， $R(x, x')$  為矩陣  $I$  的平方根 ( $I = R R^*$  轉置矩陣)，可以根據譜定理 (spectral theory, 見 Halmos, 1958)，由  $I$  的特徵函數，特徵值的平方根，相乘組成。 ( $R(x, x') = \sum_i (\lambda_i)^{1/2} e_i(x) e_i(x')$ ， $\lambda_i, e_i$  為  $I$  的特徵值，特徵函數)。

圖四呈示不同程度的均勻雜訊對 EOF 的影響，縱軸代表任意選取之某項 EOF 與正基比值，當雜訊強度僅為信號十分之一時 (Line B)，EOF 向正基重合的速率與無雜訊時 (Line A) 相當接近

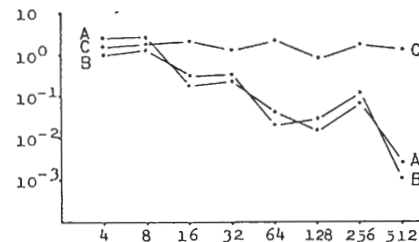
。事實上 Line B 減掉 0.1 的雜訊能量，即可與 Line A 完全一致，但當雜訊強度相當於信號 ( $f = 1.0$ , Line C) EOF 與正基乖離，不論採樣多少，兩者毫無關聯。

這種情形，實在帶給 EOF 的身價很大的打擊：Rhines (1975) 從數值實驗推論當週期性運動與紊流各佔一半能量，自然界流體達到平衡狀態，即 Rhine effect。若週期性運動以正基代表，紊流視作雜訊，則我們得到的 EOF 將失去物理含義。雖然如此，我們可輕易的從取樣的不穩定性，判斷 EOF 是否有意義。

圖五標示無雜訊 (Line A,  $f = 0.0$ )，均勻性雜訊 (Line B,  $f = 0.1$ )，與不均勻雜訊 (Line C,  $f = 0.1$ ) 比較，倘若雜訊在空間中缺乏獨立性，EOF 將被嚴重污染，甚至毫無意義。不均勻雜訊的強度雖然只及正基的十分之一，卻能完全摧毀 EOF 與正基的對稱。



圖四 縱軸為加上均勻雜訊後，在某一任意選取的測站上，EOF 與正基的比值，橫軸仍為採樣數目，A 線代表雜訊強度為零，B 線雜訊強度為信號強度之 0.1，C 線雜訊強度與信號強度相等。



圖五 同圖一，A 線無雜訊，B 線均勻雜訊強度為 0.1 信號強度，C 線為非均勻雜訊，強度亦為 0.1 信號強度。

#### 四、結 論

由動力方程式產生的 EOF，確實可還原回正基的形狀。

我們可以推廣到任何正算子：

$$L\phi = 0$$

的隨機套集，EOF 皆可以顯示正基的能譜及結構，舉例說：

羅士培波

$$\frac{\partial}{\partial t} \Delta^2 \phi + \beta \frac{\partial}{\partial x} \phi = 0$$

在理想邊界條件下，即符合這限制。我們目前正進行紊流羅士培波 (turbulent Rossby waves) 的分析。

至於有關 (North, 1984) 等人的研究，隨機外力或隨機初始條件，除了在“各態歷程”(ergodic) 假設下，並非相同的題目，實際觀測的氣候資料，應該同時包括這兩種情形。

如果 EOF 即正基，為何大氣中尚未有確切證據，證明這兩者關係呢？大氣雖然具二度空間性的流體，能譜的間隔相當大，但具有相同能階的正基亦隨波數的平方增加，如果將不同垂直結構，不同時間尺度的 EOF 加以區分，則形成一龐大的協方差矩陣，遠遠超過我們的計算能力。

但我們如果採取本文的方法或觀點，進一步了解 EOF 的限制與正基的關係，逐次引進誤差，非線性擾動，也許 EOF 可用來印證正基或一叢正基的特性，為將來大氣模式的資料同化，提供一正確的線索。

#### 誌 謝

本文圖片由黃拔源先生描繪，大部份工作在蔡清彥教授借用的 IBM PC 完成特此申謝。

#### 參考資料

- Barnett T. P., and K. Hasselmann 1979 : Techniques of Linear prediction with application to oceanic and atmospheric fields in the tropic Pacific. *Rev. Geophys. and Space Phys.* 1978, 949 — 968.
- Buell, C. E., 1979 : On the physical interpretation of empirical orthogonal function, Preprints 6th Conf. on the Probability and Statistics in Atmospheric Sciences. Amer. Meteor. Soc., 112-117.
- Daley, R., 1981 : Normal Mode Initialization Dynamic Meteorology, Springer & Verlag.
- Daley, R., and Williamson D. : The existence of Free Rossby waves during January 1979. Oct. 1985, 2121-2141.
- Halmos P. R., 1958 : Finite-Dimensional Vector Spaces. Van Nostrand, 200 pp.
- Lorenz, E. N., 1956 : Empirical orthogonal functions and statistical weather prediction. Rep. 1, Statist. Forecasting project. MIT.
- North, G., 1984 : Empirical Orthogonal Functions and Normal Modes. Vol. 41, No. 5, 1, March, *JAS*, 879-887.
- Obukhov, A. M., 1947 : Statistically homogeneous fields on a sphere. Usp. Mat. Navk. 2, 196-198.
- Onsager, 1949 : Statistical Hydrodynamics. Suppl. Vol. VI, Nuovo Climento.

- Preisendorfer, R. W., 1979: Principle components and the motions of simple dynamical systems. SIO Ref. Ser. 70-11. Scripps Institution of Oceanography, 46 pp.
- Preisendorfer, R. W., F. W. Zwiers and T.P. Barnett, 1981: Foundations of principal component selection rules. SIO Ref. Ser. 81-4. Scripps Institution of Oceanography, 192 pp.
- Rhines, P., 1975: Waves and turbulence on a Beta-plane. Vol. 69 Part 3, JFM, 417-443.
- Wallace, J. M., and R. Dickinson, 1972: Empirical orthogonal representation of time series in the frequency domain. Part I: Theoretical considerations. J. Appl. Meteor., 11, 887-892.

## A Straightforward Physical Interpretation of Empirical Orthogonal Function

Lin Ho

Department of Atmospheric Science, National Taiwan University

### ABSTRACT

Any linear System governed by a normal operator the Empirical Orthogonal Functions (EOF) will be identical to the normal modes if the stochasticity is introduced from randomly chosen initial amplitudes, provided that the samples are abundant.

At each station the appearance of an additional level of noise can interfere the convergent rate of the EOF toward the normal modes, especially when the noise is inhomogeneous (noise covariance matrix has nonzero off-diagonal elements), the EOF will soon lose any resemblance of the normal modes.

The physical interpretation of EOF in this study is considered to be unprecedentedly straightforward.