Tạp chí Khoa học và Công nghệ biển T8 (2008). Số 3. Tr 30 - 42

XÁC ĐỊNH CÁC THÀNH PHẦN CHÍNH CỦA TRƯỜNG GIÓ BỀ MẶT BIỂN ĐÔNG BẰNG PHƯƠNG PHÁP PHÂN TÍCH HÀM TRỰC GIAO THỰC NGHIỆM

NGUYỄN VĂN TUÂN, VÕ VĂN LÀNH, TỐNG PHƯỚC HOÀNG SƠN

Tóm tắt: *Bài báo trình bày các kết quả sử dụng phương pháp phân tích hàm trực giao thực nghiệm EOF để nghiên cứu các thành phần cấu trúc chính của trường gió bề mặt (ở độ cao 10m trên mặt biển) Biển Đông trên cơ sở phân tích một tập hợp lớn số liệu gió trung bình tháng theo ô vuông 0.25°x0.25°, trong khoảng thời gian 8 năm từ năm 1999 đến năm 2007. Kết quả phân tích cho thấy rằng, trường gió bề mặt Biển Đông có thể xấp xỉ bằng 4 thành phần chính sau đây:*

- Thành phần thứ nhất EOF1 đặc trưng cho trường gió bề mặt trong mùa gió Đông Bắc với đặc điểm là sự tồn tại tâm phân kỳ gió ở vùng biển Tây Bắc đảo Luzon dẫn đến hình thành hiện tượng nước trồi mạnh ở vùng này và tâm hội tụ gió bề mặt ở vùng biển Nam Trung bộ Việt Nam dẫn đến hình thành hiện tượng nước chìm mạnh ở đây.

- Thành phần thứ hai EOF2 đặc trưng cho trường gió bề mặt trong mùa gió Tây Nam với đặc điểm là sự tồn tại tâm phân kỳ gió mạnh dẫn đến sự hình thành hiện tượng nước trồi mạnh ở vùng biển Nam Trung bộ và Đông Nam bộ.

- Các thành phần thứ ba EOF3 và thứ tư EOF4 đặc trưng cho trường gió bề mặt trong các mùa gió chuyển tiếp với đặc điểm là sự tồn tại gió Tây ở trung phần Biển Đông cùng tâm phân kỳ gió ở vùng biển phía Tây đảo Luzon, và tâm hội tụ gió trên thềm lục địa Nam Biển Đông dẫn đến hình thành những tâm phân kỳ và hội tụ nước tương ứng.

- Các thành phần khác là rất nhỏ và không ổn định nên có thể bỏ qua.

I. MỞ ĐẦU

Khi sử lý số liệu quan trắc thường gặp trường hợp là số các đặc trưng (biến số) được xét khá lớn. Trong hàng loạt bài toán, không phải lúc nào cũng lựa chọn được một cách trực tiếp các biến để tiến hành xử lý kết quả. Điều này càng quan trọng khi không có những mô hình toán học có cơ sở tốt hoặc khi cần làm rõ ràng các giả thiết đã đặt ra, và khi đó mục đích của thực nghiệm sẽ bao gồm cả việc tìm hiểu ý nghĩa của sự phụ thuộc giữa các biến và hướng sử dụng chúng sau này. Nếu số những biến được nghiên cứu là quá lớn thì thường nảy sinh những khó khăn liên quan tới việc lựa chọn hàm số và đối số của chúng.

Phép biến đổi sau đây loại bỏ những sự phụ thuộc không tuyến tính có thể có giữa các biến số. Phương pháp thành phần chính thường được dùng để nhận những thông tin phụ về nguồn gốc và bản chất của sự bất định của đối tượng được nghiên cứu.. Đồng thời, thành phần chính có thể được sử dụng trong việc mô tả sự phân lớp các kết quả quan trắc.

Ý tưởng của phương pháp các thành phần chính là biểu diễn một quá trình phức tạp thành tổng của những thành phần độc lập đơn giản hơn (gọi là những thành phần triển khai). Tuy nhiên, khác với khai triển theo các đa thức Chebusev hay theo chuỗi Fourier, trong phương pháp phân tích thành phần chính cấu trúc của các tham số khai triển không được định ra trước. Nó được xác định một cách đơn trị và khách quan chỉ bởi nội dung của quá trình được nghiên cứu và mức tương quan của các chuỗi quan trắc với nhau. Vì vậy người ta gọi đây là phương pháp phân tích thành phần chính khai triển theo những hàm trực giao thực nghiệm, hay còn gọi là phương pháp phân tích hàm trực giao thực nghiệm (Empirical Orthogornal Function - EOF).

Trong khuôn khổ bài báo này chúng tôi sử dụng phương pháp phân tích hàm trực giao thực nghiệm EOF để phân tích các thành phần cấu trúc chính của trường gió bề mặt Biển Đông trong thời gian 8 năm từ 7/1999 đến 6/2007.

II. PHƯƠNG PHÁP PHÂN TÍCH HÀM TRỰC GIAO THỰC NGHIỆM

Phương pháp phân tích EOF của Kelly 1985, 1989 về bản chất là phương pháp phân tích hàm trực giao thực nghiệm EOF, tương tự với phương pháp phân tích yếu tố (Factor Analysis - FA) hay ở bậc đơn giản là phương pháp phân tích thành phần chính (Principle Component Analysic – PCA). Đó là một xấp xỉ để ước lượng các kiểu phân bố không gian của các biến theo sự biến đổi về thời gian của một tập mẫu khá lớn Y_i (t_j). Ở đây Y_i là giá trị của biến ở vị trí i, và t_j là giá trị của biến thời gian ở thời điểm j. Một cách tổng quát có thể biểu diễn sự biến thiên của trường theo không gian và thời gian thành một ma trận D_{ij} có kích thước m*n .

$D = \|Dij\|$

- i = 1, 2, 3,....., m là số lượng biến

- j = 1, 2, 3,..., n là số lượng các thời gian

Biểu diễn số của hàm ngẫu nhiên D_{ij} có thể xác định thông qua các phép phân tích ma trận được xác định ở dạng:

$D_{ij} = \sum F_{hi} * X_{hj}$

$h = 1,n; i = 1,m; j = 1, n$

X_{hj} : là hàm véc tơ riêng (eigent vector) thể hiện thông qua các giá trị riêng (eigent) của ma trận tương quan (hoặc ma trận tương tự, ma trận hiệp phương sai,...) của yếu tố gió cần phân tích

F_{hi} : là hệ số phân tích các thành phần cấu trúc chính (đại lượng không thứ nguyên).

Để xác định X_{hi} cần giải hệ phương trình tuyến tính viết dưới dạng ma trận

$(R - \lambda_h * I) X_h = 0$

R - là ma trận tương quan (ma trận tương tự, ma trận hiệp phương sai,...) nhận được từ ma trận D_{ij}.

I - là ma trận đơn vị cùng bậc với R

X_h - vecto riêng của ma trận R

λ_h – là nghiệm của phương trình đặc trưng $|R - \lambda_h * I| = 0$

Thành phần chính là tích của ma trận biến vị X' (ma trận biến vị của X) với ma trận dữ liệu D_{ij}

$F = D * X'$

Các giá trị riêng λ_h là phương sai của véc tơ F_h. Nếu xếp chúng theo thứ tự giảm dần thì có thể đánh giá sự đóng góp tương đối (%) của các thành phần chính lên phương sai của tập mẫu chung.

Độ chính xác của phép phân tích với số thành phần tùy ý được xác định bởi chỉ số

$$E = \sum_{j}^{h} \lambda j / \sum_{j}^{n} \lambda j$$

Khi dùng hết tất cả các thành phần (h = n) thì phép phân tích là chính xác 100% . Trên thực tế chúng ta chỉ khống chế vài thành phần chính đầu tiên (1, 2, 3, 4,...) vì chúng bao gồm tới 60 – 70% phương sai (thông tin) của trường ban đầu.

Sự khống chế đó chính là phép lọc thống kê cần thiết để thu được các quá trình cỡ lớn và lọc đi các quá trình ngẫu nhiên kích thước nhỏ.

Một sự khác biệt đặc trưng của phép phân tích EOF so với phương pháp PCA là sử dụng đồng thời cả véc tơ riêng của biến gió (được gọi là tải trọng nhân tố - factor loading) và cả véc tơ riêng của biến thời gian (được gọi là trị số nhân tố - factor value).

III. NGUỒN SỐ LIỆU VÀ CÁC BƯỚC PHÂN TÍCH

Số liệu trường gió được lấy từ trang web ftp://ftp.ssmi.com. Đây là nguồn số liệu gió thu thập, xử lý từ các số liệu vệ tinh. Số liệu gió được đưa về độ cao chuẩn 10 m trên mặt biển với độ phân giải 0.25^ox0.25^o và được gọi là gió bề mặt Biển Đông. Trường gió trên Biển Đông được thu thập là số liệu gió trung bình tháng trong thời gian 8 năm từ 7/1999 đến 6/2007. Nguồn số liệu này tuy có độ chính xác không cao (có thể sai số từ 1 đến 2 m/s) nhưng đây là nguồn số liệu có tính chất đồng bộ trên quy mô lớn cho phép tìm hiểu một số đặc trưng quan trọng của trường gió khu vực.

Để có thể tính toán được các thành phần EOF ta phải thiết lập được ma trận số liệu. Ở đây số liệu trường gió được phân tích thành hai thành phần kinh tuyến (U) và vĩ tuyến (V), sau đó xắp xếp chúng thành từng cột tương ứng với từng tháng:

$$
2M\left\{\begin{array}{l}
U\left\{\begin{array}{cccc}
\overbrace{u_{11} \quad u_{12} \quad \cdots\cdots\cdots \quad u_{1N}}^{N} \\
u_{21} \quad u_{22} \quad \cdot \quad u_{2N} \\
\cdot \quad \cdot \quad \cdot \quad \cdot \\
\cdot \quad \cdot \quad \cdot \quad \cdot \\
u_{M1} \quad u_{M2} \quad \cdots\cdots\cdots \quad u_{MN}
\end{array}\right. \\
V\left\{\begin{array}{cccc}
v_{11} \quad v_{12} \quad \cdots\cdots\cdots \quad v_{1N} \\
v_{21} \quad v_{22} \quad \cdot \quad v_{2N} \\
\cdot \quad \cdot \quad \cdot \quad \cdot \\
\cdot \quad \cdot \quad \cdot \quad \cdot \\
v_{M1} \quad v_{M2} \quad \cdots\cdots\cdots \quad v_{MN}
\end{array}\right.
\end{array}\right.
$$

(Ma trận số liệu để tính toán)

Trong đó:

M: số điểm theo không gian

N: số liệu theo thời gian

Với trường số liệu trung bình tháng thu thập trong 8 năm tương ứng với 96 tháng liên tục, ta thiết lập ma trận số liệu với số cột là N = 96. Khu vực tính toán trên Biển Đông được giới hạn: $100^oE \div 125^oE$, $0^oN \div 25^oN$; với độ phân giải 0.25^ox0.25^o. Sau khi lọc những ô ở trên bờ và những ô không có số liệu chúng ta nhận được 5951 ô có số liệu đầy đủ trong 8 năm liên tục và cuối cùng chúng ta ghép lại thành ma trận số liệu hai chiều có kích thước là 2*5951 x 96.

Toàn bộ quá trình tính toán được thực hiện theo từng bước như trong [3, 4].

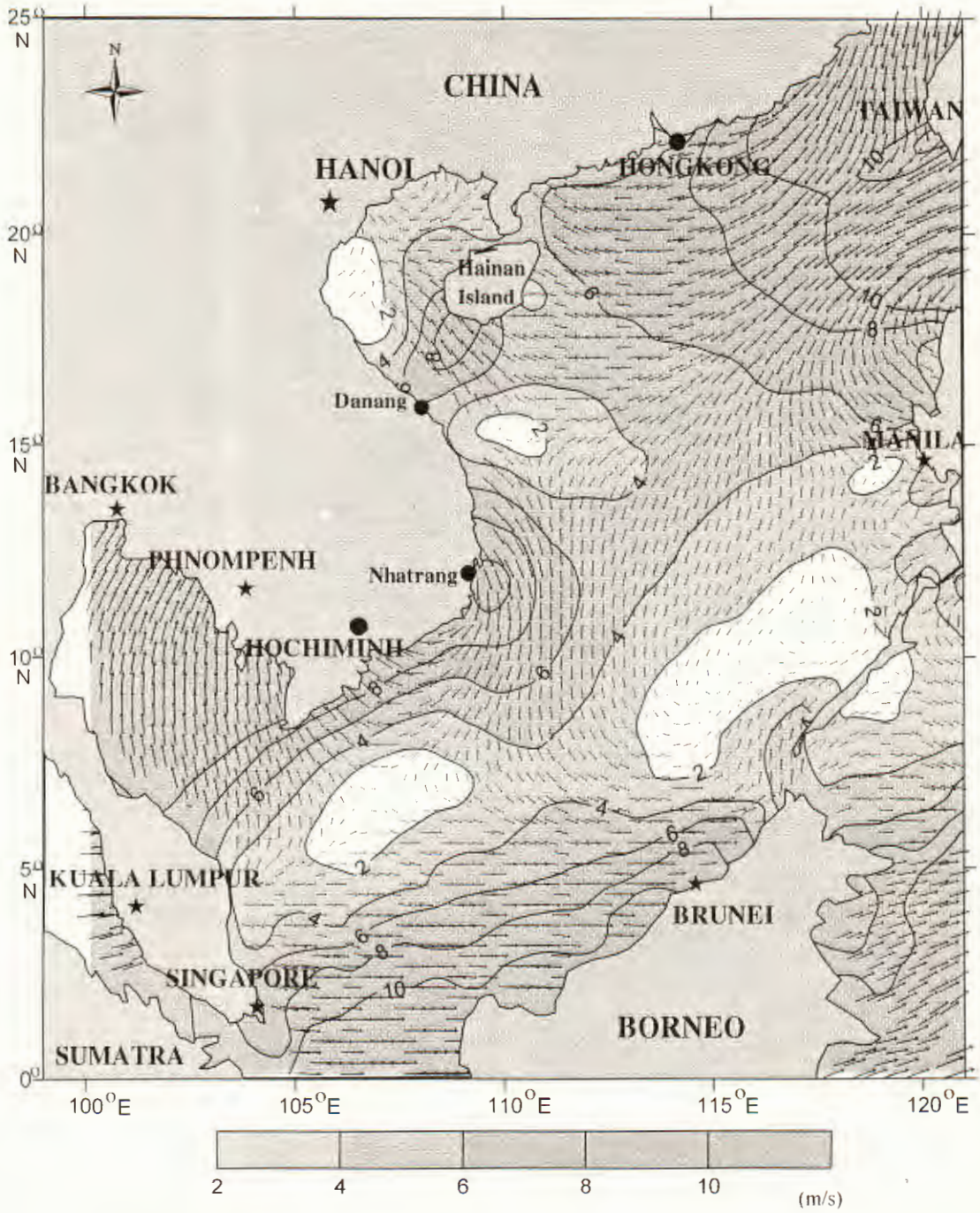
IV. KẾT QUẢ PHÂN TÍCH VÀ THẢO LUẬN

Các kết quả tính toán cho thấy: nếu dùng 4 thành phần (biến số) đầu tiên thì mức độ tập trung của số liệu là trên 77% (bảng 1), tương ứng với trường hợp dùng 4 thành phần này để phân tích lại trường số liệu ban đầu thì độ chính xác đạt được là 77%. Nếu dùng đến 9 thành phần thì độ chính xác đạt được 86%. Còn nếu dùng hết tất cả các thành phần để phân tích lại trường ban đầu thì độ chính xác là 100%. Như vậy, bốn thành phần đầu tiên có thể xấp xỉ khá tốt trường ban đầu.

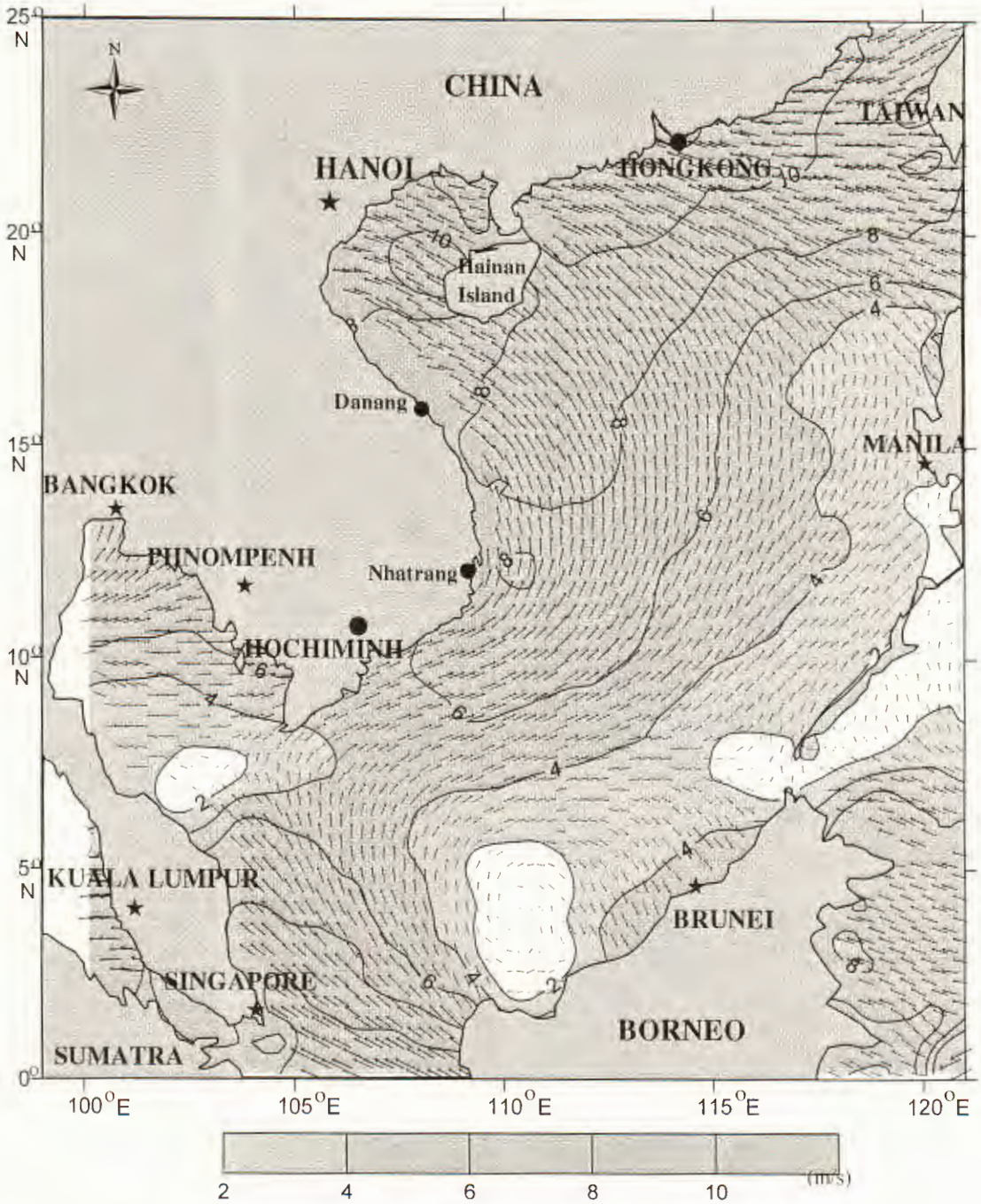
Thành phần đầu tiên có mức độ tập trung thông tin lớn nhất, nó chiếm 34% trong toàn bộ các thông tin ban đầu. Thành phần thứ hai chiếm 25%, thành phần thứ ba 13% và thành phần thứ tư 4%. Thành phần EOF4 tuy có độ tập trung thông tin thấp (4%) nhưng phân bố không gian và tải trọng của nó có những đặc điểm cần được lưu ý. Còn lại các thành phần khác chiếm tỉ lệ rất nhỏ (bảng 1).

Bảng 1: Mức độ tập trung thông tin của trường gió theo 9 thành phần đầu tiên

Thành phần	Độ tập trung (%)	Tổng tích lũy (%)
1	34	34
2	25	59
3	13	72
4	4	77
5	3	80
6	2	82
7	2	83
8	1	85
9	1	86



Hình 1: Phân bố độ lớn (m/s) và hướng (mũi tên) của thành phần EOF1 theo không gian

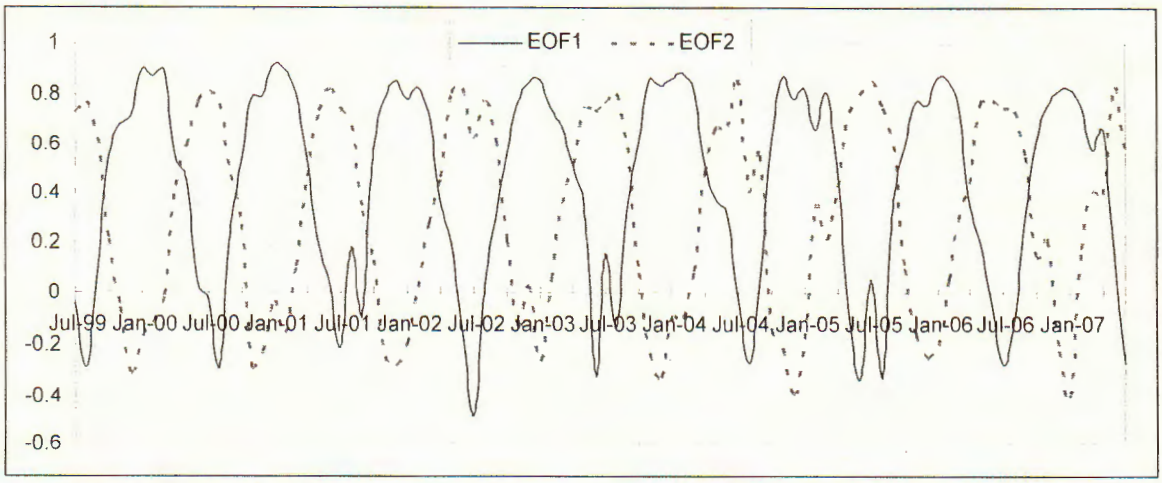


Hình 2: Phân bố độ lớn (m/s) và hướng (mũi tên) của thành phần EOF2 theo không gian

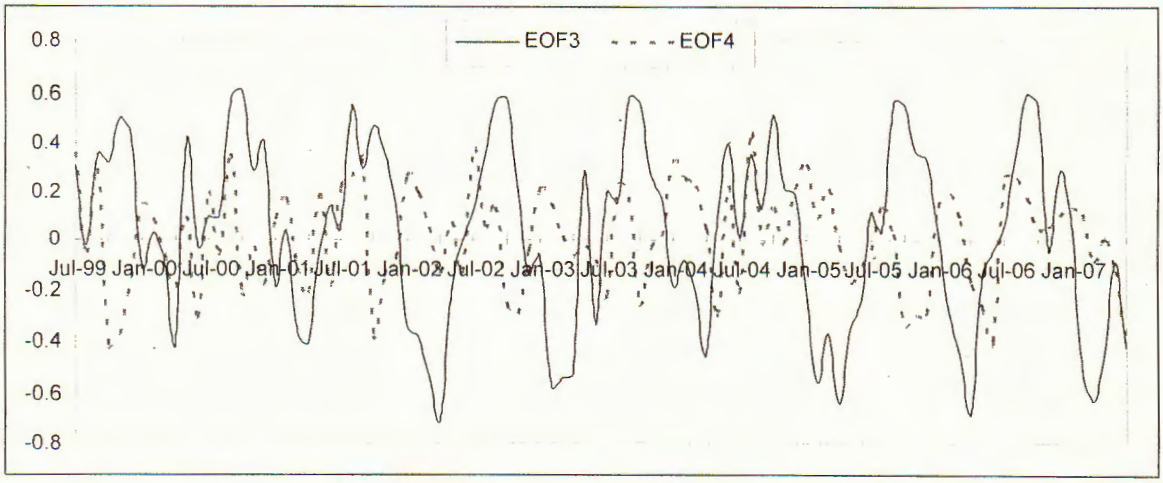
Hình 3: Phân bố độ lớn (m/s) và hướng (mũi tên) của thành phần EOF3 theo không gian

Hình 4: Phân bố độ lớn (m/s) và hướng (mũi tên) của thành phần EOF4 theo không gian

Từ hình 5 thấy rằng các nhân tố tải trọng của thành phần thứ nhất và thứ hai biến đổi có tính chu kỳ năm và có xu thế biến đổi ngược pha với nhau. Thành phần thứ nhất có tải trọng cực đại vào thời kỳ mùa đông, còn mùa hè đạt giá trị cực tiểu. Thành phần thứ hai có xu thế biến đổi ngược lại. Từ các kết quả này thấy rằng: thành phần EOF thứ nhất có thể đặc trưng cho thời kỳ gió mùa Đông – Bắc, thành phần thứ hai đặc trưng cho thời kỳ gió mùa Tây – Nam. Tải trọng của thành phần thứ ba và thứ tư (hình 6) cũng có tính biến đổi theo chu kỳ năm, với độ lệch pha nhất định so với hai thành phần trên. Cực đại tải trọng của các thành phần này đạt được vào thời kỳ chuyển tiếp giữa hai mùa: từ mùa hè sang mùa đông và ngược lại. Vì vậy các thành phần này có thể xem là đặc trưng cho các mùa chuyển tiếp. Các thành phần khác chiếm tỉ lệ nhỏ, khó xác định được quy luật biến đổi và có thể xem là do các nhiễu động nhỏ hay phi tuyến gây ra. Trong nhiều bài toán thực tế các thành phần này có thể bỏ qua.



Hình 5: Biến đổi tải trọng của thành phần chính thứ nhất và thứ hai theo thời gian



Hình 6: Biến đổi tải trọng của thành phần chính thứ ba và thứ tư theo thời gian

Như vậy, bằng phương pháp phân tích hàm trực giao thực nghiệm, trường gió bề mặt Biển Đông có thể được xấp xỉ bằng 4 thành phần đầu tiên EOF1, EOF2, EOF3 và EOF4. Thành phần đầu tiên EOF1 đặc trưng cho cấu trúc trường gió bề mặt (ở độ cao 10m trên mặt biển) trong mùa gió Đông – Bắc với đặc trưng cấu trúc đáng lưu ý là sự tồn tại tâm phân kỳ gió ở vùng biển Tây Bắc đảo Luzon và tâm hội tụ gió ở vùng biển Nam Trung bộ và Đông Nam bộ Việt Nam (hình 1). Chính tâm phân kỳ gió này là nguyên nhân gây nên vùng phân kỳ nước (nước trồi) mạnh ở vùng Tây – Bắc đảo Luzon, và tâm hội tụ gió là nguyên nhân gây nên vùng hội tụ nước (nước chìm) mạnh ở vùng biển Nam Trung bộ, như các tác giả đã phát hiện trước đây [2].

Thành phần thứ hai EOF2 đặc trưng cho cấu trúc trường gió bề mặt trong mùa gió Tây – Nam với đặc trưng cấu trúc đáng lưu ý là sự tồn tại tâm phân kỳ gió rộng lớn bao trùm lấy toàn dải ven bờ và thềm lục địa Trung và NamViệt Nam với trung tâm là vùng biển Nam Trung bộ (hình 2), chính đây là nguyên nhân gây nên hiện tượng nước trồi mạnh ở vùng biển này [2].

Thành phần thứ ba EOF3 và thứ tư EOF4 với tải trọng cực đại trong các mùa gió chuyển tiếp, khi mà ở vùng cực bắc Biển Đông còn gió mùa Đông Bắc yếu, thì ở vùng giữa Biển Đông thịnh hành gió Tây kéo theo hiện tượng phân kỳ gió ở vùng biển phía Tây đảo Luzon (hình 3) và hội tụ gió trên thềm lục địa Nam Biển Đông (hình 4).

Như vậy, rõ ràng là mối liên hệ giữa các kết quả phân tích EOF của trường gió bề mặt (cụ thể là 4 thành phần chính đầu tiên) với một số đặc trưng cấu trúc thủy văn (ví dụ như hiện tượng hội tụ và phân kỳ nước tầng mặt hay hiện tượng nước trồi, nước chìm) là rõ ràng và tường minh hơn nhiều so với trường hợp sử dụng trường gió tổng hợp trung bình. Ngoài ra, từ kết quả phân tích trên đây, có thể rút ra nhận xét quan trọng là khác với các trường gió tổng hợp trung bình, các thành phần gió chính trên Biển Đông hoàn toàn không đồng nhất theo không gian và thường tồn tại những vùng gió khá mạnh cùng với những vùng khá lặng gió (xem hình 1 -4). Dạng chuyển động phổ biến của trường gió bề mặt là chuyển động xoáy.

Trên đây là kết quả phân tích hàm trực giao thực nghiệm EOF cho trường gió bề mặt. Phương pháp này có thể áp dụng cho các trường khí tượng - thủy văn khác như nhiệt độ, độ muối, áp suất, lượng mưa, mây, dòng chảy, Chlorophyll ... hoặc có thể áp dụng cho nhiều trường (yếu tố) kết hợp lại, nhằm tìm ra những quy luật biến đổi, những nhân tố chính quyết định hoặc gây ảnh hưởng tới các quá trình biến đổi trong tự nhiên. Các kết quả phân tích EOF có thể được sử dụng làm số liệu đầu vào cho các mô hình tính toán và dự báo tương ứng. Với số liệu đầu vào như vậy các kết quả tính toán chắc chắn sẽ tốt hơn và sát thực tế hơn khi sử dụng số liệu tổng hợp trung bình.

V. KẾT LUẬN

Bằng phương pháp phân tích hàm trực giao thực nghiệm (EOF) trên cơ sở sử dụng tập hợp số liệu gió trung bình tháng ở độ cao 10m trên mặt biển với độ phân giải $0.25^{o}x0.25^{o}$ và trong thời hạn 8 năm (1999 – 2007), các thành phần cấu trúc chính của trường gió bề mặt Biển Đông đã được xác định với 4 thành phần chính đầu tiên chiếm 77% toàn bộ thông tin của trường số liệu ban đầu. Cụ thể là:

Thành phần thứ nhất EOF1 với tải trọng cực đại vào mùa gió Đông Bắc, phản ánh cấu trúc trường gió bề mặt trong mùa này.

Thành phần thứ hai EOF2 với tải trọng cực đại vào mùa gió Tây Nam, phản ánh cấu trúc trường gió bề mặt trong mùa này.

Thành phần thứ ba EOF3 và thứ tư EOF4 với tải trọng cực đại vào các mùa chuyển tiếp, phản ánh cấu trúc trường gió bề mặt trong các mùa tương ứng.

Đặc trưng quan trọng của thành phần EOF1 là sự tồn tại vùng phân kỳ gió gây nên hiện tượng nước trồi mạnh ở vùng biển tây bắc đảo Luzon và vùng hội tụ gió gây nên hiện tượng nước chìm mạnh ở vùng biển Nam Trung bộ và Đông Nam bộ Việt Nam trong mùa gió Đông Bắc.

Đặc trưng quan trọng của thành phần EOF2 là tồn tại vùng phân kỳ gió rộng lớn gây nên hiện tượng nước trồi mạnh ở vùng biển Nam Trung bộ và Đông Nam bộ Việt Nam.

Đặc trưng quan trọng của thành phần EOF3 và EOF4 là sự tồn tại gió Tây ở trung phần Biển Đông và vùng phân kỳ gió ở vùng biển phía Tây đảo Luzon cũng như vùng hội tụ gió trên thềm lục địa Nam Biển Đông trong mùa chuyển tiếp.

Các thành phần khác là rất nhỏ và không ổn định, nên có thể bỏ qua.

Lời cảm ơn: Bài báo được hoàn thành với sự hỗ trợ kinh phí của Hội đồng Khoa học Tự nhiên và Hội đồng chuyên ngành các khoa học Trái đất. Các tác giả xin chân thành cảm ơn

TÀI LIỆU THAM KHẢO

1. **Kelly K.A. 1988:** Components on the EOF analysis of AVHRR – SST pattern in Santa Barbara Chanel by Lagerloef and Bernstein .J. Geophysic Res.
2. **Võ Văn Lành, Tống phước Hoàng Sơn, Nguyễn Văn Tuân, 2007,** "Những vùng nước trồi, nước chìm mạnh và ổn định trong lớp nước tầng mặt Biển Đông". Tạp chí Khoa học và Công nghệ biển: 2(T7): 49 - 58.
3. **Smirnov N.P. , Vainovsky P.A.,Titov Iu.E.,** Chuẩn đoán và dự báo thống kê các quá trình hải dương học *(người dịch: Phạm Văn Huấn)*, NXB Đại học Quốc gia Hà Nội, 2005, 276 trg
4. **William J. Emery, Richard E. Thomson,** Data Analysis Methods in Physical Oceanography, 1998, 634pp

ANALYSING THE MAIN COMPONENT STRUCTURE OF THE SURFACE WIND FIELD OF THE SOUTH CHINA SEA BY METHOD OF EMPIRICAL ORTHOGONAL FUNCTION ANALYSIS

NGUYEN VAN TUAN, VO VAN LANH, TONG PHUOC HOANG SON

Summary: *The paper presents the results of analyzing the main component structure of the surface (at the height of 10m above sea surface) wind field of the South China Sea by method of Empirical Orthogonal Function analysis. The used dataset is collection of mean monthly wind data for the meshes of $0.25^o x 0.25^o$ of all South China Sea during 8 years from 1999 to 2007. The results of analyzing indicated that the surface wind field of the South China Sea can be approximated by following 4 main components:*

- The first component EOF1 is characterized for surface wind field in the season of the North – East monsoon with the existence of the strong wind divergence causing the strong water upwelling in the region north – west of the Luzon Island, and the strong wind convergence causing strong water downwelling in the region east of the Central and South Vietnam [2].

- The second component EOF2 is characterized for surface wind field in the season of the South – West monsoon with existence of strong wind divergence causing strong water upwelling in the region east of the Central and South Vietnam [2].

- The third and fourth components EOF3 and EOF4 are characterized for surface wind field in the transitional seasons with existence of West wind in central part of the South China Sea and the wind divergence in the region west of the Luzon Island, and the wind convergence in continental shelf of the south part of the South China Sea.

- Other components are small and unstable, that can be neglected.

Ngày nhận bài: 4 - 5 - 2008

Người nhận xét: TS. Nguyễn Bá Xuân

Địa chỉ: Viện Hải Dương học