

SỬ DỤNG PHƯƠNG PHÁP NGOẠI SUY THỐNG KÊ TỐI ƯU ĐỂ DỰ BÁO DÀI HẠN CÁC ĐẶC TRƯNG KHÍ TƯỢNG

Phạm Văn Huấn

Trường Đại học Khoa học Tự nhiên, ĐHQG Hà Nội

Tóm tắt. Dự báo dài hạn các yếu tố khí tượng trên biển nhằm có được một thông tin định hướng nào đó về điều kiện biển phục vụ các hoạt động kỹ thuật của con người trên biển. Trong bài này giới thiệu những kết quả thử nghiệm phương pháp ngoại suy tuyến tính tối ưu để thực hiện dự báo một số yếu tố khí tượng biển quy mô lớn.

Kết quả thử nghiệm dự báo thấy rằng để đạt hệ số tương quan chung giữa các giá trị thực và các giá trị dự báo ở mức không nhỏ hơn 0,98 và độ đảm bảo dự báo trên 95% thường cần tối thiểu hai chục số hạng ở vế phải của phương trình dự báo. Để đạt được những dự báo với độ đảm bảo cao nhất có thể sử dụng tối trên 50, thậm chí trên 60 số hạng ở vế phải của phương trình dự báo. Đáng chú ý là thử nghiệm dự báo với những thời hạn dự báo khác nhau (đến 12 bước thời gian) đều cho những kết quả tốt như nhau.

1. Về phương pháp ngoại suy tuyến tính tối ưu và áp dụng vào dự báo dài hạn

Phương pháp này, còn gọi là phương pháp thống kê động lực, dựa trên giả thiết về sự tồn tại những quy luật nội tại trong sự biến thiên thời gian của các quá trình khí tượng, thủy văn vĩ mô [1, 4] chịu tác động đồng thời của nhiều yếu tố ảnh hưởng.

Tư tưởng của phương pháp do I.M. Alekhin đề xướng nhằm đối tượng là những quá trình cỡ lớn, tức quá trình được lấy trung bình trên quy mô rộng theo không gian hoặc (và) theo thời gian để đảm bảo nó là hệ quả của nhiều nguyên nhân, trong đó các nguyên nhân cùng có ảnh hưởng đều như nhau, không trội hẳn so với nhau. Những nguyên nhân này về phần mình lại là hệ quả của hàng loạt các quá trình khác, tức có sơ đồ hình cây của các nguyên nhân tác động tới yếu tố mà ta cần dự báo. Biến động nhiều hướng của vô số những nguyên nhân ấy thiết lập trong yếu tố được dự báo một chế độ dao động ổn định trong thời gian, đặc trưng bởi tính liên hệ nội tại giữa những giá trị của nó trong tiền sử, hiện tại và tương lai. Tính liên hệ nội tại này thể hiện ở sự ổn định của hàm tự tương quan. Một khi hàm tự tương quan của yếu tố ổn định, ta có thể ngoại suy yếu tố đó một cách tin cậy.

Trong thực tế, nếu chuỗi quan trắc đủ dài, ta có thể kiểm tra sự ổn định của hàm tự tương quan bằng cách tính hàm này trong những giai đoạn quan trắc khác nhau và so sánh với nhau. Vì vậy, với yếu tố khí tượng hải văn lấy trung bình theo tháng, mùa hoặc năm, hoặc những đặc trưng trung bình của cả một vùng biển, của một mặt cắt với hàm tự tương quan ổn định đều có thể sử dụng phương pháp dự báo này. Xét theo nghĩa đó, phương pháp dự báo ta đang nghiên cứu có tính vạn năng, nghĩa là nó có thể sử dụng để dự báo nhiều yếu tố tự nhiên quy mô lớn.

Giá trị cần dự báo q_t (là giá trị quy chuẩn theo trị số trung bình của đại lượng cần dự báo Q) có thể được biểu diễn dưới dạng một quan hệ tuyến tính với các giá trị đã biết của nó ở những thời điểm trước đó bằng phương trình

$$q_t = K_{m,1}q_{t-m} + K_{m,2}q_{t-m-1} + \dots + K_{m,\theta}q_{t-m-\theta+1} \quad (1)$$

trong đó m – thời gian báo trước của dự báo, gọi tắt là thời hạn dự báo, $m = 1, 2, \dots$; θ – số lượng các giá trị đã biết của đại lượng q được dùng trong phương trình dự báo.

Những hệ số ngoại suy tuyến tính $K_{m,1}, K_{m,2}, \dots, K_{m,\theta}$ ứng với một giá trị xác định của m ,

làm thành hàm các hệ số ngoại suy tuyến tính K_m , được xác định thực nghiệm từ quan trắc thực tế. Người ta thường sử dụng phương pháp bình phương nhỏ nhất để xác định những trị số của hàm K_m . Theo phương pháp này, những trị số $K_{m,i}$, $i = 1, 2, \dots, \theta$, được xác định sao cho tổng của các bình phương của sai số ngoại suy theo công thức (1) so với các quan trắc thực tế đạt cực tiểu

$$\sum_{t=1}^{N-\theta} (q_t - K_{m,1}q_{t-m} - K_{m,2}q_{t-m-1} - \dots - K_{m,\theta}q_{t-m-\theta+1})^2 = \min \quad (2)$$

với N – tổng số các quan trắc về đại lượng Q .

Khảo sát điều kiện cực trị của (2) sẽ dẫn tới một hệ phương trình chuẩn tắc sau đây để tính những trị số của hàm K_m :

$$\begin{aligned} K_{m,1}r_0 + K_{m,2}r_1 + \dots + K_{m,\theta}r_{\theta-1} &= r_m \\ K_{m,1}r_1 + K_{m,2}r_0 + \dots + K_{m,\theta}r_{\theta-2} &= r_{m+1} \\ \dots & \\ K_{m,1}r_{\theta-1} + K_{m,2}r_{\theta-2} + \dots + K_{m,\theta}r_0 &= r_{m+\theta-1} \end{aligned} \quad (3)$$

trong đó r – hàm tự tương quan.

Thấy rằng việc xác định các trị số của hàm các hệ số ngoại suy tuyến tính K_m quy về việc giải hệ các phương trình đại số tuyến tính gồm θ phương trình với θ ẩn số. Với những m khác nhau, các hệ phương trình ấy sẽ chỉ khác nhau ở những số hạng tự do vế phải.

Như vậy các bước tính toán để thực hiện mô hình dự báo bao gồm:

a) Thiết lập chuỗi thời gian những giá trị quan trắc của đại lượng q quy chuẩn theo trị số trung bình của chuỗi

$$q_i = Q_i - \sum_{t=1}^N Q_t, \quad i = 1, 2, \dots, N;$$

b) Tính các giá trị của hàm tự tương quan chuẩn hoá theo công thức

$$r_k = \frac{\sum_{i=1}^{N-k} q_i q_{i+k}}{\sqrt{\sum_{i=1}^{N-k} (q_i)^2 \sum_{j=k}^N (q_j)^2}}, \quad k = 0, 1, \dots, m + \theta - 1;$$

c) Giải hệ phương trình chuẩn tắc (3) bằng một phương pháp quen thuộc trong phương pháp tính như phương pháp Gauss hoặc phương pháp lặp Seiden.

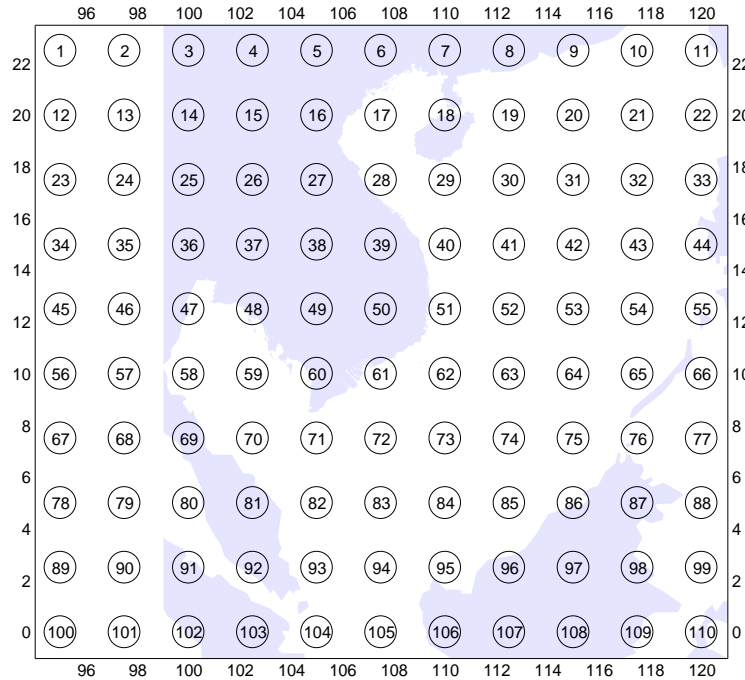
Kinh nghiệm dự báo các quá trình tự nhiên quy mô lớn bằng phương pháp ngoại suy tối ưu cho thấy rằng ứng với số lượng θ các số hạng ở vế phải của (1) khác nhau sẽ cho hiệu quả dự báo khác nhau. Người ta cho rằng, tùy thuộc vào cấu trúc biến động dao động của mỗi quá trình dự báo mà tồn tại những giá trị θ tối ưu làm cho dự báo quá trình đó đạt hiệu quả cao nhất. Tác giả của phương pháp và nhiều người áp dụng phương pháp này vào các quá trình trong thủy văn và hải dương học đã chú ý khảo sát nhằm xác định giá trị tối ưu của θ đối với từng yếu tố dự báo cụ thể và tìm được những giá trị tối ưu nằm trong khoảng từ 8 đến 40 bước thời gian (tháng hoặc năm, tùy thuộc độ gián đoạn quan trắc hay quy mô lấy trung bình các đại lượng).

Phương pháp thống kê động lực đã được nhiều tác giả dùng để dự báo giá trị trung bình tháng, trung bình mùa và năm của nhiệt độ nước Bắc Đại Tây Dương. Tác giả báo cáo này cũng đã từng thử nghiệm phương pháp cho một số chuỗi nhiệt độ nước biển, nhiệt độ không khí tại một số trạm hải văn ven biển và thấy rằng phương pháp cho kết quả khá tốt [2, 3].

2. Số liệu quan trắc

Trường các yếu tố khí tượng trung bình ngày tại vùng biển Đông thu thập được trong giai đoạn từ 1979 đến 2002. Trên miền nghiên cứu giới hạn trái bởi kinh tuyến 95°Đ, giới hạn phải bởi kinh tuyến 120°Đ, giới hạn dưới bởi vĩ tuyến 0° và giới hạn trên bởi vĩ tuyến 22,5°B được chia

thành mạng lưới các ô vuông cạnh 2,5° kinh vĩ. Tại mỗi điểm nút lưới ghi lại giá trị trung bình ngày của từng yếu tố cần dự báo. Từ dạng số liệu này đã tính ra các giá trị trung bình tháng và lập thành ma trận số liệu trung bình tháng gồm 288 dòng ứng với các tháng từ tháng 1 năm 1979 đến tháng 12 năm 2002 và 110 cột ứng với 110 điểm phân bố đều trên miền tính (hình 1).



Hình 1. Số hiệu các điểm trên miền dự báo

Cách thức xây dựng các ma trận số liệu như trên áp dụng cho tất cả các yếu tố quan trắc như áp suất khí quyển, nhiệt độ không khí, độ ẩm không khí...

3. Kết quả thử nghiệm dự báo

Những thủ tục tính toán theo sơ đồ phương pháp ngoại suy thống kê tối ưu được thực hiện trong chương trình ALE do chúng tôi xây dựng chuyên dụng cho phương pháp để xử lý các chuỗi thời gian của các yếu tố khí tượng hoặc thủy văn quy mô lớn.

Chương trình tính bao gồm cả thủ tục tự động khảo sát số lượng tối ưu các số hạng ở vế phải của phương trình dự báo. Trị số tối ưu của θ được xác định bằng cách thiết lập các phương trình dự báo với các θ khác nhau, biến đổi từ 1 đến 60-70, ứng với mỗi phương trình dự báo thực hiện dự báo kiểm tra lại trên chuỗi số liệu phụ thuộc cho từng số hạng trong chuỗi, tính các hệ số tương quan giữa chuỗi quan trắc và chuỗi nhận được, tính độ đảm bảo của dự báo rồi xác định θ tối ưu.

Với mỗi chuỗi đã thiết lập những phương trình dự báo với số θ biến đổi từ 1 đến 65 cho những thời hạn dự báo từ 1 đến 12 bước thời gian. Thực hiện các dự báo kiểm tra trên tất cả các chuỗi. Đồng thời cũng thử thực hiện những dự báo độc lập, bằng cách trích phần cuối các chuỗi gồm cỡ vài chục số liệu ra khỏi chuỗi, không đưa vào tính các hàm tự tương quan, sau đó dự báo và kiểm tra trên các số liệu đó. Với tất cả các chuỗi số liệu, để đạt hệ số tương quan chung R_0 giữa giá trị thực và các dự báo ở mức không nhỏ hơn 0,98 và độ đảm bảo dự báo trên 95 % thường cần tối cỡ hơn hai chục số hạng ở vế phải của phương trình dự báo (1). Để đạt được những dự báo với độ đảm bảo cao nhất có thể sử dụng tới cỡ 30-60 số hạng của phương trình. Những dự báo với độ đảm bảo cao đạt được khi sử dụng trên 50, thậm chí trên 60 số hạng là do với số lượng các số hạng lớn như vậy có thể bao quát đủ các thông tin về biến trình năm của các yếu tố, đồng thời tính đến cả những chu trình dao động rõ nhất của đại đa số các yếu tố khí tượng hải văn vùng biển là chu trình cỡ hai năm và chu trình cỡ 5-7 năm.

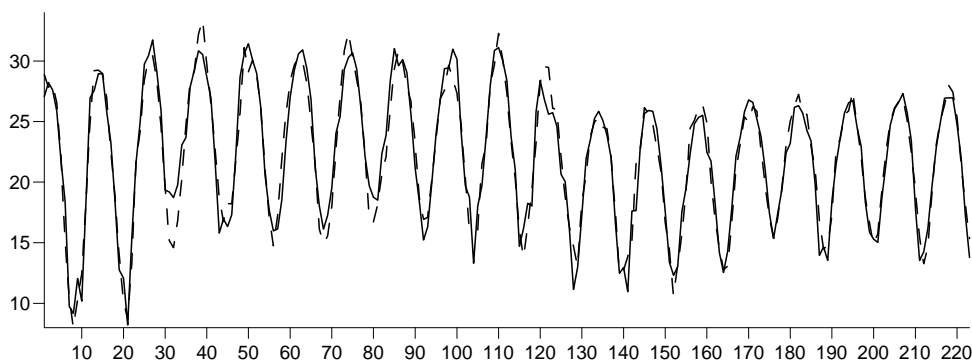
Đáng chú ý là thử nghiệm dự báo với những thời hạn dự báo khác nhau (có thể đến 12 bước thời gian) đều cho những kết quả khá như nhau. Bởi vì về thực chất sơ đồ này cho phép khôi phục đúng xu thế của quá trình không những trong tương lai kế cận thời điểm dự báo mà cả một thời đoạn khá dài.

Khi kiểm tra dự báo phụ thuộc, toàn bộ độ dài các chuỗi số liệu nhiệt độ không khí (288 tháng) được sử dụng để tính các hàm tự tương quan. Để kiểm tra dự báo độc lập, một phần của chuỗi quan trắc, ở đây lấy bằng 36 tháng cuối cùng, đã bị loại khỏi chuỗi, không được đưa vào tính hàm tự tương quan.

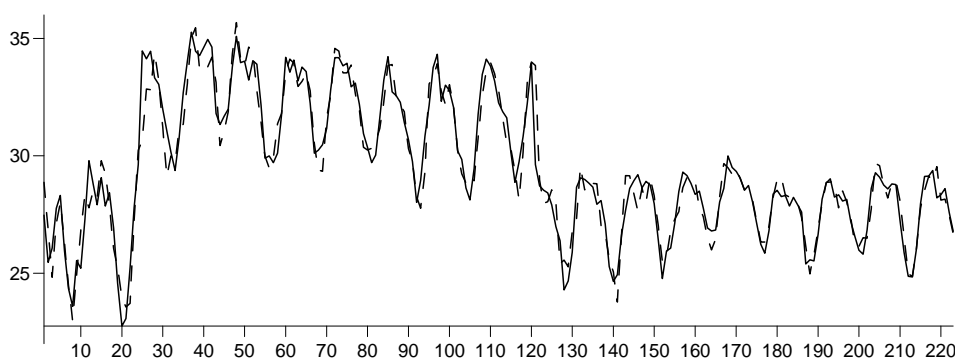
Trong bảng 1 sẽ giới thiệu kết quả thử nghiệm dự báo nhiệt độ không khí mực mặt biển tại một số vùng điển hình có đặc điểm biến thiên theo thời gian khác nhau. Trên các hình 2-5 là thí dụ so sánh các trị số quan trắc và dự báo phụ thuộc đối với chuỗi nhiệt độ không khí mực mặt biển tại một số điểm của miền tính. Thời gian dự báo từ tháng 6 năm 1984 đến 12 năm 2002. Trên các hình 6-10 so sánh các trị số quan trắc và dự báo độc lập trong ba năm 2000-2002 tại một số điểm.

Đã thực hiện dự báo độc lập nhiệt độ, áp suất không khí mực mặt biển với thời hạn dự báo 1 tháng cho tất cả các điểm còn lại trên miền tính cho năm 2002. Trên các hình 11-14 trực quan thể hiện kết quả khôi phục các trường bằng phương pháp dự báo ngoại suy thống kê tối ưu. Trên các bản đồ cũng dẫn các đường đẳng trị nhiệt độ hoặc khí áp quan trắc để tiện nhận định về sự trùng hợp khá cao giữa dự báo và quan trắc.

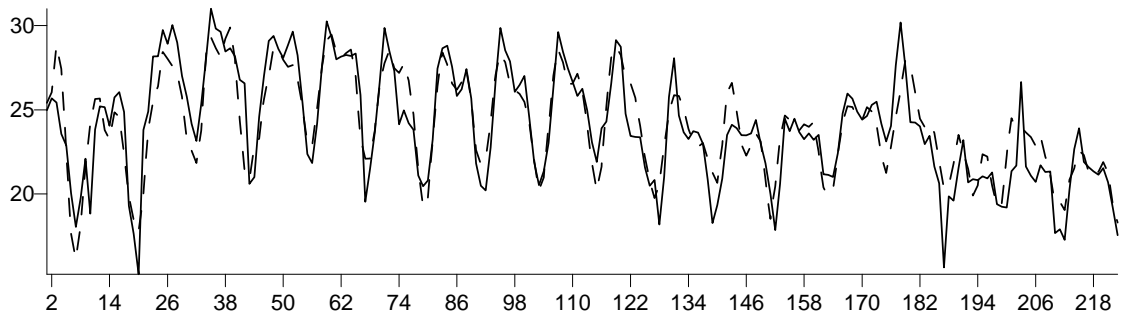
Thấy rằng, hệ số tương quan giữa các giá trị dự báo và giá trị quan trắc khá cao, tại phần lớn các điểm hệ số tương quan đều lớn hơn 0,90. Sai số bình phương trung bình thường nhỏ hơn 1°C . Độ đảm bảo dự báo trong tất cả các vùng đều đạt trên 90% (bảng 1). Đặc biệt không có sự khác nhau về hiệu quả giữa dự báo độc lập và dự báo phụ thuộc. Điều đó cho thấy phương pháp dự báo có tính tin cậy. Từ các bản đồ so sánh thấy rằng các trường dự báo khá phù hợp với các trường số liệu quan trắc.



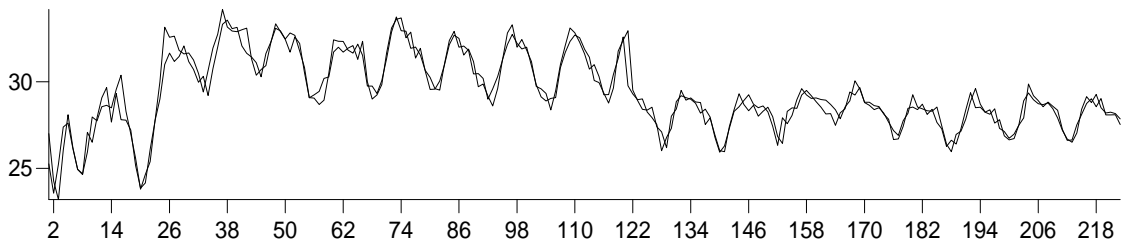
Hình 2. So sánh quan trắc (đường liền nét) và dự báo phụ thuộc (đường gạch nối):
điểm số 15 ($102,5^{\circ}\text{Đ} - 20,0^{\circ}\text{B}$) thời kỳ 1984 - 2002



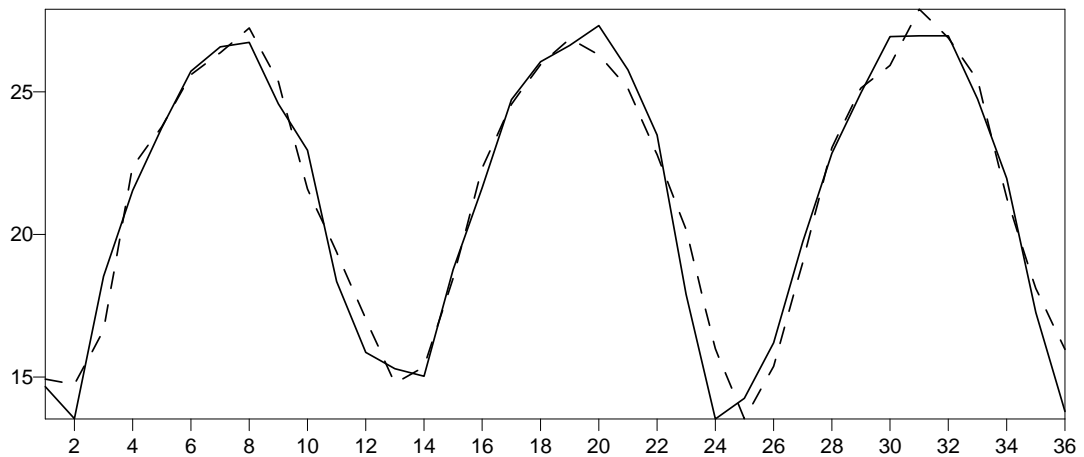
Hình 3. So sánh quan trắc (đường liền nét) và dự báo phụ thuộc (đường gạch nối):
điểm số 49 ($105,0^{\circ}\text{Đ} - 12,5^{\circ}\text{B}$) thời kỳ 1984 - 2002



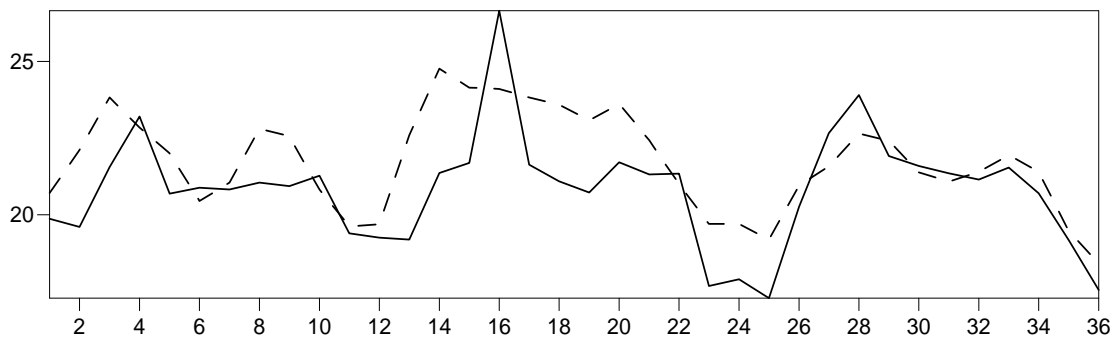
Hình 4. So sánh quan trắc (đường liền nét) và dự báo phụ thuộc (đường gạch nối):
điểm số 20 (115,0°Đ - 20,0°B) thời kỳ 1984 - 2002



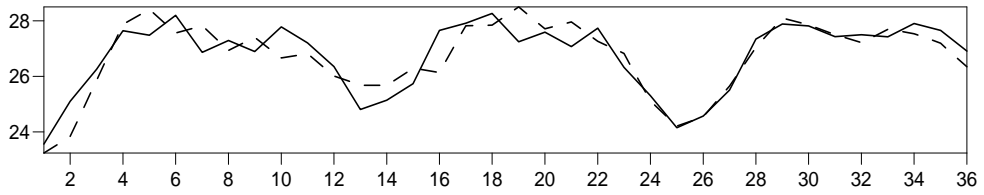
Hình 5. So sánh quan trắc (đường liền nét) và dự báo phụ thuộc (đường gạch nối):
điểm số 70 (102,5°Đ - 7,5°B) thời kỳ 1984 - 2002



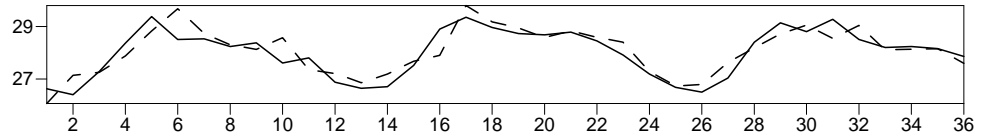
Hình 6. So sánh quan trắc (đường liền nét) và dự báo độc lập (đường gạch nối):
điểm số 15 (102,5°Đ - 20,0°B) năm 2000 - 2002



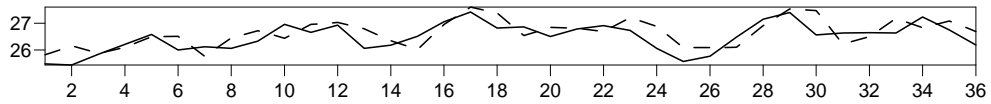
Hình 7. So sánh quan trắc (đường liền nét) và dự báo độc lập (đường gạch nối):
điểm số 20 (115,0°Đ - 20,0°B) năm 2000 - 2002



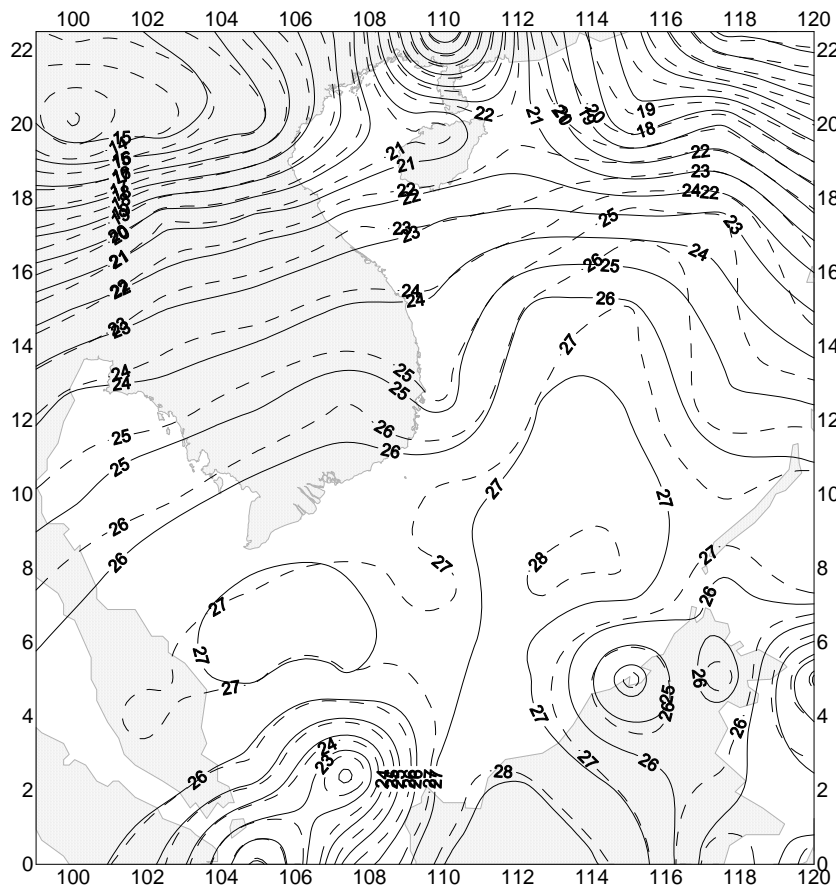
Hình 8. So sánh quan trắc (đường liền nét) và dự báo độc lập (đường gạch nối):
điểm số 40 (110,0°Đ - 15,0°B) năm 2000 - 2002



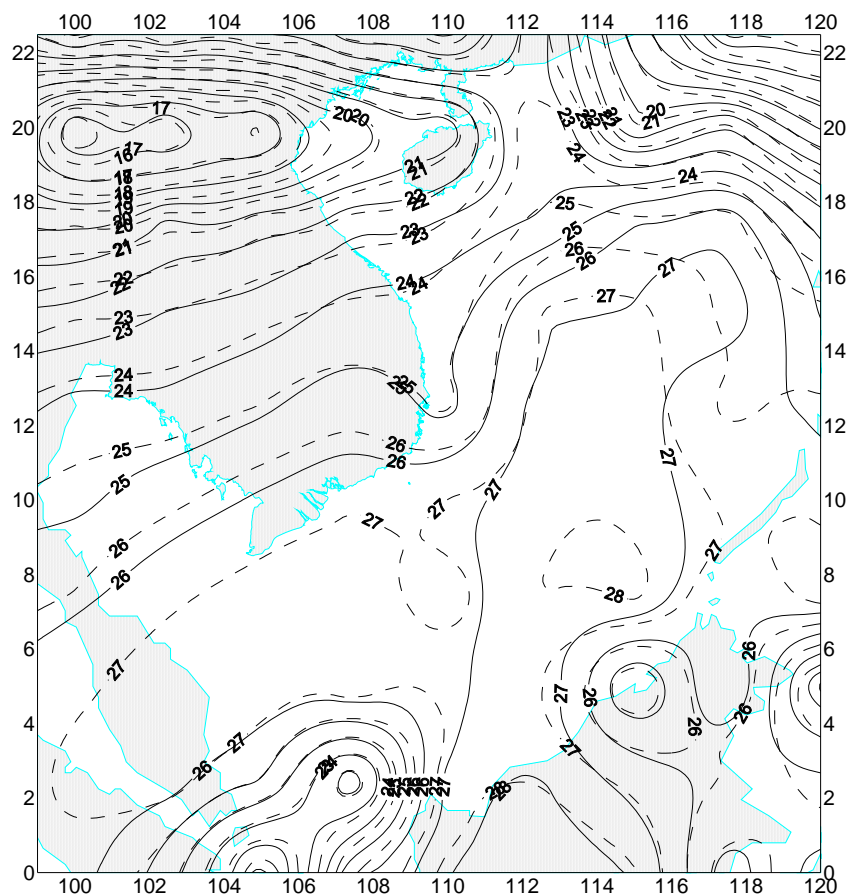
Hình 9. So sánh quan trắc (đường liền nét) và dự báo độc lập (đường gạch nối):
điểm số 70 (102,5°Đ - 7,5°B) năm 2000 - 2002



Hình 10. So sánh quan trắc (đường liền nét) và dự báo độc lập (đường gạch nối):
điểm số 105 (107,5°Đ - 0,0°B) năm 2000 - 2002



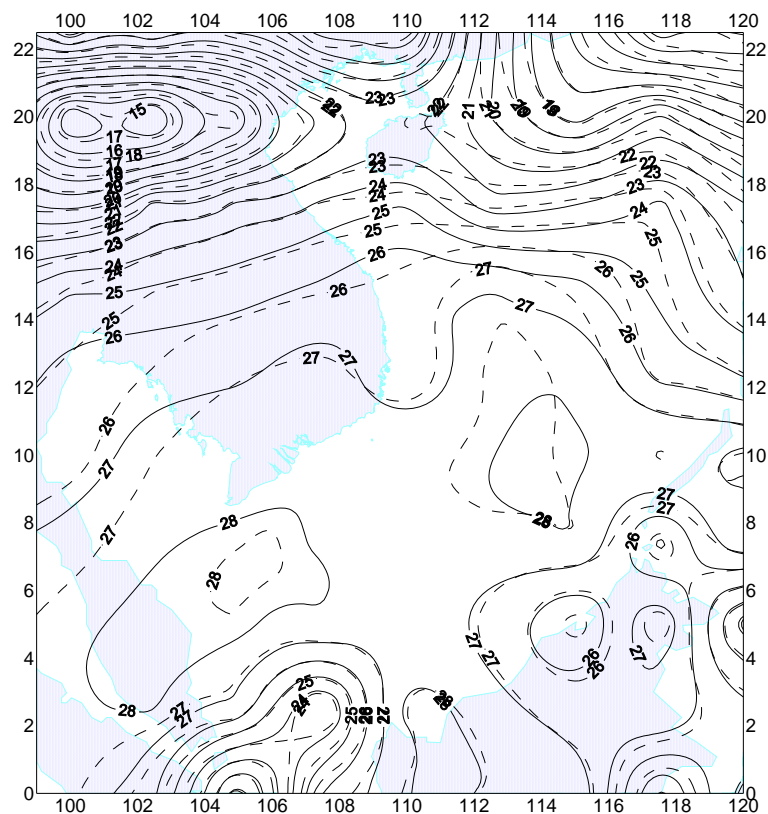
Hình 11. So sánh trường nhiệt độ không khí tháng 1 - 2003
(đường liền nét - quan trắc, đường gạch nối - dự báo)



Hình 12. So sánh trường nhiệt độ không khí tháng 2 - 2003
(đường liền nét - quan trắc, đường gạch nổi - dự báo)

Bảng 1. Kết quả đánh giá dự báo nhiệt độ không khí mực mặt biển với thời hạn dự báo 1 tháng

Số hiệu điểm	Các đặc trưng thống kê		Số số hạng tối ưu	Hiệu quả dự báo		
	Biên độ, °C	Phương sai		Hệ số tương quan giữa dự báo và quan trắc	Sai số bình phương trung bình, °C	Độ đảm bảo, %
<i>Dự báo phụ thuộc</i>						
15	28,48	31,06	49	0,97	1,43	100
49	17,85	8,27	59	0,96	0,79	100
20	21,42	10,95	35	0,92	1,30	99
40	15,16	4,26	24	0,93	0,74	98
70	15,04	4,66	52	0,95	0,65	99
105	10,41	2,80	65	0,93	0,61	99
Tính chung cho tất cả 110 điểm				0,97	0,95	
<i>Dự báo độc lập</i>						
15				0,97	0,99	
49				0,93	0,49	
20				0,72	1,48	
40				0,88	0,62	
70				0,86	0,46	
105				0,69	0,42	
Tính chung cho tất cả 110 điểm				0,98	0,63	



Hình 13. So sánh trường nhiệt độ không khí tháng 12-2003
(đường liền nét - quan trắc, đường gạch nối - dự báo)

4. Kết luận

Phương pháp ngoại suy tối ưu với thủ tục tính toán đơn giản đã cho phép dự báo hiệu quả các yếu tố khí tượng thủy văn biển. Với chuỗi số liệu lịch sử cỡ vài chục tháng có thể cho phép ta ngoại suy một cách khá tin cậy nhiều yếu tố khí tượng.

Phương pháp dự báo tốt đối với tất cả các vùng đất liền cũng như vùng biển với đặc điểm biến thiên thời gian rất khác nhau của các yếu tố khí tượng. Phương pháp cho phép dự báo khá tin cậy với các thời hạn dự báo khác nhau, thậm chí với thời hạn dự báo tới 12 bước thời gian và hơn nữa, đảm bảo có thể áp dụng vào dự báo dài hạn.

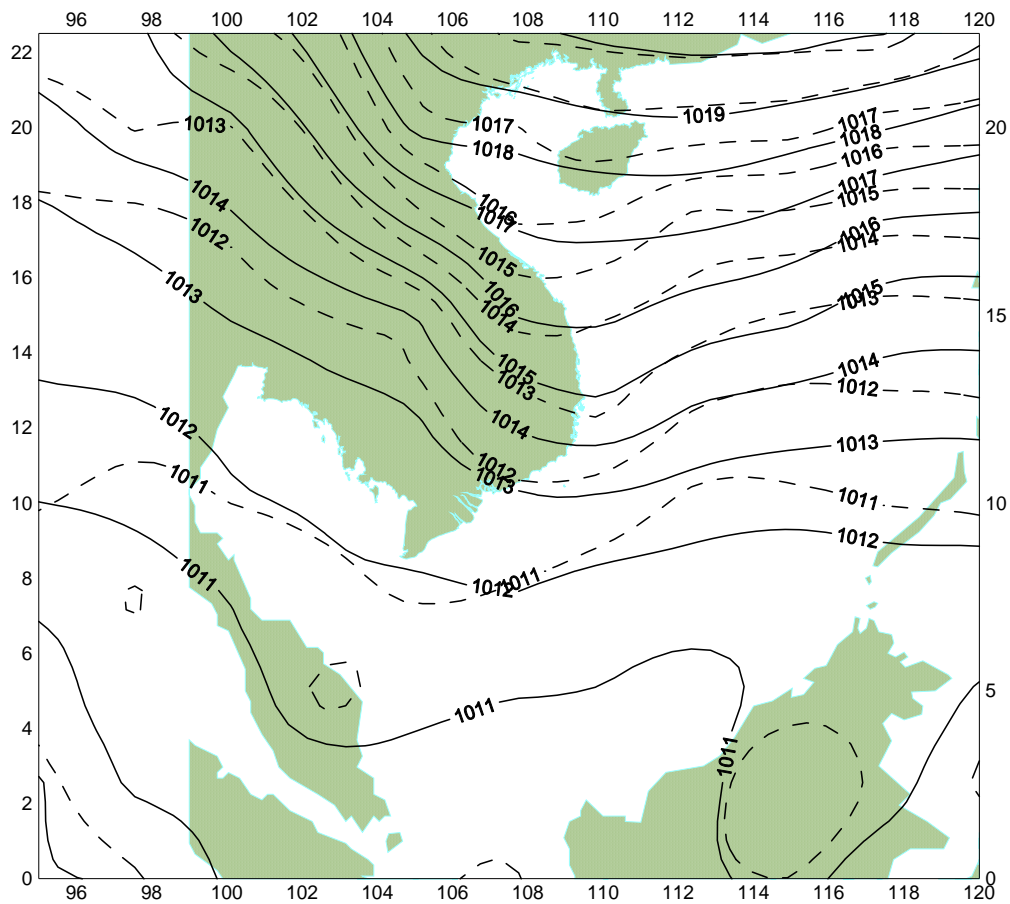
Cơ sở dữ liệu và chương trình máy tính được xây dựng là một công cụ tiện ích để thực hiện nghiên cứu thử nghiệm dự báo với nhiều yếu tố khí tượng khác. Đồng thời nó cũng có thể sử dụng để triển khai các dự báo nghiệp vụ trong thực tế.

Công trình này là kết quả nghiên cứu trong khuôn khổ đề tài cấp nhà nước "Xây dựng và áp dụng những mô hình dự báo các trường khí tượng, thủy văn vùng Biển Đông" mã số KC.09.04 và đề tài "Phân tích và dự báo các trường khí tượng thủy văn biển Đông" mã số 74.30.04 thuộc Chương trình nghiên cứu cơ bản, 2004.

Tài liệu tham khảo

- [1] Phạm Văn Huấn. *Dự báo thủy văn biển*. Nxb Đại học Quốc gia Hà Nội. Hà Nội, 2003, 165 tr.
- [2] Nguyễn Tài Hối, Phạm Văn Huấn. *Thử nghiệm phương pháp ngoại suy thống kê tuyến tính để dự báo những yếu tố khí tượng thủy văn biển*. Tạp chí Khí tượng thủy văn, 2 (434), 1997, tr. 30-35.
- [3] Nguyễn Tài Hối, Phạm Văn Huấn. *Thử nghiệm phương pháp động lực thống kê để dự báo những yếu tố khí tượng thủy văn biển*. Tổng cục KTTV, Viện KTTV, Tập báo cáo công trình NCKH Hội nghị khoa học lần thứ VI, tập 1, 1997, tr. 140-145.

[4] Вопросы статистико-динамического прогноза и спектрального анализа временных рядов. Сборник трудов. ЛГМИ, Ленинград, № 56, 1975



Hình 14. So sánh trường khí áp mực biển tháng 1 - 2002
(đường liền nét - quan trắc, đường gạch nối - dự báo)

APPLICATION OF THE OPTIMAL STATISTICAL EXTRAPOLATION FOR LONG-TERM PROGNOSIS OF METEOROLOGICAL PARAMETERS

Pham Van Huan

University of Sciences, VNUH

The prognosis of meteorological parameters over the sea surface aims at the obtainment of the some information on hydrometeorological conditions useful for technique activities of men on the sea. This paper presents the results of experiments of using the optimal linear extrapolation method for predicting macro-scale marine meteorological processes.

The experiments pointed out that in order to obtain a high correlation between observed and calculated values and a sufficient efficiency of the prognoses a number of terms in the right of the prognosis equations must be more than twenty. And for the highest prognosis efficiency this number would be larger than 50, even larger than 60. In addition, the prognoses with different periods of notice (up to 12 time steps) gave the same level of success.