

THỬ NGHIỆM PHƯƠNG PHÁP NGOẠI SUY THỐNG KÊ TUYẾN TÍNH ĐỂ DỰ BÁO NHỮNG YẾU TỐ KHÍ TƯỢNG THỦY VĂN BIỂN

KS. Nguyễn Tài Hoi – Trung tâm KTTV biển
PTS. Phạm Văn Huân - Đại học quốc gia Hà Nội

Cho tới nay các phương pháp tính hoặc dự báo những yếu tố khí tượng thủy văn biển trên cơ sở thủy động số trị chưa phát triển do độ chính xác còn thấp và những khó khăn liên quan tới việc thiết lập điều kiện biên và điều kiện đầu. Trong khi đó các phương pháp thống kê thuận tiện cho việc tính toán thực tế và với nhiều trường hợp tỏ ra hiệu quả. Dưới đây chúng tôi giới thiệu một số kết quả thử nghiệm với phương pháp ngoại suy thống kê hay phương pháp thống kê động lực của Alekhin [1] và xây dựng chương trình tính toán thực hành thuận lợi để dự báo một số yếu tố hải văn và khí tượng biển lấy trung bình trên quy mô thời gian cỡ tháng hoặc năm.

1. Về phương pháp thống kê động lực dự báo các quá trình vĩ mô

Tư tưởng của phương pháp thống kê động lực do 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 chúng ta cần dự báo. Biến động nhiều hướng của vô số những nguyên nhân ấy thiết lập trong yếu tố chúng ta cần 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 tương quan của yếu tố ổn định, có thể ngoại suy yếu tố đó một cách tin cậy; chúng tôi gọi phương pháp này là phương pháp ngoại suy thống kê tuyến tính theo bản chất tính toán của nó.

Trong thực tế nếu chuỗi quan trắc đủ dài, chúng ta có thể kiểm tra sự ổn định của hàm tương quan bằng cách tính hàm này trong những đoạn quan trắc 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 tương quan ổn định đều có thể sử dụng phương pháp dự báo này. Xét theo nghĩa đó phương pháp dự báo chúng 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.

2. Phương trình dự báo của phương pháp ngoại suy thống kê tuyến tính và thủ tục tính toán

Giá trị dự báo q_t (là đại lượng quy tâm 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 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 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 $k_1, k_2, \dots, k_\theta$ 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 qua quan trắc thực tế. Người ta thường sử dụng phương pháp bình phương nhỏ nhất để xác định những giá trị 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, tức là

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

với N – tổng số các quan trắc của 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 :

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

Ở đây r là hàm tự tương quan của chuỗi thời gian q . 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 trong vế phải.

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

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

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

b) Tính các giá trị của hàm tự tương quan chuẩn hóa 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 Zeiden.

Những thủ tục tính toán theo sơ đồ này được thực hiện trong chương trình do chúng tôi xây dựng chuyên dụng cho phương pháp.

3. Xác định hàm các hệ số ngoại suy tối ưu

Kinh nghiệm dự báo các quá trình tự nhiên quy mô lớn [1] bằng phương pháp ngoại suy thống kê tuyến tính 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 thống kê động lực và nhiều người áp dụng phương pháp này vào các quá trình 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 30 bước thời gian (tháng hoặc năm).

Chương trình tính của chúng tôi cũng bao gồm cả thủ tục tự động khảo sát số lượng tối ưu các số hạng ở vế phải của phương trình dự báo (1). Trị số tối ưu của θ được xác định bằng cách thiết lập các phương trình dự báo dạng (1) 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 bằng công thức (1), tính độ đảm bảo của dự báo rồi xác định θ tối ưu.

Hệ số tương quan chung giữa những giá trị quan trắc q_i và những giá trị dự báo tương ứng q'_i được tính theo công thức

$$R_0 = \frac{\sum_{i=1}^S q_i q'_i}{\sqrt{\sum_{i=1}^S (q_i)^2 \sum_{i=1}^S (q'_i)^2}}$$

ở đây S – tổng số dự báo thử, $S = N - \theta$. Còn độ đảm bảo của dự báo thử (dự báo phụ thuộc) được xác định bằng tỷ số phần trăm giữa số dự báo đúng và tổng số các dự báo thử S đã được thực hiện.

4. Một số kết quả thử nghiệm phương pháp dự báo đối với những yếu tố khí tượng thủy văn biển

Trong mục này nêu một số kết quả thử nghiệm áp dụng phương pháp đối với một số chuỗi thời gian các yếu tố khí tượng hoặc thủy văn biển mà chúng tôi có được. Bảng 1 tóm tắt các đặc trưng của chuỗi số liệu xuất phát.

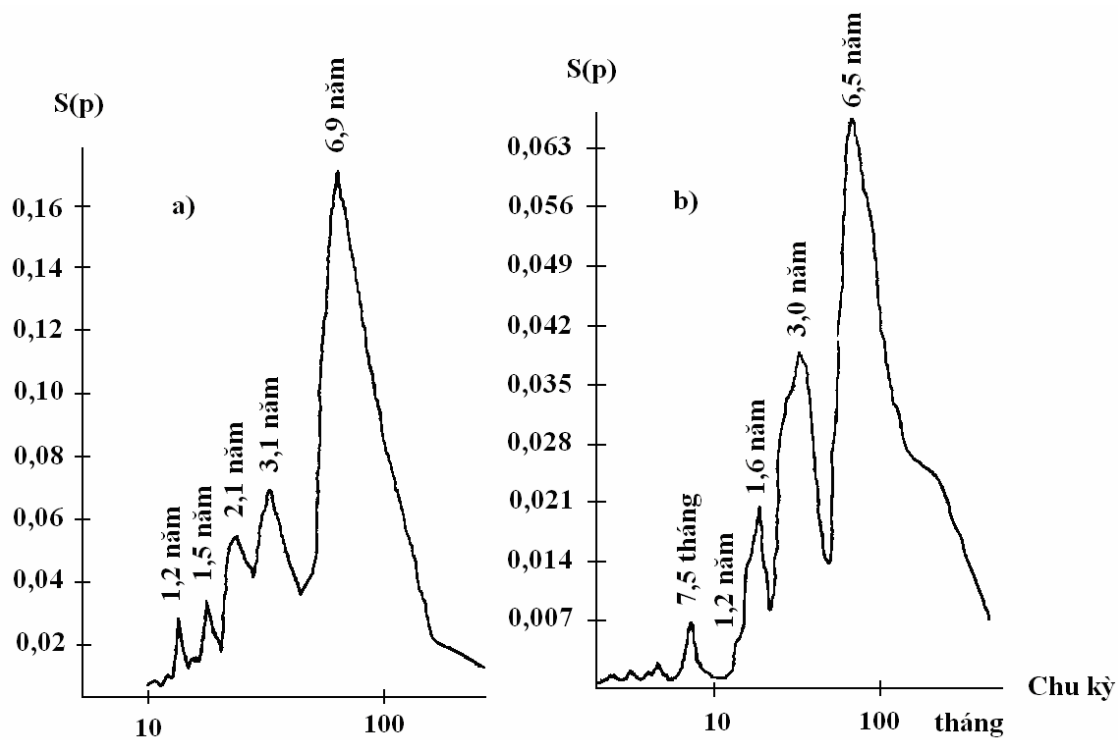
Bảng 1. Những chuỗi số liệu sử dụng để thử nghiệm

TT	Tên chuỗi	Độ gián đoạn	Độ dài chuỗi
1	Nhiệt độ nước biển trạm Hòn Dấu	tháng	441 (4/1956 – 12/1992)
2	Nhiệt độ nước biển trạm Phú Quý	tháng	139 (6/1979 – 12/1990)
3	Nhiệt độ nước biển trạm Côn Đảo	tháng	140 (5/1979 – 12/1990)
4	Nhiệt độ không khí trạm Hòn Dấu	tháng	418 (4/1956 – 12/1990)
5	Nhiệt độ không khí trạm Phú Quý	tháng	139 (6/1979 – 12/1990)
6	Nhiệt độ không khí trạm Côn Đảo	tháng	140 (5/1979 – 12/1990)
7	Lượng mưa tháng trạm Hải Phòng	tháng	480 (1/2005 – 12/1944)

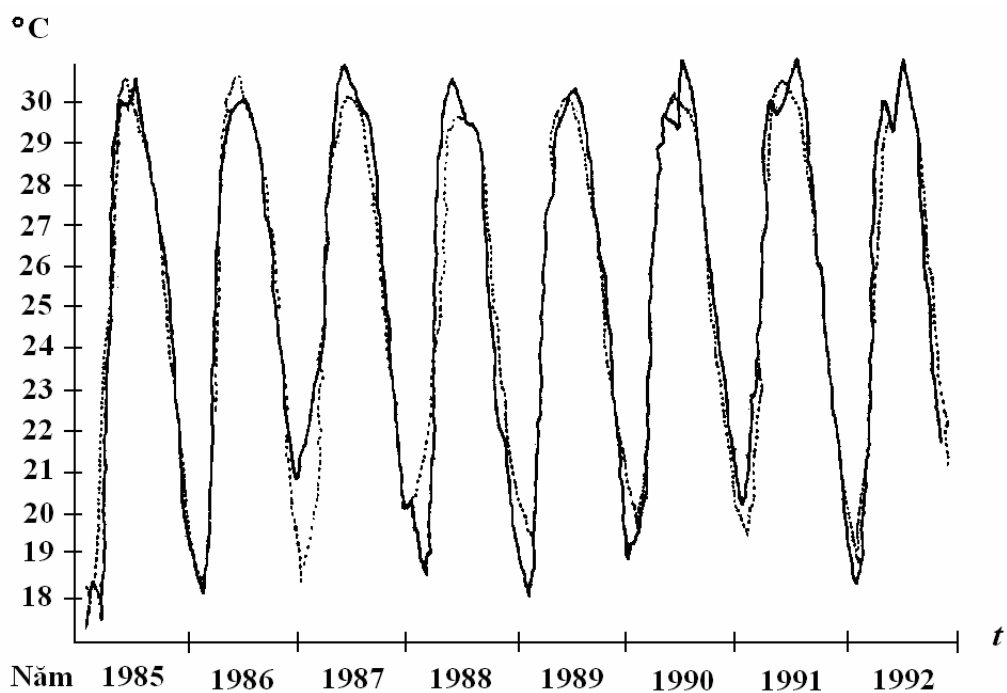
Với những chuỗi này đã thử thiết lập những phương trình dự báo dạng (1) với số θ biến đổi từ 1 đến 65 cho những thời hạn dự báo từ 1 đến 12 bước thời gian. Đã thực hiện các dự báo kiểm tra trên tất cả các chuỗi. Đồng thời cũng thử thực hiện những dự báo độc lập, bằng cách trích phần cuối các chuỗi gồm vài chục số liệu ra khỏi chuỗi, không tham gia vào tính các hàm tự tương quan, sau đó dự báo và kiểm tra trên các số liệu.

Trong khi đánh giá dự báo, chúng tôi sử dụng sai số cho phép bằng một phần năm biên độ tính toán là chỉ tiêu khắt khe hơn so với biên độ tự nhiên của các yếu tố. Biên độ tính toán được tính bằng tổng các giá trị tuyệt đối của các chênh lệch dương và âm lớn nhất của yếu tố đang xét giữa hai lần quan trắc cách nhau khoảng thời gian bằng thời hạn dự báo. Thí dụ, với chuỗi nhiệt độ nước biển ở Hòn Dấu, biên độ tự nhiên tính được bằng 15,5 °C trong khi biên độ tính toán chỉ bằng 12,0 °C.

Thấy rằng, với tất cả các chuỗi số liệu, để đạt hệ số tương quan chung 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 tới 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 hơn nữa có thể sử dụng tới cỡ 60-70 số hạng. Những dự báo với độ đảm bảo cao đạt được khi sử dụng trên 20 hoặc trên 60 số hạng là do với số lượng các số hạng lớn như vậy chúng ta có thể bao quát hết các thông tin về biến trình năm, đồng thời tính đến cả những chu trình dao động nhiều năm rõ nhất của đại đa số các yếu tố khí tượng hải văn vùng biển là chu trình 2-3 năm và 5-7 năm (hình 1).



Hình 1. Phổ nhiệt độ nước các trạm Hòn Dầu (a) và Phú Quý (b). Nhận thấy rằng các yếu tố này dao động với chu kỳ dài nhất khoảng gần 7 năm



Hình 2. So sánh giá trị nhiệt độ nước quan trắc ở trạm Hòn Dầu (đường liền nét) với dự báo (đường gạch nổi)

Khảo sát cũng thấy rằng nếu sử dụng số lượng các số hạng phương trình dự báo nhỏ cỡ vài ba số như trong các phương pháp dự báo thống kê đơn giản thì khó có thể đạt kết quả cao. Đặc biệt trường hợp dự báo quán tính thuần túy (khi $\theta = 1$) thì độ đảm bảo của dự báo thấp hơn nhiều, mặc dù các quá trình nhiệt ở biển nói riêng và các quá trình khí tượng hải văn nói chung, như chúng ta đã biết, có quán tính khá lớn.

Hình 2 so sánh các giá trị thực và dự báo độc lập thời hạn một tháng trên thí dụ với nhiệt độ nước Hòn Dấu. Đa số các dự báo đều có sai số nhỏ hơn nhiều so với sai số cho phép. Những sai số lớn chỉ xảy ra với những tháng mà yếu tố đạt cực đại hoặc cực tiểu ở những năm dị thường lớn. Đối với tất cả các chuỗi khác ở bảng 1 kết quả dự báo hoàn toàn tương tự: hệ số tương quan giữa giá trị thực và dự báo đạt trên 0,95 và độ đảm bảo dự báo độc lập đạt trên 90 %.

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. Điều này là do về thực chất mô hình thống kê này cho phép khôi phục đúng xu thế của quá trình không những trong tương lai kế cận thời điểm lập dự báo mà cả một thời đoạn khá dài nếu trong khoảng thời gian đó cấu trúc dao động của yếu tố dự báo tương đối ổn định. Trong chương trình tính và dự báo thực hành, ngoài những chức năng đã nêu ở trên, chúng tôi đã chú ý xây dựng thủ tục tính toán sao cho những số liệu quan trắc mới nhất được tham gia vào quá trình tính hàm tương quan và thiết lập hệ phương trình chuẩn tắc xác định trị số của hàm ngoại suy tối ưu k_m và như vậy nếu hàm tương quan có sự biến động nào đó ở thời đoạn dự báo thì nó tự động được tính đến làm cho dự báo hiệu quả hơn.

Thay cho kết luận, chúng tôi nhận xét rằng phương pháp ngoại suy thống kê tuyến tính hoàn toàn có thể sử dụng hiệu quả để dự báo nhiều đặc trưng khí tượng hải văn. Điều quan trọng nữa là nó dễ sử dụng, không đòi hỏi nhiều dữ liệu ban đầu. Ngoài ra, nó cũng có ích trong các công tác thực tiễn xử lý số liệu khí tượng thủy văn như khôi phục những quan trắc khuyết trong quá khứ, tạo lập những biến trình năm của các yếu tố nhằm những mục đích khác nhau.

TÀI LIỆU THAM KHẢO

1. Alekhin Iu.M. *Dự báo thống kê trong địa vật lý*. LGU, L, 1963 (tiếng Nga).