

KHẢ NĂNG NÂNG CAO MỨC BẢO ĐẢM CỦA PHƯƠNG PHÁP TƯƠNG TỰ TRONG DỰ BÁO MỨC NƯỚC LŨ CAO NHẤT NĂM

PTS. BUI VĂN ĐỨC

KS. LÊ VĂN THẠCH

Cục Dự báo KTTV

Việc nâng cao mức bảo đảm của dự báo hạn dài mực nước lũ cao nhất năm ở nước ta đang có những đòi hỏi cấp bách, nhằm đáp ứng nhu cầu sản xuất nông nghiệp và nâng cao hiệu quả kinh tế. Các công trình thủy lợi - thủy điện đã và đang mọc lên ngày một nhiều trên khắp đất nước. Điều này dẫn tới sự cần thiết sử dụng các phương pháp mới và hoàn thiện các phương pháp cũ đang dùng trong dự báo nghiệp vụ.

Quá trình hình thành đỉnh lũ năm của các sông trên lãnh thổ nước ta rất phức tạp. Hiện nay các mô hình thủy lực còn đang bất lực trong việc mô tả quá trình vật lý phức tạp này. Vì vậy, trong dự báo hạn dài mực nước lũ cao nhất, năm người ta đang sử dụng rộng rãi các mô hình vật lý thống kê [2, 4].

Trong nhiều năm qua do điều kiện tính toán không cho phép trong dự báo mực nước lũ lớn nhất năm, phương pháp tương tự mới được sử dụng ở mức đơn giản [1]. Mức tương tự được đánh giá bằng hàm khoảng cách theo dấu (mang nhiều định tính hơn định lượng).

Trong khuôn khổ bài này chúng tôi đề cập khả năng nâng cao mức bảo đảm của phương pháp tương tự trong dự báo mực nước lũ cao nhất năm (H_{max}). Mức tương tự được xác định bằng khoảng cách Ocôlit mở rộng [3] cùng với sự chuẩn hóa tập nhân tố trước khi đưa vào xác định mức tương tự.

Sơ đồ khối và chương trình tính bằng ngôn ngữ basic cho máy Comodore-64 được minh họa bằng ví dụ dự báo mực nước lũ cao nhất năm trên sông Cả tại trạm Nam Đàn.

I- CƠ SỞ LÝ LUẬN CỦA PHƯƠNG PHÁP TƯƠNG TỰ

1. Tóm tắt thuật toán của phương pháp tương tự

Giả sử ta đã chọn được tập nhân tố $x_1, x_2, \dots, x_m$ tối ưu để dự báo yếu tố y . Qua nhiều năm quan trắc ta có ma trận trị số nhân tố và yếu tố sau:

$$\begin{pmatrix} x_{11} & x_{12} & \dots & x_{1m} \\ x_{21} & x_{22} & \dots & x_{2m} \\ \dots & \dots & \dots & \dots \\ x_{n1} & x_{n2} & \dots & x_{nm} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix} \quad (1)$$

Ở đây m - số nhân tố; n - độ dài dãy quan trắc (số tập mẫu).

Trong nhiều bài toán dự báo thủy văn hạn dài, yếu tố (Y) có thể không chỉ là một trị số, mà còn là một vectơ đặc trưng cho dòng chảy mùa lũ (mức nước lũ lớn nhất, năm H_{max} , mức nước trung bình năm H_b ; thời gian xuất hiện mức nước lớn nhất H_{max} ;). Vectơ nhân tố $X (x_1, x_2, \dots, x_m)$ - các đặc trưng khí tượng trong mùa đông xuân, được tuyển chọn theo các chỉ tiêu thống kê [1] qua số liệu quan trắc nhiều năm.

Thuật toán của phương pháp tương tự có thể tóm tắt như sau:

Ta coi m nhân tố dự báo là những phần tử của vectơ trong không gian m chiều. Mức tương tự được biểu diễn qua khoảng cách giữa các điểm, xác định bằng các vectơ nhân tố tương ứng $X_i (x_{i1}, x_{i2}, \dots, x_{im})$.

Người ta thường dùng những khoảng cách sau trong việc xác định mức tương tự [3].

a) Hàm khoảng cách theo dấu

$$D(X_0, X_i) = \frac{1}{m} \sum_{j=1}^m G_j \quad (2)$$

$$\text{với } G_j = \begin{cases} 1 & \text{khi } (x_{ij} - \bar{x}_j)(x_{0j} - \bar{x}_j) \geq 0 \\ 0 & \text{khi } (x_{ij} - \bar{x}_j)(x_{0j} - \bar{x}_j) < 0 \end{cases} \quad (3)$$

b) Hàm khoảng cách Ocolit mở rộng

$$D(X_0, X_i) = \left\| \frac{1}{m} \sum_{j=1}^m K_j^2 (x_{ij} - x_{0j})^2 \right\| \frac{1}{2} \quad (4)$$

c) Hàm khoảng cách Hemming

$$D(X_0, X_i) = \frac{1}{m} \sum_{j=1}^m K_j |x_{ij} - x_{0j}| \quad (5)$$

Ở đây: X_0 - vectơ nhân tố xuất phát dự báo, không chứa trong tập số liệu lưu trữ; x_{0j} và x_{ij} những phần tử (nhân tố) có chỉ số j của các vectơ nhân tố X_0 và X_i và K_j - hệ số tỉ trọng của nhân tố dự báo j .

Do điều kiện không đồng nhất về đơn vị đo và sự khác nhau về trị số tuyệt đối, người ta thường không dùng trực tiếp vectơ nhân tố $X (X_1, X_2, \dots, X_m)$, mà dùng nó ở dạng chuẩn hóa $Z (z_1, z_2, \dots, z_m)$.

$$Z_{ij} = \frac{x_{ij} - \bar{x}_j}{\sigma_j}$$

trong đó

$$\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij} \quad (7)$$

$$\sigma_j = \left\| \frac{1}{n} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2 \right\| \frac{1}{2} \quad (8)$$

Hệ số tỷ trọng K_j có nhiều phương pháp xác định [4]. Trong trường hợp đơn giản có thể dùng hệ số tương quan tuyến tính (R_{xy}) giữa nhân tố (X_j) với yếu tố dự báo (y) thay cho hệ số tỷ trọng K_j .

$$K_{j\infty} R_{xiy} = \frac{\sum_{i=1}^n (x_{ij} - \bar{x})(y_i - \bar{y})}{\left\{ \sum_{i=1}^n (x_{ij} - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2 \right\}^{\frac{1}{2}}} \cdot (y_i - \bar{y})^2 \quad (9)$$

Quy tắc quyết định dự báo theo phương pháp tương tự là:

$$Y'(X_0) = Y(X_i) \quad (10)$$

$$\text{nếu như } D(x_0, x_i) = \min D(X_0, X_e) \quad (11)$$

với $e=1, 2, \dots, n$.

Ở đây i là chỉ số hình thể, khi ở đó X_i so với bất kỳ X_e với $e \neq i$ có khoảng cách ngắn nhất tới X_0 .

2. Phân tích dự báo

Trong nhiều năm qua, hàm tương tự dấu (2) được sử dụng rộng rãi trong dự báo thủy văn hạn dài [1]. Phương pháp xác định mức tương tự theo hàm dấu (2) khá đơn giản, khối lượng tính toán không nhiều. Tuy vậy, hàm dấu (2) còn có những nhược điểm nhất định: mang nhiều định tính hơn định lượng, chưa chú ý tới tỷ trọng (K_j) của từng nhân tố trong hình thành yếu tố dự báo (y).

Để khắc phục những tồn tại nói trên và nâng cao mức bảo đảm của phương pháp tương tự, mức tương tự được xác định bằng khoảng cách Geolit mở rộng (4) với chuẩn hóa (6) và hệ số tỷ trọng K_j (9). Sơ đồ và kết quả tính toán được trình bày cụ thể ở phần sau, trên ví dụ số liệu của sông Cả tại trạm thủy văn Nam Đàn.

Bài toán dự báo thủy văn hạn dài có những đặc thù riêng, điều kiện hình thành dòng chảy quá phức tạp, tập nhân tố được tuyển chọn $X(x_1, x_2, \dots, x_m)$ không mô tả nổi một cách đầy đủ quá trình hình thành mực nước lũ cao nhất năm của các sông. Hơn nữa, ngay trong tập nhân tố $X(x_1, x_2, \dots, x_m)$ đã chứa những sai số nhất định ảnh hưởng trực tiếp tới chất lượng dự báo.

Chính vì những nguyên nhân trên mà khi quyết định trị số dự báo mực nước lũ lớn nhất năm ($H_{\max}$) phải dựa trên sự phân tích khoảng cách $D(X_0, X_i)$ và $H_{\max}$ tương ứng của một nhóm năm lân cận gần nhất với nhân tố xuất phát dự báo X_0 .

+ Trường hợp thứ nhất — khoảng cách $D(X_0, X_i)$ của năm gần nhất (i) nhỏ hơn nhiều lần so với những khoảng cách khác $D(X_0, X_e)$ với $e \neq i$, thì trị số dự báo $H_{\max}$ được quyết định theo $H_{\max}(X_i)$.

+ Trường hợp thứ hai — khoảng cách $D(X_0, X_i)$ của năm gần nhất (i) không nhỏ hơn nhiều so với một số khoảng cách $D(X_0, X_e)$, tạo thành nhóm năm tương tự ngang mức. Trường hợp này lại được phân tiếp thành hai trường hợp nhỏ.

— Nếu trị số $H_{\max}$ của các năm trong nhóm tương tự gần nhất không khác nhau nhiều (năm ở cùng trong một pha: dưới trung bình; xấp xỉ trung bình và trên trung bình), thì giá trị dự báo $H_{\max}(X_0)$ được tính theo công thức (10)

$$H'_{\max}(X_0) = \frac{1}{m_0} H_{\max e}, \quad (10)$$

trong đó m_0 —số năm của nhóm tương tự gần nhất.

— Nếu trị số mực nước lớn nhất năm $H_{\max}$ của các năm trong nhóm tương tự gần nhất ngang mức lại nằm ở những pha khác nhau, thì cần tham khảo kết quả dự báo của một số phương pháp khác, hoặc $H_{\max}$ của năm « tương tự ngược » để loại bớt đi những giá trị kém tin cậy. Số giá trị $H_{\max}$ còn lại được sử dụng để xác định trị số $H'_{\max}(X_0)$ theo công thức (10).

II — DỰ BÁO MỰC NƯỚC LŨ CAO NHẤT NĂM TRÊN SÔNG CẢ TẠI TRẠM THỦY VĂN NAM ĐÀN

1. Nhân tố dự báo

Theo số liệu thống kê nhiều năm dễ dàng nhận thấy mực nước lớn nhất trên các sông ở miền Bắc Việt Nam trong mùa lũ có quan hệ với các trạng thái thời tiết của mùa đông và mùa xuân trên lưu vực. Theo kinh nghiệm [1] các nhân tố dự báo dài hạn mực nước lũ cao nhất năm có thể như sau:

- + x_1 — Tổng lượng mưa trong các tháng XII, I, II,
- + x_2 — Số ngày mưa trong các tháng XII, I, II,
- + x_3 — Độ ẩm trung bình tương đối của không khí trong các tháng XII, I, II,
- + x_4 — Lượng bốc hơi từ mặt đất trong các tháng XII, I, II,
- + x_5 — Áp suất trung bình của không khí trong các tháng XII, I, II,
- + x_6 — Áp suất lớn nhất của không khí trong các tháng XII, I, II,
- + x_7 — Áp suất thấp nhất của không khí trong các tháng XII, I, II,
- + x_8 — Nhiệt độ trung bình của không khí trong các tháng XII, I, II,
- + x_9 — Nhiệt độ cao nhất của không khí trong các tháng XII, I, II,
- + x_{10} — Nhiệt độ thấp nhất của không khí trong các tháng XII, I, II.

Trên sông Cả tại trạm thủy văn Nam Đàn, mực nước được quan trắc liên tục từ năm 1927, nhưng chuỗi số liệu khí tượng quan trắc tại trạm khí tượng Vinh lại không đồng bộ. Vì vậy, dãy số liệu dùng làm tập mẫu chỉ còn lại 31 năm kể từ năm 1957. Các đặc trưng khí tượng $X(x_1, x_2, \dots, x_{10})$ được chọn trong các tháng mùa đông (XII, I, II) vì thời tiết trong các tháng này có độ ổn định cao và mang nhiều dấu hiệu biểu hiện các hoạt động hoàn lưu trong mùa hè kế tiếp [1]

2. Sơ đồ dự báo

Giả sử ta có véctor nhân tố xuất phát dự báo $X_0(x_{01}, x_{02}, \dots, x_{0,10})$. Gộp véctor nhân tố xuất phát dự báo với tập véctor nhân tố lưu trữ $[X]_{n,10}$ ta tìm các đặc trưng (thống kê $\bar{x}_j$; σ_j và hệ số tỷ trọng K_j theo các công thức (7), (8) và (9). Sau đó tiến hành chuẩn hóa chuỗi nhân tố $[X]_{n+1,10}$ theo công thức (8). Bước tiếp theo xác định khoảng cách $D(X_0, X_e)$ giữa các véctor nhân tố dự báo X_0 với véctor nhân tố xuất phát dự báo X_0 . Dãy khoảng cách $D(X_0, X_e)$ được sắp xếp lại theo thứ tự lớn dần với mục đích phân cấp mức tương tự. Trị số mực nước lũ cao nhất $H'_{\max}(X_0)$ được quyết định dựa trên cơ sở

phân tích dãy $D(X_0, X_e)$ và $H_{max}(X_e)$ tương ứng. Sơ đồ dự báo mực nước cao nhất năm được trình bày ở hình 1.

3. Kết quả tính toán và đề xuất hoàn thiện

Từ sơ đồ khối (hình 1) quá trình tính toán được cụ thể hóa trên ngôn ngữ Basic của máy tính Comodode-64.

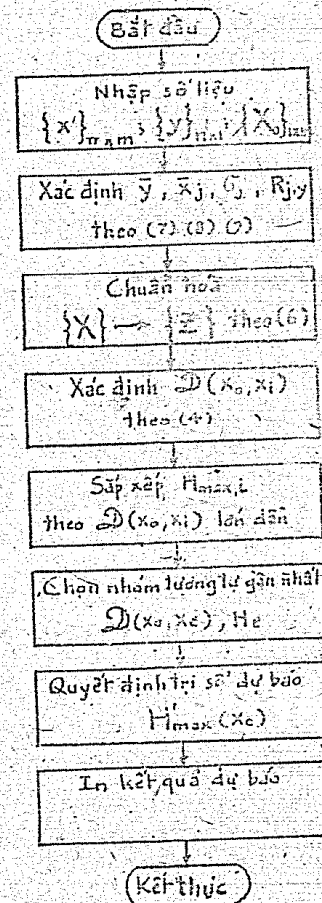
Phương pháp được đánh giá trên chuỗi số liệu dòng chảy sông Cả tại trạm thủy văn Nam Đàn và chuỗi số liệu khí tượng tại trạm khí tượng Vinh. Dãy số liệu độc lập được tạo nên bằng cách trượt dần [4]. Kết quả tính cho dự báo thử như sau:

— Dự báo trị số mực nước lũ cao nhất năm H_{max} với mức đảm bảo là 68% (cao hơn phương án cũ [1] là 4%).

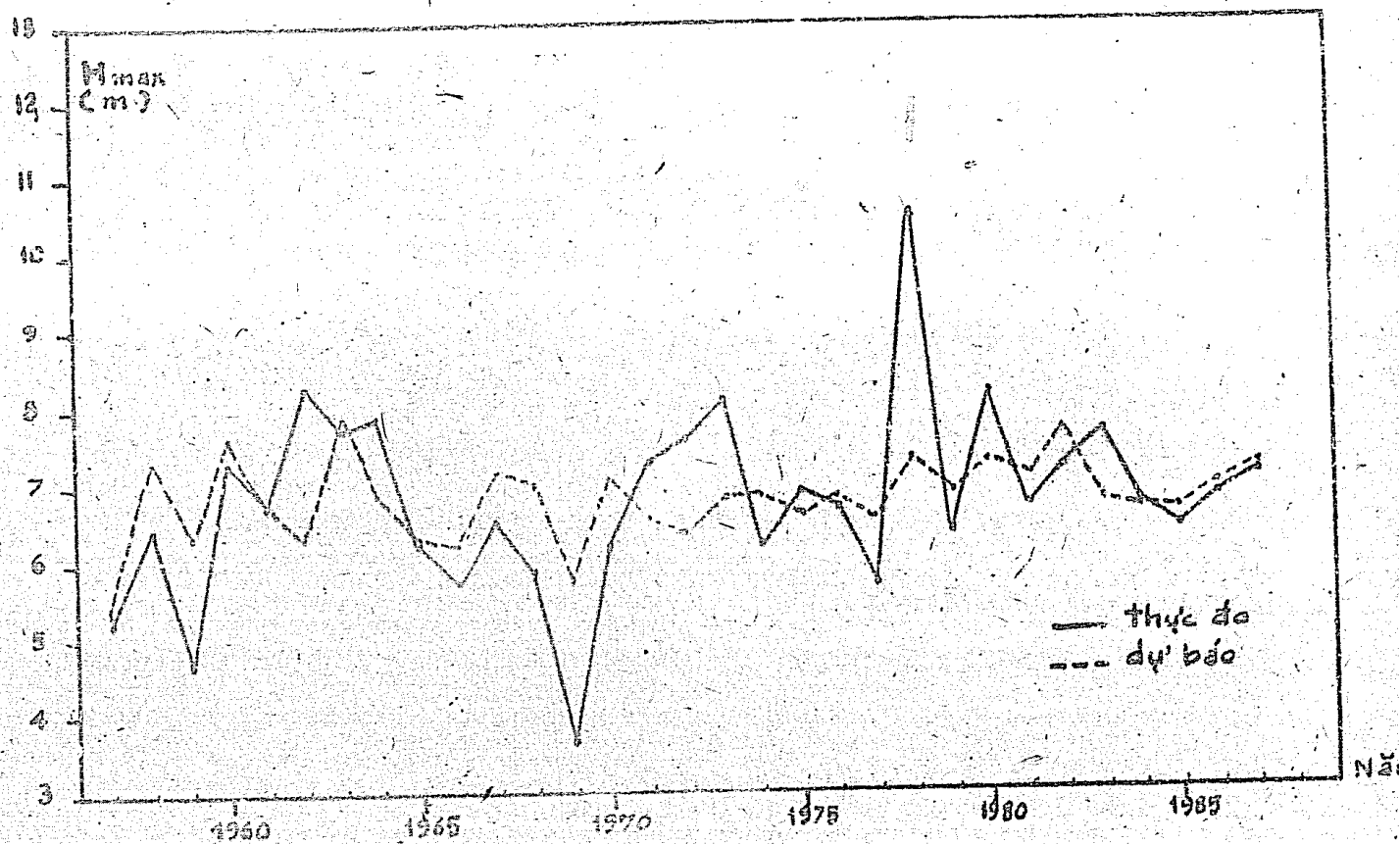
— Dự báo thời gian xuất hiện mực nước lũ cao nhất năm ($t_{H_{max}}$) có mức bảo đảm là 64% (dự báo $t_{H_{max}}$ đúng khi thời gian dự báo ($t_{H_{max}}$) và thời gian xảy ra H_{max} ($t_{H_{max}}$) cũng nằm trong một tháng).

Từ cơ sở lý luận và thực tế tính toán trên ví dụ dự báo mực nước lũ cao nhất năm tại trạm Nam Đàn (sông Cả) cho thấy sơ đồ dự báo hình 1 có độ ổn định tốt và mức bảo đảm cao hơn phương án dự báo trong [1]. Tuy nhiên, đối với những năm lũ đặc biệt lớn (1968) hoặc đặc biệt thấp (1959, 1969) dự báo mực nước lũ cao nhất năm chỉ đúng về xu thế (hình 2).

Để nâng cao hơn nữa mức bảo đảm của phương pháp tương tự, cần nghiên cứu thêm việc tuyển chọn tập nhân tố dự báo $X(x_1, x_2, \dots, x_m)$. Trong điều kiện cho phép, cần thiết đưa thêm vào các nhân tố đặc trưng cho hoàn lưu khí quyển.



Hình 1. Sơ đồ dự báo mực nước cao nhất năm (H_{max})



Hình 2. Biến trình mực nước cao nhất năm trên sông Cả tại Nam E

TÀI LIỆU THAM KHẢO

1. Dự báo mực nước lũ cao nhất năm trên các sông miền Trung bằng hàm tương tự dân. Phương án dự báo hạn dài của Phòng Dự báo thủy văn, Cục Dự báo KTTV, 1984.
2. Nguyễn Lan Châu. Đề tài dự báo hạn dài đỉnh lũ trên lưu vực sông Hồng. Tổng cục KTTV, 1985.
3. Trịnh Văn Thư, Phạm Ngọc Hiến. Dự báo sự di chuyển của bão bằng phương pháp tương tự. Phân tích dự báo bão, tập 1, Tổng cục KTTV, 1985.
4. Bùi Văn Đứ. Sự hình thành và dự báo dòng chảy tháng mùa hè đầu của các sông vùng Primorski — Liên Xô. Luận án phó tiến sĩ khoa học địa lý. Qactra, 1984 (tiếng Nga).