

DỰ BÁO DÒNG CHẢY LỚN NHẤT NĂM BẰNG PHƯƠNG PHÁP HÀM PHÂN LỚP

PTS. Nguyễn Việt Thi
Cục Dự báo KTTV

I. MỞ ĐẦU

Nhu cầu về dự báo dòng chảy lớn nhất năm với thời gian dự kiến từ 4 đến 6 tháng ngày càng tăng trong nhiều ngành kinh tế quốc dân, nhất là trong công tác phòng chống lũ lụt bảo vệ tính mạng và tài sản của nhân dân. Đây là một vấn đề rất phức tạp và còn ít được nghiên cứu. Trong những năm gần đây, do tính cấp bách của vấn đề, chúng tôi có tiến hành nghiên cứu thử nghiệm một số phương pháp dự báo khí tượng - thủy văn hạn dài để giải quyết vấn đề này. Trong bài này sẽ trình bày về kết quả áp dụng phương pháp hàm phân lớp để xây dựng các phương án dự báo dòng chảy lớn nhất năm của sông Hồng tại Hà Nội.

2. CƠ SỞ CỦA PHƯƠNG PHÁP HÀM PHÂN LỚP

Mô hình toán học của phương pháp hàm phân lớp được trình bày khá đầy đủ trong nhiều tác phẩm và công trình nghiên cứu của các nhà khoa học thế giới. Ở nước ta, trong lĩnh vực khí tượng thủy văn có thể tham khảo trong các công trình [1]. Do vậy, ở đây chỉ trình bày phần nội dung cơ bản, cần thiết để xây dựng phương án dự báo.

Về mặt vật lý có thể giả thiết rằng dòng chảy lớn nhất năm được đặc trưng bởi một số vô hạn các dấu hiệu. Nhưng trên thực tế, do khả năng hiểu biết của con người và do hạn chế về thông tin, ta chỉ có thể tính được hoặc xét đến một số hữu hạn các dấu hiệu:

$$X = (X_1, X_2, \dots, X_n) \quad (1)$$

Nếu ký hiệu $R^n = \{X = (X_1, X_2, \dots, X_n)\}$ là các giá trị có thể của X , khi chia không gian R^n làm 2 lớp rời nhau $R_1^n (Q_1)$ là tập hợp các hình thế X mà đã quan sát thấy cấp dòng chảy Q_1 và R_2^n - tập đã quan sát thấy cấp dòng chảy Q_2 . Dưới đây toán học ta có:

$$\begin{cases} X \in R_1^n (Q_1) \Leftrightarrow Q_1 \text{ xảy ra} \\ X \in R_2^n (Q_2) \Leftrightarrow Q_2 \text{ xảy ra} \end{cases} \quad (2)$$

Giả sử X là vectơ ngẫu nhiên và có phân phối chuẩn n chiều dòng thời phân bố bên trong của từng lớp cũng là chuẩn với mật độ phân bố $f_1(X)$ và $f_2(X)$.

Bài toán dự báo đặt ra là nếu có quan sát X về x thì phải xác định xem x là quan sát của vectơ ngẫu nhiên X có phân bố $f_1(X)$ hay $f_2(X)$, tức là ứng với việc xảy ra cấp dòng chảy Q_1 hay Q_2 .

Để xây dựng được quy tắc quyết định đối với phương án dự báo, vấn đề đặt

ra là phải phân chia R_1^n và R_2^n sao cho tối ưu. Từ tài liệu đã tích lũy được có thể biết được tần số xuất hiện biến cố Q_1 hay Q_2 và ký hiệu là q_1 hay q_2 tương ứng. Thiệt hại do dự báo sai là $C(Q_1/Q_2)$ ứng với trường hợp dự báo Q_2 nhưng thực do là Q_1 và $C(Q_2/Q_1)$ ứng với trường hợp ngược lại. Khi đó hàm phân lớp tối ưu được xác định như sau:

$$g(X) = \frac{f_1(X)q_1C(Q_2/Q_1)}{f_2(X)q_2C(Q_1/Q_2)} \quad (3)$$

từ đây ta có 2 lớp:

$$\begin{cases} R_1^n \equiv [X: g(X) \geq 1] \\ R_2^n \equiv [X: g(X) < 1] \end{cases} \quad (4)$$

Với giả thiết ma trận hiệp biến của các lớp là như nhau và các lớp được chia sao cho tần suất xuất hiện trong các lớp là bằng nhau, đồng thời coi thiệt hại do dự báo sai cũng như nhau. Khi đó ta thu được biểu thức giải tích của hàm phân lớp bậc 1:

$$U(X) = X'V^{-1} [M(Q_1) - M(Q_2)] - \frac{1}{2}[M(Q_1) + M(Q_2)]' +$$

$$+ \frac{1}{2}X'V^{-1}[M(Q_1) - M(Q_2)]X = \sum a_i x_i + g(X) \quad (5)$$

Nếu $U(X) \geq 0 \Leftrightarrow Q_1$ xảy ra

Nếu $U(X) < 0 \Leftrightarrow Q_2$ xảy ra

3. CHỌN LỌC CÁC NHÂN TỐ DỰ BÁO

Cũng như hầu hết các phương pháp vật lý thống kê kết quả của phương pháp hàm phân lớp phụ thuộc rất nhiều vào các số liệu lưu trữ và sự lựa chọn đúng đắn các nhân tố dự báo. Đối với dự báo dòng chảy lớn nhất năm hạn dài, các nhân tố dự báo tác động rất phức tạp, cho đến nay vẫn chưa có thể xác định được một hệ nhân tố dự báo hoàn chỉnh. Do vậy, đối với từng bài toán cụ thể thường có những biện pháp xử lý thông tin riêng.

Với bài toán hàm phân lớp từng bước để lựa chọn các nhân tố có chứa thông tin về yếu tố dự báo, chúng tôi đã sử dụng kết hợp hai quá trình sau:

Quá trình 1: dựa trên cơ sở vật lý, phương trình cân bằng nước, các phương pháp luận synopsis phân tích cơ chế ảnh hưởng giữa các hiện tượng tự nhiên để sơ tuyển các nhân tố dự báo.

Quá trình 2: sử dụng các phương pháp và chỉ tiêu thống kê để tuyển chọn bộ nhân tố và xây dựng phương trình hàm phân lớp tối ưu. Quá trình này gồm 2 bước sau:

Bước 1: tuyển lại các nhân tố đã thu được trong quá trình 1. Chỉ tiêu chọn là hệ số tương quan giữa yếu tố với các nhân tố dự báo và giữa các nhân tố dự báo với nhau.

Bước 2: sử dụng phương pháp lọc từng bước, thực hiện thủ thuật đưa vào và loại ra các thành phần của phương trình hàm phân lớp thông qua các chỉ tiêu thống kê. Cụ thể ở đây chúng tôi đã sử dụng chỉ tiêu Fisher.

Dựa vào phương trình tính hàm phân lớp $U(X)$ (5) biến X , ký hiệu:

$$\lambda(X) = \frac{\det W(X)}{\det T(X)} \quad (7)$$

trong đó X là véc-tơ (X_1, X_2, \dots, X_p) , T - tổng các bình phương chung khi chưa phân lớp, W - tổng các bình phương bên trong từng lớp. Biến X được đưa vào phải là biến có $\lambda(X)$ là nhỏ nhất.

Từng bước ta đưa vào thêm biến mới K ta tính được các giá trị $\lambda(X, K)$ và lượng giảm giá trị này so với khi chưa thêm biến mới là:

$$\lambda(K, X) = \frac{\lambda(X, K)}{\lambda(X)} \quad (8)$$

và chỉ tiêu Fisher được tính như sau:

$$F = \frac{n-q-p}{q-1} \cdot \frac{1-\lambda(K, X)}{\lambda(X)} \quad (9)$$

với q - số biến có thể có

n - số quan sát

p - số biến đã có trong phương trình.

Biến được đưa vào phải có giá trị $\lambda(X, K)$ nhỏ nhất và nhỏ hơn $\lambda(X)$ hay tương đương với việc biến đó có giá trị F lớn nhất.

Sau khi đã chọn được biến thứ hai, tiếp tục đưa vào biến thứ 3, thứ 4 v.v. cho đến khi không tìm được biến có giá trị λ mới nhỏ hơn giá trị λ ở bước trước. Khi đó quá trình lọc sẽ được dừng và phương trình tính hàm phân lớp cuối cùng được coi là phương trình tối ưu.

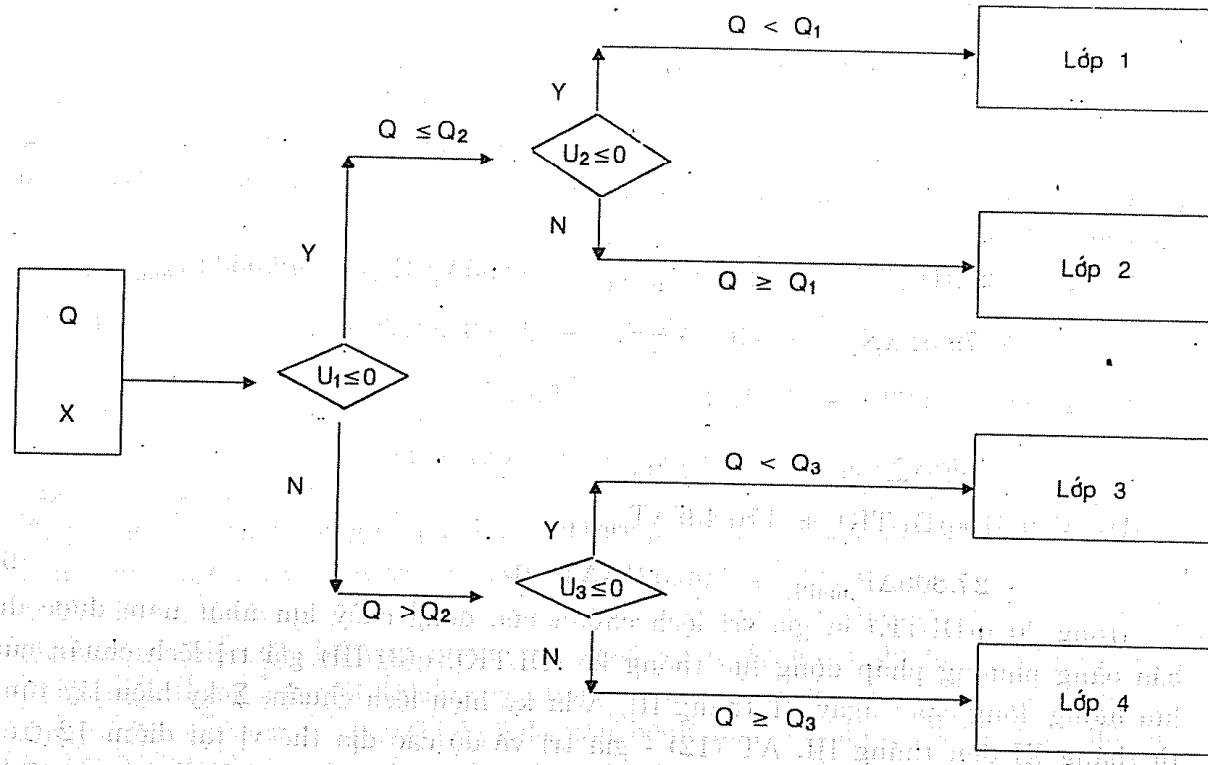
4. SƠ ĐỒ TÍNH

Chuỗi dòng chảy lớn nhất trong năm của sông Hồng tại trạm Hà Nội được phân ra làm 4 lớp đều nhau với các ngưỡng $Q_1 = Q_{75\%}$, $Q_2 = Q_{50\%}$ và $Q_3 = Q_{25\%}$. Cụ thể như sau:

- Lớp 1: $Q < Q_1$
 - Lớp 2: $Q_1 \leq Q < Q_2$
 - Lớp 3: $Q_2 \leq Q < Q_3$
 - Lớp 4: $Q \geq Q_3$
- $$(10)$$

Từ tập số liệu thống kê với các công thức từ (5) đến (9) và các ngưỡng phân lớp (10) quá trình dự báo được thực hiện như sau:

Tìm hàm phân lớp U_1 theo công thức (5) tối ưu với ý nghĩa chỉ tiêu Fisher đạt cực đại. Từ các giá trị của U_1 chuỗi yếu tố dự báo được phân làm 2 lớp với $Q < Q_2$ và $Q \geq Q_2$. Sau đó từ mỗi lớp tiến hành xác định hàm phân lớp U_2 để phân lớp thứ nhất với $Q < Q_2$ ra làm 2 lớp $Q < Q_1$ và $Q \geq Q_1$ và hàm phân lớp U_3 để phân lớp thứ hai với $Q \geq Q_2$ thành 2 lớp $Q < Q_3$ và $Q \geq Q_3$. Từ đó dựa vào các giá trị tính được của U_1 , U_2 hoặc U_3 ta tìm được cấp dự báo dòng chảy lớn nhất năm. Sơ đồ lôgic dự báo cấp dòng chảy lớn nhất năm được trình bày trên hình 1.



Hình 1 - Sơ đồ lôgic dự báo cấp dòng chảy lớn nhất năm

5. DỰ BÁO DÒNG CHÁY LỚN NHẤT NĂM

5.1. Yếu tố dự báo

Yếu tố dự báo là dòng chảy lớn nhất năm của sông Hồng tại Hà Nội với thời gian phát báo vào đầu tháng IV hàng năm. Từ chuỗi số liệu quan trắc 88 năm (từ năm 1902 đến năm 1989) đã xác định được các ngưỡng phân lớp:

$$Q_1 = 10000 \text{m}^3/\text{s}, Q_2 = 12000 \text{m}^3/\text{s}, Q_3 = 14000 \text{m}^3/\text{s}$$

5.2. Các nhân tố dự báo

Các nhân tố dự báo gồm các nhóm sau:

Nhóm 1: chuỗi dòng chảy lớn nhất năm được phân tích bằng phương pháp động lực thống kê với thời gian trễ từ 1 đến 20 năm, để tính thành phần dòng chảy tạo nên bởi các nguyên nhân bên trong của bản thân quá trình dòng chảy lớn nhất năm. Phương pháp động lực thống kê có thể tham khảo trong [2].

Nhóm 2: các chuỗi dòng chảy tháng III và các chuỗi dòng chảy 5 ngày và 1 ngày cuối tháng III, chúng đặc trưng cho tính quán tính và lượng trữ nước trong lưu vực sông.

Nhóm 3: các yếu tố khí hậu mùa đông (mưa, ẩm, nhiệt, áp) từ tháng XI đến tháng III của trạm Lai Châu và Phù Liễn. Nhóm nhân tố này đặc trưng cho ảnh hưởng của thành phần quán tính khí hậu đến dòng chảy lớn nhất năm.

Nhóm 4: các chuỗi độ cao địa thế vị trung bình mùa đông (từ tháng I đến tháng III) của 20 trạm trên mức 500hPa. Nhóm nhân tố này biểu thị tác động của các hệ

thống nhiễu động khí quyển ở trạng thái ban đầu và khả năng phát triển của chúng trong tương lai.

5.3. Các phương trình phân lớp

Từ những chuỗi số liệu về dòng chảy lớn nhất năm và các nhân tố dự báo quan trắc đồng bộ từ năm 1958 đến 1989 theo các công thức từ (5) đến (9) và sơ đồ lôgic dự báo cấp dòng chảy lớn nhất năm (hình 1) đã tìm được các phương trình hàm phân lớp tối ưu sau:

$$U_1 = 0,863q(DLT) + 1,131\Delta P_{maxLC} + 0,895\Delta \sum E_{LC} + 46,533\Delta T_{minLC} \\ + 79,51\Delta N_{PL} + 0,951\Delta \sum E_{PL} + 235,358\Delta AT \quad (12)$$

$$U_2 = 0,66q(DLT) + 7,91q(31/3) + 0,56\Delta N_{LC} \\ - 4,98\Delta \sum P_{PL} + 29,83\Delta N_{PL} - 1,44\Delta AT \quad (11)$$

$$U_3 = 0,915q(DLT) + 156,439\Delta T_{minLC} \\ - 27,806\Delta P_{minPL} + 790,521\Delta AT \quad (9)$$

trong đó $q(DLT)$ là giá trị lệch chuẩn của dòng chảy lớn nhất năm được dự báo bằng phương pháp động lực thống kê (DLT), $q(31/3)$ giá trị lệch chuẩn của lưu lượng dòng chảy ngày 31 tháng III, Δ là ký hiệu lệch chuẩn, Σ ký hiệu lấy tổng từ tháng XI đến tháng III, AT (12) - giá trị về độ cao địa thế vị tại điểm 12 trên mức 500hPa. Các yếu tố khí hậu có kèm theo ký hiệu LC hoặc PL được quan trắc tại trạm Lai Châu hoặc Phù Liễn.

5.4. Kết quả dự báo

Chất lượng dự báo của phương pháp được đánh giá qua dự báo phụ thuộc và dự báo phục vụ từ năm 1958 đến năm 1991. Kết quả dự báo phụ thuộc trong 32 năm từ năm 1958 đến năm 1989 đúng 28 năm và sai 4 năm, mức đảm bảo của phương pháp đạt 87%. Sơ đồ này đã được thử nghiệm trong dự báo tác nghiệp 3 năm 1990, 1991 và 1992. Kết quả dự báo 2 năm 1990 và 1991 là phù hợp với thực tế.

6. KẾT LUẬN

Cho đến nay trên thế giới ngay cả các nước phát triển việc dự báo dòng chảy lớn nhất năm từ mưa còn gặp rất nhiều khó khăn. Các mô hình dự báo chủ yếu là dựa trên cơ sở của các phương pháp phân tích chuỗi thời gian và phương pháp hồi qui nhiều biến. Chất lượng dự báo còn chưa cao. Việc đưa vào nghiên cứu ứng dụng phương pháp hàm phân lớp đã tăng thêm cơ sở khoa học và kỹ thuật cho công tác này. Kết quả dự báo trên cơ sở số liệu thực đo về dòng chảy lớn nhất năm của sông Hồng tại Hà Nội cho thấy khả năng có thể sử dụng phương pháp này như một trong những công cụ dự báo tác nghiệp hàng năm đồng thời có thể nghiên cứu áp dụng cho các sông lớn khác.

TÀI LIỆU THAM KHẢO

1. Nguyễn Việt Phong. Dự báo sự xâm nhập của front lạnh bằng phương pháp phân tích phân biệt. Luận án phó tiến sĩ. HN, 1983.
2. Nguyễn Việt Thi. Sử dụng phương pháp chuỗi thời gian để dự báo dòng chảy tháng. Hội thảo Quốc gia về ứng dụng mô hình toán thủy văn - thủy lực trong phát triển và quản lý tài nguyên nước. Hà Nội, 1988.