

SỬ DỤNG PHƯƠNG PHÁP PHÂN TÍCH TỔNG HỢP DỰ BÁO DÒNG CHẢY NĂM SÔNG HỒNG

PTS. Nguyễn Viết Thi

Cục Dự báo KTTV

1. MỞ ĐẦU

Vấn đề dự báo dòng chảy năm đã được nhiều quốc gia quan tâm. Có nhiều mô hình và phương pháp đã được nghiên cứu sử dụng, trong đó các phương pháp phân tích chuỗi thời gian và hồi qui tuyến tính được sử dụng nhiều hơn cả. Tuy nhiên, do thời gian dự kiến dài và do hạn chế về số liệu quan trắc, cho đến nay chất lượng dự báo của các phương pháp này còn chưa đáp ứng được yêu cầu ngày càng tăng của nhiều ngành kinh tế quốc dân. Đặc biệt đối với các vùng nhiệt đới gió mùa như ở Việt Nam, dòng chảy sông thường biến động rất phức tạp, việc dự báo chúng lại càng khó khăn hơn. Để tăng cường về kỹ thuật dự báo, khắc phục những hạn chế của các phương pháp đang sử dụng, chúng tôi đã thử nghiệm áp dụng phương pháp phân tích tổng hợp để xây dựng các phương án dự báo dòng chảy năm của sông Hồng và các sông nhánh Đà, Thao, Lô.

2. CƠ SỞ CỦA PHƯƠNG PHÁP PHÂN TÍCH TỔNG HỢP

Phương pháp phân tích tổng hợp có cơ sở trên quan điểm thông tin - dự báo xét quá trình tự nhiên dưới dạng:

$$A_t = \sum_{i=1}^m f_i(B_{it}) + \sum_{j=1}^n f_j(C_{jt}) + \sum_{k=1}^p r_k \quad (1)$$

Như vậy, quá trình tự nhiên A được tạo nên bởi 3 thành phần cơ bản: thành phần tạo nên bởi các nguyên nhân bên trong (B_{it}) đặc trưng cho qui luật biến đổi của quá trình A, thành phần này được tính đến qua tính chu kỳ và tính quán tính của quá trình, thành phần thứ hai tạo nên bởi các yếu tố tác động bên ngoài (C_{jt}) và thành phần thứ ba - thành phần ngẫu nhiên sinh ra trong quá trình hình thành và phát triển của quá trình A.

Với quan điểm thông tin - dự báo trên, để xây dựng mô hình dự báo, đã kết hợp hai xu hướng chính trong lĩnh vực dự báo hạn dài hiện nay là xu hướng căn nguyên và xu hướng thống kê toán lý xây dựng sơ đồ phân tích tổng hợp khí hậu - thủy văn - quán tính với tỷ trọng thực nghiệm và thời gian trễ khác nhau.

Mô hình dự báo dòng chảy năm bằng phương pháp phân tích tổng hợp có dạng sau:

$$Q_t = \bar{Q} + \sum_{i=1}^l \alpha_{it} q(t-i) + \sum_{j=1}^m \sum_{k=1}^n \beta_{ikt} X_j(t-k) + U_t \quad (2)$$

Trong biểu thức (2), Q_t - lưu lượng dòng chảy năm thứ t ; $\bar{Q}$ - giá trị trung bình của chuỗi Q_t ; $q(t-i)$ - các giá trị lệch trung bình của lưu lượng năm thứ $(t-i)$ với hệ số tự hồi qui α_{it} , tính đến đóng góp thông tin của $q(t-i)$ vào Q_t , l - số thành phần tự hồi qui tham gia vào phương trình; $X_j(t-k)$ - giá trị của nhân tố thứ j tại thời điểm $(t-i)$; m - số nhân tố ảnh hưởng; n - thời gian trễ mà nhân tố đó có ảnh hưởng đến Q_t , β_{ikt} - hệ số hồi qui, tương ứng với nhân tố $X_j(t-k)$; U_t - thành phần ngẫu nhiên, ở đây được xem như sai số tính toán.

3. THUẬT TOÁN TỔNG QUÁT CỦA PHƯƠNG PHÁP PHÂN TÍCH TỔNG HỢP

Các phương hướng cơ bản của việc xây dựng phương án dự báo bằng phương pháp phân tích tổng hợp là sử dụng các phương pháp vật lý thống kê chủ yếu đã được nghiên cứu và phát triển trong lĩnh vực dự báo khí tượng thủy văn hạn dài. Theo [1], thuật toán tổng quát của phương pháp này có thể mô tả như sau:

Giả sử Q là yếu tố phải dự báo và $X(n)$ là n nhân tố dự báo, có thể tiến hành đánh giá các thành phần của quá trình Q và tổ hợp nó như sau:

3.1 Thành phần tạo nên do các nguyên nhân bên trong

Từ giả thiết về sự tồn tại mối quan hệ bên trong của các quá trình vĩ mô tự nhiên, quá trình dòng chảy năm được xác định thông qua mối quan hệ bên trong của quá trình:

$$q(t) = \sum_{\tau=1}^l K(\tau) q(t-\tau) \quad (3)$$

Trong đó $q(t)$ - giá trị lệch chuẩn của lưu lượng năm thứ t được tính thông qua các giá trị lệch chuẩn của lưu lượng các năm thứ $(t-\tau)$; $K(\tau)$ - hệ số tự hồi qui, tính đến đóng góp của $q(t-\tau)$ vào $q(t)$; τ - thời gian trễ.

Mối quan hệ (3) được thể hiện qua hàm tự tương quan và dẫn đến giải hệ phương trình tuyến tính:

$$r = \sum_{\tau=1}^l K(\tau) r(i-\tau); \quad i = 1, 2, \dots \quad (4)$$

Ngoài những ký hiệu đã sử dụng ở các phương trình trên, $r(i)$ và $r(i-\tau)$ là giá trị của hàm tự tương quan tại thời điểm i và $(i-\tau)$.

Hệ phương trình (4) được giải bằng phương pháp Gauss với điều kiện tối thiểu phương sai dư của phương trình (3).

Chất lượng mô phỏng của phương trình (3) được đánh giá qua 3 chỉ tiêu:

- Chỉ tiêu tương quan, hệ số tương quan giữa tính toán và thực đo

$$R_{TD}(l) = \frac{\sum_{i=1+1}^n q_i \hat{q}_i}{(n-1)\sigma^2} \quad (5)$$

Với $R_{TD}(l)$ - hệ số tương quan giữa chuỗi thực đo q_i và chuỗi tính toán $\hat{q}_i$ tính theo phương trình (3) khi có l tham số; n - độ dài chuỗi số liệu ban đầu; $(n-1)$ - độ dài chuỗi tính toán; σ^2 và $\hat{\sigma}^2$ là phương sai của chuỗi q_i và $\hat{q}_i$.

- Chỉ tiêu ngẫu nhiên:

$$\delta(l) = \frac{\sigma^2 \Delta(l)}{\sigma^2} \quad (6)$$

Trong đó $\delta(l)$ - chỉ tiêu ngẫu nhiên ứng với trường hợp phương trình (3) có l tham số; $\sigma^2 \Delta(l)$ - phương sai dư và σ^2 - phương sai của chuỗi thực đo.

- Mức đảm bảo dự báo của phương trình:

$$P(l) = \frac{m}{(n-1)} 100\% \quad (7)$$

$P(l)$ - mức đảm bảo dự báo của phương trình (3); m - số lần dự báo có sai số nhỏ hơn sai cho phép; $(n-1)$ - tổng số lần dự báo; l - số tham số của phương trình (3).

Phương trình thỏa mãn các điều kiện sau:

$$\left. \begin{aligned} R_{TD}(l) &\rightarrow \max \\ \delta(l) &\rightarrow \min \\ P(l) &\rightarrow \max \end{aligned} \right\} \quad (8)$$

được chọn là phương trình tối ưu, có tổ hợp thông tin tốt nhất từ các nhân tố bên trong đóng góp vào sự phát triển của quá trình trong tương lai.

3.2 Thành phần tạo nên do các nhân tố bên ngoài và các nhân tố đặc trưng cho trạng thái ban đầu của lưu vực được đánh giá bằng phương pháp hồi qui nhiều chiều lọc từng bước.

Mô hình hồi qui tuyến tính nhiều biến số có dạng:

$$Y_i = \beta_0 + \sum_{j=1}^m \beta_j X_{ji} + \mu_i \quad (9)$$

trong đó Y_i - giá trị của biến phụ thuộc, X_{ji} - các giá trị thứ i của biến độc lập thứ j ; β_j - các hệ số hồi qui và μ - sai số tính toán.

Các tham số của (q) được ước lượng bằng phương pháp bình phương nhỏ nhất.

Ý nghĩa của phương trình hồi qui được xác định qua các chỉ tiêu thống kê Fisher (F) và Student (t).

Chỉ tiêu thống kê F kiểm định giả thiết tồn tại của phương trình hồi qui, được tính bằng công thức sau:

$$F = \frac{n-p-1}{p} \times \frac{R^2}{1-R^2} \quad (10)$$

Với n là độ dài chuỗi tính toán; p - số biến của phương trình (9); R^2 - hệ số xác định của phương trình hồi qui.

Chỉ tiêu Student t_i , kiểm định giả thiết về tồn tại của từng hệ số hồi qui:

$$t_i = \frac{\hat{\beta}_i}{S_{\beta_i}} \quad (11)$$

Trong đó $\hat{\beta}_i$ là hệ số hồi qui của biến thứ i , được ước lượng bằng phương pháp bình phương nhỏ nhất; S_{β_i} - là sai số chuẩn của hệ số hồi qui β_i .

Qui tắc chọn các nhân tố dự báo được xây dựng trên cơ sở kết hợp hai phương pháp tuyến chọn thông tin sau:

a) Trên cơ sở vật lý, các phương pháp luận synóp, thủy văn phân tích mối quan hệ và ảnh hưởng của các hiện tượng khí tượng thủy văn để sơ tuyển các nhân tố ban đầu.

b) Sử dụng các phương pháp, chỉ tiêu thống kê toán lý hoặc các phép biến đổi thông tin để tuyển chọn tổ hợp các nhân tố có chứa lượng thông tin có ý nghĩa dự báo lớn nhất. Quá trình này được thực hiện qua hai bước thứ tự như sau:

Bước 1, dùng chỉ tiêu tương quan, đánh giá mối quan hệ của các nhân tố với yếu tố dự báo. Các nhân tố được chọn khi có hệ số tương quan lớn hơn 0,20, chúng tạo nên bộ nhân tố cơ sở.

Bước 2, kết hợp phân tích phương sai với phân tích dự báo trong quá trình lọc từng bước để xây dựng các phương án dự báo.

Nội dung của phép phân tích hồi qui từng bước là thực hiện các quá trình đưa vào và đưa ra các nhân tố dự báo để chọn nhân tố và tổ hợp tuyến tính của chúng có ý nghĩa thông tin dự báo tốt nhất.

Việc đưa vào và đưa ra các nhân tố dự báo được thực hiện qua chỉ tiêu giảm phương sai dư khi biến đó được đưa vào hay đưa ra khỏi phương trình:

$$C_{j+1,i} = \frac{R_{j+1,i}^2 - R_j^2}{1 - R_j^2} \quad (10)$$

Trong đó $C_{j+1,i}$ là chỉ số giảm phương sai dư khi đưa thêm biến thứ i vào phương trình hồi qui, i là chỉ số của các biến chưa được đưa vào phương trình ở bước trước (bước thứ j); R_j^2 và $R_{j+1,i}^2$ là hệ số xác định khi phương trình có j tham số và khi được đưa vào thêm một biến thứ i .

Chỉ tiêu dự báo được đánh giá qua hiệu quả dự báo khi có thêm hoặc bớt biến mới:

$$\Delta P_{j+1} = P_{j+1} - P_j \quad (11)$$

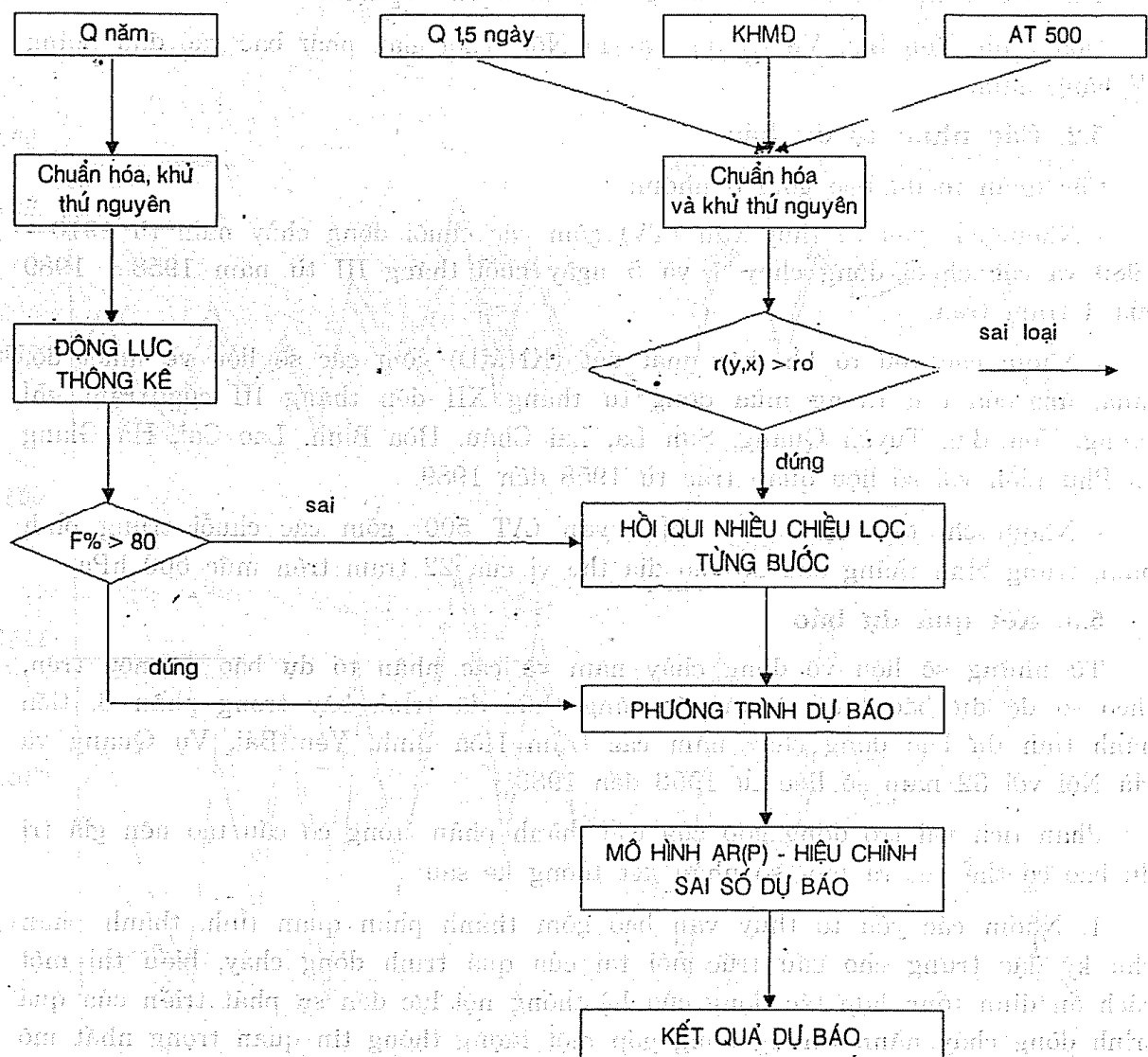
với ΔP_{j+1} là hiệu quả dự báo của phương trình (9) khi có $j+1$ tham số; P_{j+1} mức đảm bảo dự báo của phương trình khi có $j+1$ biến và P_j mức đảm bảo dự báo của phương trình khi có j biến.

Biến được đưa vào phương trình ở bước thứ $j+1$ là biến có $C_{j+1,i}$ lớn nhất và có ΔP_{j+1} dương. Sau khi đưa thêm biến mới vào phương trình ta lại thực

hiện phép thử đưa dần các biến đã có trong phương trình ở các bước trước vào khỏi mô hình. Nếu khi đưa ra một biến nào đó mà chỉ tiêu phương sai dự không tăng và mức đảm bảo dự báo không giảm thì chứng tỏ biến đó cần được đưa ra. Nếu ngược lại, biến thử đưa ra đó cần được giữ lại. Quá trình đưa vào, đưa ra các biến cứ lặp đi lặp lại cho đến khi không còn biến nào thỏa mãn các chỉ tiêu thống kê đã nêu.

4. SƠ ĐỒ DỰ BÁO

Với những thuật toán mô tả trên, sơ đồ dự báo bằng phương pháp phân tích tổng hợp được trình bày trong hình 1.



Hình 1. Sơ đồ dự báo

Giả sử ta có chuỗi số liệu về yếu tố dự báo Q, sử dụng phương pháp động lực thống kê (DLTK) tính theo các công thức (3), (4) và các chỉ tiêu (5), (6), (7) và (8) để chọn tổ hợp thông tin tốt nhất từ các nhân tố bên trong. Từ các chuỗi số liệu về các nhân tố dự báo, tiến hành đánh giá tương quan để chọn các nhân tố tác động bên ngoài. Liên kết các nhân tố chọn được bằng phương pháp tương quan với các thành phần của tổ hợp thông tin đã chọn bằng phương pháp động lực thống kê lập nên bộ nhân tố cơ sở. Từ đây dựa vào phương pháp hồi qui từng bước (9) và các chỉ tiêu thống kê (10) và (11) xây dựng các phương trình dự báo tối ưu. Sai số dự báo được hiệu chỉnh bằng mô hình AR (P) trên cơ sở phân tích chuỗi sai số dự báo phụ thuộc.

5. DỰ BÁO DÒNG CHẢY NĂM TRÊN LƯU VỰC SÔNG HỒNG

5.1. Các yếu tố dự báo

Các yếu tố dự báo bao gồm dòng chảy năm của 4 trạm:

Hòa Bình, Yên Bái, Vụ Quang và Hà Nội. Thời gian phát báo vào đầu tháng IV hàng năm.

5.2. Các nhân tố dự báo

Các nhân tố dự báo gồm 3 nhóm:

- Nhóm các yếu tố thủy văn (TV) gồm các chuỗi dòng chảy năm từ 1910 - 1989 và các chuỗi dòng chảy 1 và 5 ngày cuối tháng III từ năm 1958 - 1989 của 4 trạm trên.

- Nhóm các yếu tố khí hậu mặt đất (KHMD) gồm các số liệu về nhiệt độ, mưa, ẩm của các tháng mùa đông từ tháng XII đến tháng III của 8 trạm khí tượng: Yên Bái, Tuyên Quang, Sơn La, Lai Châu, Hòa Bình, Lào Cai, Hà Giang và Phù Lỗ với số liệu quan trắc từ 1958 đến 1989.

- Nhóm chỉ tiêu hoàn lưu khí quyển (AT 500) gồm các chuỗi trung bình năm, trung bình tháng của độ cao địa thế vị của 22 trạm trên mức 500 hPa.

5.3. Kết quả dự báo

Từ những số liệu về dòng chảy năm và các nhân tố dự báo đã nêu trên, theo sơ đồ dự báo hình 1 và các công thức đã trình bày trong phần 3, tiến hành tính dự báo dòng chảy năm các trạm Hòa Bình, Yên Bái, Vụ Quang và Hà Nội với 32 năm số liệu từ 1958 đến 1989.

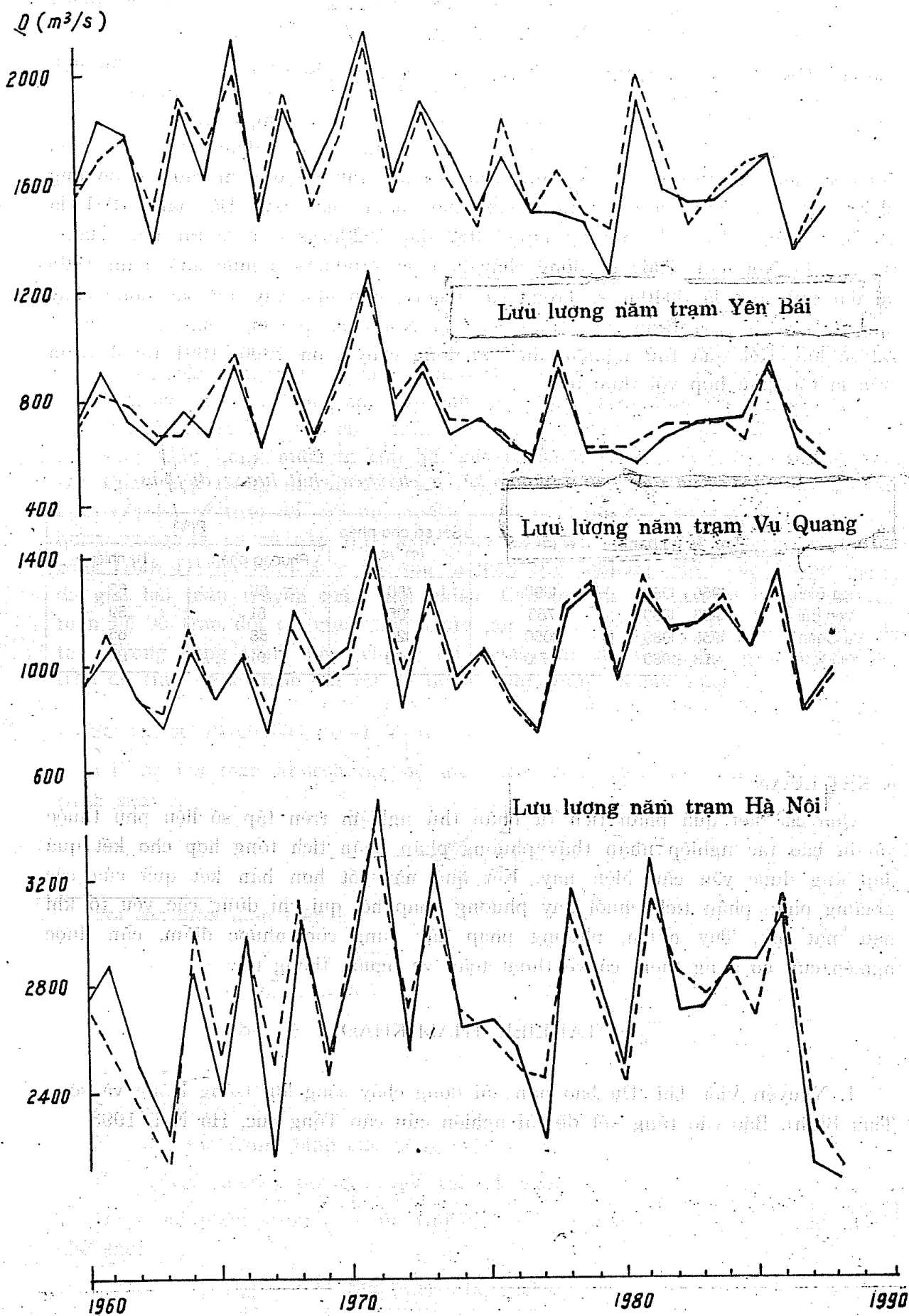
Phân tích vai trò đóng góp của các thành phần trong cơ cấu tạo nên giá trị dự báo có thể rút ra một số nhận xét thống kê sau:

1. Nhóm các yếu tố thủy văn bao gồm thành phần quán tính, thành phần chu kỳ đặc trưng cho cấu trúc nội tại của quá trình dòng chảy, biểu thị một cách ổn định tổng hợp tác động của hệ thống nội lực đến sự phát triển của quá trình dòng chảy năm. Chúng đóng góp một lượng thông tin quan trọng nhất mô phỏng được từ 60 đến 70% phương sai.

2. Yếu tố khí hậu mặt đất biểu thị sự ảnh hưởng của chế độ khí hậu đối với biến đổi của quá trình dòng chảy năm.

3. Nhóm thứ 3 gồm các nhân tố trên mặt đẳng áp AT 500 hPa và các dẫn

Lưu lượng năm trạm Hòa Bình



Hình 2. Quá trình dòng chảy năm (1960 - 1988).

— thực đo - - - dự báo

xuất của chúng là những chỉ tiêu biến đổi theo ngoại lực của các hệ thống nhiều động khí quyền đến dòng chảy năm.

Các chỉ tiêu chất lượng của dự báo thống kê trong bảng 1 và sự phù hợp giữa đường quá trình dự báo và thực đo trong hình 2 đã chứng tỏ khả năng cao của phương pháp phân tích tổng hợp trong dự báo dòng chảy năm của các trạm chính trên sông Hồng. Tuy nhiên, đối với những năm dòng chảy năm biến đổi đặc biệt lớn, từ rất nhỏ ở năm trước và đạt rất lớn ở năm sau thì phương pháp này còn hạn chế. Ví dụ, dòng chảy năm của sông Đà năm 1981 là $1250\text{m}^3/\text{s}$ thuộc loại nhỏ nhưng năm 1982 đạt $1920\text{m}^3/\text{s}$ ở mức rất lớn. Tương tự, tại Hà Nội năm 1967 có dòng chảy năm là $2150\text{m}^3/\text{s}$ ở mức nhỏ, năm 1968 lại đạt mức cao là $3040\text{m}^3/\text{s}$. Trong các trường hợp như vậy, dự báo dòng chảy năm 1982 của Hòa Bình hay 1968 của Hà Nội bằng phương pháp này còn có sai số lớn. Kết quả thử nghiệm dự báo dòng chảy năm 1990, 1991 tại 4 trạm trên là tốt, phù hợp với thực tế.

Bảng 1. Các đặc trưng thống kê và chỉ tiêu chất lượng dự báo

Trạm	Thời kỳ dự báo	$\bar{Q}$ (m^3/s)	Sai số cho phép (m^3/s)	P(%)	
				Phương pháp	Tự nhiên
Hòa Bình	1958 - 1989	1690	150	88	62
Yên Bái	1958 - 1989	765	105	81	59
Vũ Quang	1958 - 1989	1050	120	88	66
Hà Nội	1958 - 1989	2710	250	91	66

6. KẾT LUẬN

Qua các kết quả phân tích từ phần thử nghiệm trên tập số liệu phụ thuộc và dự báo tác nghiệp nhận thấy phương pháp phân tích tổng hợp cho kết quả đáp ứng được yêu cầu hiện nay. Kết quả này tốt hơn hẳn kết quả của các phương pháp phân tích chuỗi hay phương pháp hồi qui chỉ dùng các yếu tố khí hậu mặt đất. Tuy nhiên, phương pháp này cũng còn nhược điểm, cần được nghiên cứu bổ sung thêm cả về thuật toán và nguồn thông tin.

TÀI LIỆU THAM KHẢO

1. Nguyễn Viết Thi. Dự báo hạn dài dòng chảy sông lớn (sông Hồng và sông Thái Bình). Báo cáo tổng kết đề tài nghiên cứu cấp Tổng cục, Hà Nội, 1993.