

PHƯƠNG PHÁP PHÂN TÍCH TỔNG HỢP TRONG DỰ BÁO THỦY VĂN HẠN DÀI

PTS. Nguyễn Viết Thi

Trung tâm quốc gia Dự báo KTTV

Dự báo hạn dài dòng chảy sông với thời gian dự kiến từ 1 tháng trở lên đang là một nhu cầu cấp bách đối với công tác quy hoạch, quản lý và khai thác có hiệu quả nguồn nước sông. Nhưng việc nghiên cứu xây dựng các phương pháp tính toán dự báo hạn dài vẫn còn là một khó khăn cả về lý thuyết lẫn cơ sở dữ liệu thử nghiệm. Cho đến nay đã có nhiều phương pháp và mô hình được đề xuất, song kết quả áp dụng chúng trong dự báo hạn dài dòng chảy sông lớn của Việt Nam vẫn chưa đạt yêu cầu. Để giải quyết vấn đề này, trên cơ sở lý thuyết về các quá trình vĩ mô tự nhiên chúng tôi đã xây dựng được phương pháp phân tích tổng hợp (PTTH) [6] và áp dụng thành công trong dự báo hạn dài dòng chảy sông lớn của Việt Nam. Trong bài báo này xin trình bày tóm tắt về phương pháp và một số kết quả ứng dụng nó trong dự báo thủy văn hạn dài.

1. Cơ sở của phương pháp PTTH

Trên cơ sở lý thuyết về các quá trình vĩ mô tự nhiên [1,2,3], xuất phát từ quan điểm thông tin - dự báo, qui luật hình thành và phát triển của các quá trình thủy văn được xét đến có dạng sau:

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

Trong đó tổng thứ nhất tạo nên bởi tác động của các nguyên nhân bên trong (B_{it}), tổng thứ hai do các nguyên nhân bên ngoài (C_{jt}) và tổng thứ ba - tác động chung của các thành phần ngẫu nhiên (U_{kt}).

Như vậy, một quá trình tự nhiên được phát triển liên tục do tác động của chính bản thân quá trình (các nguyên nhân bên trong), đồng thời chịu tác động của các nguyên nhân bên ngoài và các thành phần ngẫu nhiên - sinh ra trong suốt thời gian tạo nên quá trình đó, qua nhiều giai đoạn và do nhiều nguyên nhân tạo thành.

Rõ ràng theo qui luật tự nhiên, đến lượt mình các nguyên nhân tác động bên ngoài của quá trình lại là những quá trình tự nhiên, có đầy đủ 3 thành phần của mình tuân theo dạng tổng quát (1) có các nguyên nhân của mình và đồng thời cũng có quy luật phát triển nội tại và thành phần ngẫu nhiên sinh ra trong quá trình tạo nên chúng.

2. Mô hình dự báo

Dạng tổng quát của mô hình dự báo bằng phương pháp PTTH là:

$$Q_t = Q_{-t}^* + \sum_{i=1}^l a_{it} q(t-i) + \sum_{j=1}^m \sum_{k=1}^n b_{jkt}^i \lambda^j (t-i) + U_t \quad (2)$$

Trong biểu thức (2), Q_t là lưu lượng dòng chảy trung bình trong thời đoạn t ; Q^* - giá trị trung bình nhiều năm của chuỗi Q_t ; $q(t-i)$ - chuẩn sai của Q tại thời đoạn $(t-i)$; a_{it} - hệ số tự hồi quy, tính đến đóng góp của $q(t-i)$ vào Q_t ; l - số thành phần tự hồi qui đóng góp vào phương trình; $X^j(t-i)$ - giá trị của nhân tố thứ j tại thời đoạn $(t-i)$; n - thời gian trễ mà nhân tố thứ j có ảnh hưởng đến Q_t ; b_{it}^j - hệ số hồi qui tương ứng với nhân tố $X^j(t-i)$ và U_t là thành phần ngẫu nhiên gồm cả sai số do thiếu thông tin.

3. Thuật toán tổng quát của phương pháp PTTH

Thuật toán chủ yếu của phương pháp PTTH là kết hợp các mặt mạnh của 3 phương pháp: phương pháp động lực thống kê, phương pháp hồi qui nhiều biến lọc từng bước và mô hình tự hồi qui AR(p).

3.1. Phương pháp động lực thống kê (DLTK)

Từ giả thiết về sự tồn tại của mối quan hệ bên trong của các quá trình vĩ mô tự nhiên, nội dung cơ bản của phương pháp DLTK là xác định giá trị tương lai của quá trình qua các giá trị của nó trong quá khứ thông qua mối quan hệ sau:

$$q_t = f(q_{t-1}, q_{t-2}, q_{t-3}, \dots, q_{t-m}) \quad (3)$$

trong đó q_t là giá trị của quá trình tại thời điểm t cần dự báo; $q_{t-1}, q_{t-2}, q_{t-3}, \dots, q_{t-m}$ - các giá trị trong quá khứ của quá trình; m - số thành phần quá khứ tham gia vào mối quan hệ (3).

Việc thực hiện số trị mối quan hệ (3) được tiến hành trên cơ sở xác định cấu trúc nội tại của quá trình thông qua hàm tự tương quan và dẫn đến giải hệ các phương trình tuyến tính:

$$q_t = \sum_{\tau=1}^m k(\tau) q(t-\tau) \quad (4)$$

Các hệ số tự hồi qui $k(\tau)$ được xác định qua hệ các phương trình tương quan tuyến tính với điều kiện tối thiểu phương sai dư của phương trình (4). Các giá trị tối ưu của tham số m được xác định thông qua các chỉ tiêu thống kê:

- Chỉ tiêu tương quan, hệ số tương quan giữa tính toán và thực đo R_{td}

$$R_{td} = \left(\sum_{i=m+1}^n q_i q_i^* \right) / ((n-m) \sigma \sigma^*) \rightarrow \max \quad (5)$$

trong đó q_i là giá trị thực đo; q_i^* - giá trị tính toán; σ^2 và σ^{*2} - phương sai tương ứng của chuỗi q_i và q_i^* .

- Chỉ tiêu ngẫu nhiên:

$$\delta(m) = (\sigma_{\Delta m})^2 / \sigma^2 \rightarrow \min \quad (6)$$

với $(\sigma_{\Delta m})^2$ là phương sai dư hay phương sai của chuỗi sai số tính toán khi có m tham số trong phương trình (4).

3.2. Phương pháp phân tích hồi qui nhiều nhân tố lọc từng bước

3.2.1. Phương pháp hồi qui bội

Mô hình hồi qui tuyến tính bội có dạng:

$$Y_i = b_0 + \sum_{j=1}^m b_j X_{ji} + U_i \quad (7)$$

với Y_i là giá trị của biến phụ thuộc; X_{ji} - giá trị thứ i của biến độc lập thứ j ; b_j ($j = 0 \dots m$) - các hệ số hồi qui; U_i - biến ngẫu nhiên không tương quan có kỳ vọng bằng không và phương sai không đổi.

Dưới dạng ma trận có:

$$Y = Xb + U \quad (8)$$

Y là vectơ cấp ($n \times 1$) của biến phụ thuộc; X - ma trận cấp ($n \times (m+1)$) các biến độc lập; n - độ dài của chuỗi quan trắc; m - số biến độc lập; b - vectơ hàng chứa ($m+2$) tham số, gồm giá trị trung bình của chuỗi Y và các tham số $b_0 \dots b_m$.

Để ước lượng các tham số của mô hình, đã sử dụng phương pháp bình phương nhỏ nhất, trong đó vectơ tham số b được tìm dưới dạng:

$$\bar{b} = (X^T X)^{-1} X^T Y \quad (9)$$

với T là ký hiệu chuyển vị.

Vectơ tham số b tính qua biểu thức (9) là ước lượng tuyến tính không chêch. Phương sai dư được đánh giá qua biểu thức:

$$\tilde{\sigma}_u^2 = (n-m)^{-1} s^2 \quad (10)$$

$$\text{với } S^2 = Y^T Y - \bar{b}^T X^T Y$$

Hệ số xác định R^2 , tỷ số giữa phương sai giải tích và phương sai toàn phần:

$$R^2 = \left(\sum_{i=1}^n (Y_i^* - \bar{Y})^2 \right) / \left(\sum_{i=1}^n (Y_i - \bar{Y})^2 \right) = (b^T X^T - n \bar{Y}^2) / (Y^T Y - n \bar{Y}^2) \quad (11)$$

trong đó Y_i là giá trị thứ i của chuỗi dòng chảy với trung bình là \bar{Y} ; Y_i^* - ước lượng của Y_i bằng mô hình hồi qui và n - độ dài chuỗi tính toán.

Ý nghĩa của phương trình hồi qui được kiểm định thông qua các chỉ tiêu thống kê:

- Chỉ tiêu Fisher (F), kiểm định giả thiết tồn tại của phương trình hồi qui, tính theo công thức:

$$F = ((n-p-1)/p) \times (R^2/(1-R^2)) \quad (12)$$

Phương trình được chấp nhận khi giá trị tính toán của F lớn hơn giá trị tra theo bảng chỉ tiêu Fisher với mức ý nghĩa $\alpha = 5\%$ và số bậc tự do bằng $(n-p-1)$.

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

$$t_i = \left| \frac{b_i}{S_{bi}} \right| \quad (13)$$

S_{bi} là sai số chuẩn của hệ số hồi qui b_i . Nếu t_i có giá trị tính được theo (13) lớn hơn giá trị tra theo bảng phân phối Student với mức ý nghĩa $\alpha = 5\%$ và số bậc tự do là $(n-p-1)$ thì b_i là có nghĩa.

3.2.2. Tuyển chọn các nhân tố dự báo

Bản chất của việc dự báo là tìm những kết cục sẽ xảy ra của đại lượng dự báo tại một thời điểm nào đó trong tương lai từ những thông tin hiện có. Do vậy, chất lượng của dự báo sẽ được quyết định bởi chất lượng của thông tin.

Để tuyển chọn các nhân tố dự báo, phương pháp PTTH đã sử dụng kết hợp hai cách:

- Phân tích vật lý, bằng phương pháp luận synop, phương trình cân bằng nước phân tích cơ chế ảnh hưởng và các mối quan hệ của các yếu tố khí hậu, thủy văn, hoàn lưu khí quyển lựa chọn sơ bộ các nhân tố dự báo. Trong đó có thể sử dụng cả các đặc trưng thống kê, các phép biến đổi số liệu ban đầu v.v.

- Liên kết phân tích tương quan với phân tích phương sai thông qua các chỉ tiêu thống kê Fisher (12), chỉ tiêu Student (13) và mức đảm bảo của dự báo để tuyển chọn và tổ hợp thông tin trong quá trình hồi qui lọc từng bước.

Nội dung của phép hồi qui lọc từng bước là thực hiện các quá trình đưa vào và đưa ra các nhân tố dự báo để chọn các nhân tố và tổ hợp các nhân tố tốt nhất trong việc thiết lập phương trình hồi qui tuyển tính nhiều chiều với mục tiêu là giảm phương sai dư đến mức tối đa. Biến mới được chọn đưa vào phương trình là biến có chỉ số giảm phương sai dư lớn nhất trong các biến còn lại:

$$C_{j=1,i} = (R^2_{j=1,i} - R^2_j) / (1 - R^2_j) \rightarrow \max \quad (14)$$

Trong đó: $C_{j=1,i}$ là chỉ số giảm phương sai dư khi thêm biến thứ i vào phương trình hồi qui; i - chỉ số của các biến chưa được đưa vào phương trình ở bước thứ j ; R^2_j và $R^2_{j=1,i}$ - hệ số xác định trước và sau khi đưa thêm vào phương trình biến thứ i .

Để chọn số nhân tố tối ưu cho mô hình và từ đó xác định điểm dừng của quá trình lọc đã sử dụng chỉ tiêu mức đảm bảo dự báo. Quá trình chọn lọc được dừng khi tìm được phương trình có mức đảm bảo dự báo cao nhất.

3.3. Mô hình tự hồi qui AR(p)

Thành phần sai số dự báo U_t được mô phỏng bằng mô hình tự tương quan dưới dạng:

$$U_t = \sum_{i=1}^p \phi_i U_{t-i} + a_t \quad (15)$$

các hệ số ϕ_i được xác định thông qua hệ phương trình tương quan tuyến tính:

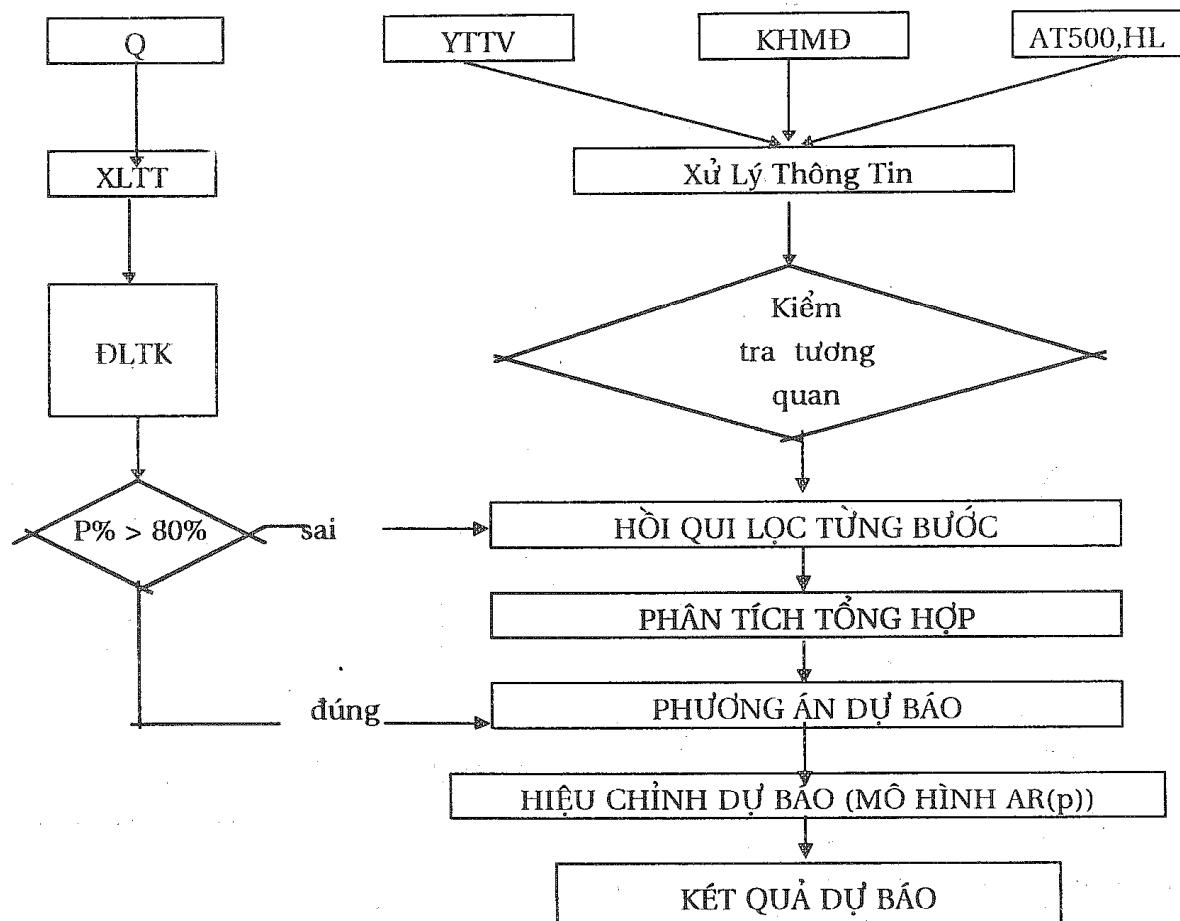
$$\rho_k = \sum_{i=1}^p \phi_i \rho_{k-i}, \quad k = 1, 2, \dots \quad (16)$$

bằng phương pháp bình phương nhỏ nhất. Trong biểu thức (16), ρ_k, ρ_{k-i} là các giá trị của hàm tự tương quan tại các bước thứ $k, k-i$.

Do tính không dừng của quá trình khí tượng thủy văn và đặc biệt do sự ảnh hưởng của biến đổi khí hậu, hoạt động của con người, vai trò của các thành phần cấu thành các quá trình tự nhiên trong mọi mảnh đất tạo nên quá trình dòng chảy sông ít nhiều có sự biến động. Vì vậy, cũng như bất cứ một mô hình nào, phương pháp PTTH cần được thích nghi thường xuyên các thông số của phương trình với nguồn thông tin mới nhất.

4. Sơ đồ tính

Dưới dạng tổng quát, sơ đồ tính toán xây dựng các phương trình dự báo bằng phương pháp PTTH được mô tả trong hình 1.



Hình 1: SƠ ĐỒ TỔNG QUÁT CỦA PHƯƠNG PHÁP PTTH

Trong đó phương pháp ĐLTK được sử dụng để phân tích các chuỗi dòng chảy đã được chuẩn hóa và khử thứ nguyên. Phương pháp này được sử dụng với hai chức năng: phân tích cấu trúc nội tại của chuỗi yếu tố dự báo, xây dựng quan hệ dự báo và đánh giá lượng thông tin đóng góp của các thành phần quá khứ vào quá trình của nó trong tương lai. Trong trường hợp khi mức đảm bảo của phương pháp ĐLTK đạt trên 80% thì phương pháp này được sử dụng như một phương án độc lập. Trong các trường hợp khác, phương pháp được sử dụng như một công cụ tuyển chọn thông tin cho phương pháp HQNB lọc từng bước. Khi đó các chuỗi dòng chảy với các thời gian trễ khác nhau, tham gia trong phương trình tối ưu của phương pháp ĐLTK sẽ được sử dụng như các nhân tố độc lập của phương pháp HQNB.

Trong nhánh thứ hai của sơ đồ, các chuỗi số liệu về nhân tố dự báo với thời gian trễ khác nhau, được chuẩn hóa, khử thứ nguyên và được tuyển lại qua sai số chuẩn của hệ số tương quan. Các nhân tố này kết hợp với các nhân tố đã được tuyển chọn bằng phương pháp ĐLTK lập nên bộ nhân tố cơ sở. Bằng phương pháp HQNB lọc từng bước tiến hành xây dựng các phương trình dự báo. Sai số của phương trình tối ưu sẽ được hiệu chỉnh bằng mô hình AR(p).

5. Kết quả ứng dụng phương pháp PTTH trong dự báo thủy văn hạn dài

Phương pháp PTTH đã được sử dụng trong nghiên cứu xây dựng các phương án dự báo dòng chảy tháng, mùa, năm của các trạm chính trên hệ thống sông Hồng, sông Thái Bình [4,5,6] và dự báo tổng lượng dòng chảy đến hồ Hòa Bình thời kỳ tích nước [8]. Kết quả dự báo được đánh giá thông qua hệ số tương quan giữa tính toán với thực đo (R_{dt}) và mức bảo đảm của dự báo độc lập (p%). Số liệu trong bảng 1 cho thấy phương pháp PTTH có chất lượng tốt đáp ứng được yêu cầu của dự báo thủy văn hạn dài. Với kết quả này, phương pháp PTTH đã được xây dựng thành công nghệ dự báo và áp dụng trong dự báo tác nghiệp dòng chảy tháng, mùa, năm và dòng chảy thời kỳ tích nước (21/VIII - 31/X) đến hồ Hòa Bình phục vụ khai thác thủy điện và phòng chống lụt hả du sông Hồng từ năm 1992.

Bảng 1. Kết quả dự báo hạn dài bằng phương pháp PTTH

Trạm	Yếu tố	$R_{dt} \times 100$	P (%)	Trạm	Yếu tố	R_{dt}	P (%)
hồ Hòa Bình	Qth	83 - 93	75-92	Hà Nội	Qth	85 - 94	75-91
	Qmùa cạn/lũ	94/93	85/77		Qmùa cạn/lũ	91/88	81/73
	Qnăm	90	86		Qnăm	88	83
	Q21/8-31/10	86	76	Phả Lại	Hth	89 - 96	77-88
Yên Bá	Qth	83 - 95	77-91		Hmùa cạn/lũ	92/91	86/85
	Qmùa cạn/lũ	91/93	88/81		Hnăm	96	90
	Qnăm	96	83				
Vụ Quang	Qth	87 - 96	72-92				
	Qmùa	87/91	88/73				
	Qnăm	94	79				

6. Kết luận

Từ kết quả nghiên cứu ứng dụng phương pháp PTTH trong dự báo hạn dài dòng chảy các sông lớn của Việt Nam có thể rút ra một số kết luận thống kê sau:

1. Trên cơ sở lý thuyết về quá trình vĩ mô tự nhiên, xuất phát từ quan điểm thông tin dự báo, phương pháp PTTH đã kết hợp được hai xu hướng phát triển chính trong lĩnh vực dự báo thủy văn hạn dài hiện nay là xu hướng cẩn nguyên và thống kê toán lý.
2. Bằng sơ đồ phân tích tổng hợp khí hậu - thủy văn quán tính với tỷ trọng thực nghiệm và bằng biện pháp tuyển chọn thông tin kết hợp giữa phân tích cẩn nguyên vật lý với sơ đồ hồi quy lọc từng bước, phương pháp PTTH đã tận dụng được tối ưu lượng thông tin hiện có về yếu tố dự báo.
3. Các kết quả thử nghiệm dự báo với số liệu phụ thuộc, số liệu độc lập, và kết quả dự báo tác nghiệp đã chứng tỏ chất lượng dự báo của phương pháp là khả quan và đáp ứng được yêu cầu hiện nay.

Tài liệu tham khảo

1. Aliôkhin Iu. M. (1963). Dự báo thống kê trong quá trình địa vật lý. NXB. MGU, Moskva (tiếng Nga).
2. Box G.E.P., Jenkins G.M. (1970). Time series analysis: forecasting and control. Holden - Day.
3. Kartvellisveli N. A. (1981). Thủy văn ngẫu nhiên. Nxb KTTV, Liên Xô (tiếng Nga).
4. Nguyễn Viết Thi (1990). Dự báo dòng chảy tháng đến hồ Hòa Bình bằng phương pháp phân tích tổng hợp. Tuyển tập các báo cáo khoa học tại Hội nghị khoa học về dự báo KTTV lần thứ III, Hà Nội.
5. Nguyễn Viết Thi (1992). Sử dụng phương pháp phân tích tổng hợp dự báo dòng chảy năm trên lưu vực sông Hồng. Báo cáo khoa học tại hội nghị khoa học của Viện KTTV, Hà Nội.
6. Nguyễn Viết thi (1993). Dự báo hạn dài dòng chảy sông lớn (Báo cáo tổng kết đề tài nghiên cứu cấp tổng cục), Hà Nội.
7. Nguyễn Viết Thi (1994). Xây dựng công nghệ dự báo dòng chảy tháng đến hồ Hòa Bình bằng phương pháp phân tích tổng hợp. Tổng kết đề tài NC thuộc dự án LAN của TTQG Dự báo KTTV. Hà Nội.
8. Nguyễn Viết Thi và các cộng tác viên (1994). Dự báo tổng lượng dòng chảy và đỉnh lũ thời kỳ 21/VIII - 31/X đến hồ Hòa Bình. Báo cáo tổng kết chuyên đề dự báo phục vụ khai thác hồ Hòa Bình.