

NGOẠI SUY SỐ LIỆU TRONG TÍNH TOÁN THỦY VĂN

PTS. TRẦN ĐỨC HẢI
Viện Khí tượng Thủy văn

Trong tính toán quy hoạch, thiết kế, khai thác các công trình thủy lợi: tính toán tài nguyên nước cho lưu vực sông hoặc vùng lãnh thổ v.v. số liệu đo dòng chảy từ lưu vực hoặc khu vực nghiên cứu là rất cần thiết và quan trọng.

Tuy nhiên trong thực tế tính toán thường gặp các trường hợp:

- Có số liệu đo nhiều năm nhưng không liên tục;
- Có số liệu, nhưng số năm quan trắc chưa đủ nhiều để bảo đảm yêu cầu về sai số khi tính toán;
- Hoàn toàn không có số liệu quan trắc dòng chảy tại lưu vực nghiên cứu.

Thực trạng này đã đặt ra yêu cầu cần ngoại suy số liệu dòng chảy để có được chuỗi số đủ dài bảo đảm các yêu cầu khi tính toán thống kê. Thực tế cũng đã chứng tỏ đây là một trong những vấn đề quan trọng và cần thiết của thủy văn, đặc biệt trong điều kiện nước ta.

Năm 1973, Tổ chức Khí tượng thế giới (WMO) và ủy ban văn hóa, khoa học, giáo dục của Liên hiệp quốc (UNESCO) đã phối hợp tổ chức một hội nghị quốc tế « Tính toán thiết kế các công trình thủy lợi trong trường hợp không đủ số liệu [5].

Ở đây khái niệm ngoại suy bao gồm việc kéo dài những chuỗi quan trắc còn ngắn, bổ sung thành phần cho các chuỗi không liên tục hoặc tạo chuỗi dòng chảy cho trường hợp không có số liệu quan trắc.

Hiện có nhiều phương pháp được đưa ra và áp dụng có hiệu quả để giải quyết vấn đề này, nhưng có thể lập hợp thành 3 nhóm:

1. Phương pháp hồi quy
2. Các mô hình tất định
3. Các mô hình ngẫu nhiên.

Ở nước ta những phương pháp này cũng đã và đang được áp dụng có kết quả. Trong bài này xin giới thiệu một số nét cơ bản qua kinh nghiệm ứng dụng những phương pháp đó.

1. Phương pháp hồi quy

Phương pháp hồi quy dựa trên cơ sở tương quan giữa chuỗi dòng chảy đo được hoặc có thể tồn tại ở lưu vực tính toán với chuỗi dòng chảy đo đồng thời tại một hoặc vài lưu vực lân cận có số năm quan trắc đầy đủ hơn và các yếu tố địa lý tự nhiên quyết định quá trình sản sinh dòng chảy (như mưa, các đặc trưng mặt đệm, hình thái lưu vực v.v.).

Trong tính toán thường sử dụng tương quan đơn biến vì chúng đơn giản. Nhưng trong nhiều trường hợp, để nâng cao độ chính xác các kết quả tính toán hoặc tương quan đơn biến không bảo đảm yêu cầu, cần sử dụng tương quan nhiều biến. Vấn đề ở đây là phải đánh giá được mức ổn định của các phương trình tương quan và độ tin cậy của các hệ số hồi quy. Kinh nghiệm sử dụng tương quan tuyến tính nhiều biến cho thấy số lượng biến độc lập thường không nên nhiều hơn 3 hoặc 4. Việc đưa vào tương quan quá nhiều biến chỉ dẫn đến sự mất ổn định các hệ số hồi quy và lãng phí số ngẫu nhiên khi tính các hệ số đó.

Độ chính xác của phương pháp hồi quy phụ thuộc nhiều vào hệ số tương quan và độ dài chuỗi quan trắc đồng thời. Khi sử dụng phương pháp này cần chú ý các điều kiện sau [3].

Số năm cùng quan trắc giữa lưu vực tính toán và lưu vực tương tự không ít hơn 10 năm.

$$n \geq 10$$

(1)

Hệ số tương quan (đơn biến hoặc nhiều biến) không nhỏ hơn 0,7

$$r, R \geq 0.70$$

(2)

Tỷ số giữa hệ số hồi quy và sai số quân phương của hệ số đó không nhỏ hơn 2

$$k_j / \sigma_{k_j} \geq 2$$

(3)

Một khía cạnh khác cần lưu ý là những thành phần trong chuỗi được ngoại suy bằng phương trình hồi quy có quan hệ hàm số với các thành phần tương ứng tại lưu vực tương tự, trong khi các trị số thực đo lại có quan hệ không đơn trị do có các thành phần ngẫu nhiên. Vì vậy, để bảo đảm sự đồng nhất giữa các thành phần thực đo và phần tính toán bằng phương trình hồi quy cần thêm thành phần ngẫu nhiên vào từng trị số ngoại suy:

$$Q'_i = Q_i + \varepsilon_i \sigma_Q \sqrt{1 - R^2} \quad (4)$$

trong đó Q_i — trị số ngoại suy theo phương trình hồi quy;

ε_i — biến ngẫu nhiên chuẩn với số trung bình bằng 0 và phương sai bằng 1:

d_Q — Phương sai của Q ; hệ số tương quan (đơn biến hoặc nhiều biến).
Cũng có thể sử dụng công thức đơn giản [4]

$$q'_i = \frac{q_i - \bar{q}}{R} + \bar{q} \quad (5)$$

Trong đó $\bar{q}$ — trị số môđun trung bình trong thời kỳ tính toán.

Kinh nghiệm sử dụng phương pháp tương quan ở nước ta cho thấy thường chỉ có đặc trưng trung bình năm mới có tương quan không gian chặt ($r \geq 0,70$). Với các đặc trưng mùa, tháng, ngày và cực trị thường rất khó tìm được trạm tương tự để xác định hệ số tương quan.

Phương pháp được sử dụng rộng rãi, dễ dàng, các bước tính toán, xác định hệ số tương quan, phương trình hồi quy v.v. đã được giới thiệu nhiều trong các sách vở trong và ngoài nước. Để việc tính toán được nhanh chóng đặc biệt khi dùng tương quan nhiều biến và phải tính cho nhiều trường hợp máy tính điện tử, nhất là máy vi tính đã được sử dụng rộng rãi.

2. Mô hình tất định—mô hình nhận thức

Theo cách phân loại của WMO mô hình tất định được chia ra 2 loại mô hình hộp đen và mô hình nhận thức (còn gọi là mô hình hộp xám).

Các mô hình nhận thức mưa—dòng chảy thường được sử dụng trong dự báo. Tuy nhiên, với khả năng tổng hợp dòng chảy từ mưa, những mô hình này còn có thể sử dụng cho mục đích ngoại suy số liệu để từ lượng mưa tương ứng tính ra dòng chảy năm, tháng, ngày thậm chí dòng chảy trên lũ nếu có số liệu mưa tự ghi đủ tin cậy.

Mô hình nhận thức là loại mô hình toán mô phỏng các quá trình hình thành và tập trung dòng chảy trên lưu vực sông. Các quá trình mưa, thấm, bốc thoát hơi; quá trình sản sinh các thành phần dòng chảy (mặt, dưới mặt, ngầm v.v.) tập trung nước trong mạng lưới sông, được mô tả như những quá trình biến đổi theo thời gian, thậm chí cả trong không gian và trong mối tương quan vật lý giữa chúng. Vì vậy, những mô hình này thường phức tạp, có nhiều thông số, phải mô phỏng trên máy tính điện tử. Các bước tính toán cơ bản bao gồm:

— Lựa chọn mô hình.

— Hiệu chỉnh mô hình.

— Dùng mô hình đã hiệu chỉnh từ quá trình mưa tính ra dòng chảy tương ứng.

Đã có rất nhiều mô hình nhận thức mưa—dòng chảy được đưa ra. Tổ chức Khí tượng thế giới đã chọn 10 mô hình tiêu biểu để kiểm tra, đánh giá [6] mô hình nào cũng có ưu và khuyết điểm. Việc lựa chọn mô hình chủ yếu

phụ thuộc vào sự hiểu biết của người sử dụng và điều kiện cơ sở vật chất sẵn có (chương trình tính và máy tính). Ở nước ta hiện có 2 mô hình nhận thức được nhiều cơ quan nghiên cứu sử dụng là mô hình TANK và SSARR [2]. Chương trình tính theo hai mô hình này đã được chuyển đổi sang hệ máy vi tính nên việc sử dụng rất thuận tiện.

Để hiệu chỉnh mô hình (xác định bộ thông số phù hợp cho một lưu vực cụ thể) cần có số liệu thực đo trong vài năm về mưa, dòng chảy, bốc hơi đủ độ tin cậy. Kinh nghiệm sử dụng mô hình SSARR cho thấy phải có khoảng 5 đến 8 năm số liệu để bảo đảm ổn định của bộ thông số được chọn. Trong thực tế ở các tuyến công trình chỉ có 2-3 năm quan trắc và số liệu điều tra nhưng vẫn có thể sử dụng để hiệu chỉnh mô hình rồi tính toán như một phương án tham khảo.

Sau khi hiệu chỉnh mô hình, có được bộ thông số ổn định nhận được coi như ổn định theo thời gian với giả thiết là những điều kiện cơ bản trên lưu vực ít thay đổi. Dùng bộ thông số đó, từ mưa có thể tính ra dòng chảy tương ứng. Như vậy, cần phải có số liệu mưa bình quân lưu vực hoặc mưa trạm đại biểu (như với mô hình SSARR). Độ dài thời gian ngoại suy phụ thuộc vào độ dài chuỗi số liệu mưa.

Thường các mô hình nhận thức (SSARR hoặc TANK đang được sử dụng ở nước ta) được dùng để tính quá trình dòng chảy bình quân ngày từ lượng mưa ngày tương ứng. Tuy nhiên, cũng có thể dùng lượng mưa tháng để tính ra dòng chảy bình quân tháng, hoặc thậm chí dùng lượng mưa thời đoạn nhỏ hơn ngày để tính ra dòng chảy trung bình thời đoạn nhỏ hơn ngày để tính ra dòng chảy trung bình thời đoạn tương ứng.

Khi tính toán cần lưu ý:

- Hiệu chỉnh mô hình theo số liệu những trạm mưa nào thì khi tính cũng phải có số liệu của những trạm mưa đó.

- Hiệu chỉnh mô hình theo thời đoạn nào (giờ, ngày, tháng) thì chỉ áp dụng để tính cho thời đoạn đó. Ví dụ: nếu hiệu chỉnh mô hình theo mưa ngày, không được dùng bộ thông số đó để tính cho mưa tháng hoặc mưa giờ.

Để quá trình hiệu chỉnh được nhanh chóng và hợp lý cần tận dụng mọi loại số liệu quan trắc có được trên lưu vực như mưa, dòng chảy, các đặc trưng khu vực v.v.. Bằng cách phân tích đánh giá mối tương quan giữa chúng để có được những ước lượng ban đầu về các thông số. Ví dụ: phân tích quá trình mưa, dòng chảy kết hợp với những số liệu về lớp phủ mặt lưu vực (thổ nhưỡng, thực vật) có thể ước lượng các thành phần dòng chảy và quan hệ giữa chúng. Các thông số liên quan đến dòng chảy dưới mặt đất có thể ước tính từ kết quả phân tích đường nước rút. Cũng cần phân tích phương trình cân bằng nước đơn giản để xem các giá trị trung bình của mưa, dòng chảy, tổn thất có nằm trong mối tương quan hợp lý không, từ đó ước tính lượng mưa bình quân lưu vực và các thông số liên quan đến mưa, tổn thất (như: thông số tỷ trọng trạm mưa và quan hệ ROP-SMI trong mô hình SSARR).

3. Mô hình ngẫu nhiên

Mô hình hóa dòng chảy sông ngòi như quá trình ngẫu nhiên là một phương pháp có hiệu quả trong tính toán thiết kế các hệ thống thủy lợi và ngày càng được sử dụng rộng rãi, đặc biệt là trong trường hợp không có số liệu đo dòng chảy tại lưu vực tính toán.

Hầu hết các mô hình ngẫu nhiên được xây dựng để mô phỏng quá trình dòng chảy năm là quá trình thường được coi là dừng. Mô hình dòng chảy các thời đoạn ngắn hơn như tháng, tuần phức tạp hơn nhiều vì số lượng thông số lằng lằng đáng kể (ví dụ mô hình Thomas Fiering cho dòng chảy năm có 3 thông số, nhưng cho dòng chảy tháng có tới 36 thông số). Tuy nhiên, trong thực tế thường cần số liệu dòng chảy tháng nên một số mô hình ngẫu nhiên cho dòng chảy tháng đã được ứng dụng rộng rãi như mô hình chuỗi Markov, mô hình tự hồi quy trung bình trượt (ARMA, ARIMA) mô hình Thomas Fiering v.v..

Khác với các mô hình nhận thức chỉ cần vài năm số liệu đo để nhận biết các thông số, khi sử dụng mô hình ngẫu nhiên cần có số năm quan trắc đủ để bảo đảm tính ổn định của các tham số thống kê. Tuy nhiên, các tham số thống kê có thể xác định theo số liệu đo tại tuyến công trình hoặc ở vị trí tương tự (lưu vực tương tự). Với mô hình nhận thức chưa có khả năng dùng số liệu lưu vực tương tự để hiệu chỉnh.

Mô hình nhận thức được kiểm nghiệm bằng cách so sánh quá trình dòng chảy tính toán và thực đo theo trình tự thời gian. Mô hình thống kê chỉ được đánh giá theo mức độ phù hợp về mặt thống kê các thông số của chuỗi số liệu quan trắc và chuỗi tổng hợp được.

Ở nước ta mô hình Thomas Fiering đã được ứng dụng để tạo chuỗi dòng chảy tháng trong 100 năm phục vụ việc quản lý hồ chứa Hòa Bình [5].

Với đơn vị thời gian là tháng, biểu thức toán của mô hình được viết dưới dạng

$$Q_{i+1} = \bar{Q}_{i+1} + \varphi_i (Q_i - \bar{Q}_i) + T_{i+1} + \sqrt{1 - R^2} V_{i+1} \quad (6)$$

Trong đó Q_{i+1} , Q_i — dòng chảy tháng thứ $i+1$ và i ,
 $\bar{Q}_{i+1}$, $\bar{Q}$ — trị số trung bình,

φ_i — hệ số hồi quy,

T_{i+1} — Biến ngẫu nhiên độc lập có luật phân phối chuẩn với kỳ vọng bằng 0 và phương sai bằng 1.

$\sigma_{Q_{i+1}}$ — độ lệch chuẩn của dòng chảy,

R — Hệ số tương quan.

Các bước tính cơ bản khi áp dụng mô hình là:

- Xác định các đặc trưng thống kê của chuỗi thực đo.
- Xác định luật phân phối của chuỗi thực đo để chọn phương pháp tạo biến ngẫu nhiên T .

- Thực hiện các phân tích tương quan và hồi qui.
- Tạo số ngẫu nhiên và tính biến ngẫu nhiên T.

Sau các bước trên sẽ nhận được hệ 12 phương trình hồi qui tuyến tính cho 12 tháng. Trong phương trình đầu i sẽ là 12 (tháng 12) còn $i+1$ là 1 (tháng 1) sau đó theo trình tự i sẽ là 1, 2, ..., 11 và $i+1$ là 2, 3 ..., 12.

Trong trường hợp cụ thể này, chuỗi thực đo được lấy lại tuyến Hòa Bình từ năm 1961 đến 1978.

Mô hình ngẫu nhiên còn được một số nhà thủy văn Ấn Độ dùng để tạo chuỗi dòng chảy trận lũ lớn nhất [4]. Bài toán đặt ra như sau: xác định quá trình lũ thiết kế cho sông Yamuna tại tuyến Okha ($F = 6811 \text{ km}^2$) trong điều kiện hoàn toàn không có số liệu đo dòng chảy tại tuyến. Để giải quyết bài toán này các nhà thủy văn Ấn Độ đã chọn mô hình chuỗi Markov không dừng, kết hợp với phương pháp đường đơn vị.

Mô hình Markov được dùng khi tạo chuỗi mưa trận lũ lớn nhất năm trong 100 năm. Sau đó dùng hệ số dòng chảy ($\alpha = 0,80$) tính chuyển sang lớp dòng chảy và dùng đường đơn vị để tính quá trình lũ thiết kế.

Thời đoạn mưa tính toán được chọn bằng 6 giờ. Số liệu mưa lớn nhất 6 giờ được trích từ số liệu đo trong 10 năm (1956-1965) tại Trạm New Delhi nằm gần tuyến công trình. Các trận mưa 6 giờ được sắp xếp sao cho mưa bắt đầu vào giờ thứ nhất sau đó kéo liên tục và kết thúc vào giờ thứ 6 (Trong các trận mưa thực tế sự sắp xếp có thể không hoàn toàn như vậy. Ví dụ có thể có ngắt quãng trong quá trình mưa).

Mô hình tính toán có dạng

$$x_{t,j} = rX_{t-1,j} + \varepsilon_{t,j} \quad (7)$$

Ở đây: $X_{t,j}$ - lượng mưa giờ thứ t của năm thứ j .

$X_{t-1,j}$ - lượng mưa giờ trước đó; r - hệ số chuỗi Markov; $\varepsilon_{t,j}$ thành phần ngẫu nhiên do lượng mưa giờ X_t ; t - thời gian tính bằng giờ ($t = 1 - 6$ và $j = 1 - 100$).

Từ phương trình (7), khi $t = 1$ nghĩa là vào giờ thứ nhất lượng mưa bằng trị số $\varepsilon_{1,j}$ vào thời điểm trước đó ($t-1 = 0$) mưa chưa bắt đầu ($X_{t-1} = 0$).

Hệ số chuỗi Markov (r) và thành phần ngẫu nhiên cho những năm có số liệu, được xác định từ lượng mưa đã biết bằng phương pháp bình phương tối thiểu khi chọn thích hợp đường quan hệ X_t và X_{t-1} .

Các thông số phân bố như trị số trung bình, độ lệch chuẩn của chuỗi số liệu mưa quan trắc, được tính cho từng giờ.

100 số hạng tựa ngẫu nhiên (pseudo-random numbers) có luật phân bố đồng nhất của giờ thứ nhất được tạo lập trên máy tính IRM1401. Các số ngẫu nhiên đó được coi là lượng mưa giờ thứ nhất của 100 trận mưa và được dùng để tính 100 trị số lượng mưa giờ thứ 2. Lượng mưa giờ thứ 2 của 100 trận mưa vừa tính lại được dùng để tính lượng mưa giờ thứ 3. Cứ tính như vậy cho đến giờ thứ 6. Kết quả được 100 trận mưa với thời gian mưa bằng 6 giờ.

TÀI LIỆU THAM KHẢO

1. Nguyễn văn Lai. Ứng dụng mô hình toán thủy văn trong tính toán dòng chảy sông. Luận án PTS. Đ.H. Thủy lợi, 1987.
2. Nguyễn Văn Thắng. Phương pháp tính toán khôi phục chuỗi tài liệu lưu lượng theo mô hình TANK. Tuyển tập hội nghị khoa học lần thứ VIII Đ.H. Thủy lợi, 1984.
3. Hướng dẫn xác định các đặc trưng thủy văn tính toán. NXB Khí tượng Thủy văn Liên Xô, 1982.
4. Sharma H.D. The use of Simulation Techniques for Sequential Generation of Short-Sized Rainfall Data and its application in the Estimation of Design Flood. «Proc of the Madrid Symp. June 1973» V.1.
5. Design of Water Resources Projects With Inadequate Data. Proc of the Madrid Symp. June 1963 V.1, V.2.
6. WMO Project on Intercomparision of Conceptual Models Used in Operational Hydrological Forecasting. Operational Hydrology Report N°7. Geneva, 1975.