

MỘT SỐ KẾT QUẢ ỨNG DỤNG MÔ HÌNH ARIMA VÀO DỰ BÁO DÒNG CHẢY THẮNG ĐẾN HỒ HOÀ BÌNH

PTS. Bùi Văn Đức

Trung tâm quốc gia dự báo KTTV

Vấn đề dự báo thủy văn phục vụ điều hành hồ Hoà Bình chống lũ hạ du và phát điện luôn luôn là trung tâm chú ý của nhiều ngành và nhiều nhà khoa học. Đã có biết bao đề tài, hội thảo bàn về vấn đề nâng cao hiệu quả trong quản lý khai thác hồ Hoà Bình. Trong số đó cần nhắc tới đề tài cấp nhà nước " Nhận dạng lũ sông Hồng trong điều hành hồ Hoà Bình chống lũ hạ du", 1991 -1993 do trường Đại học thủy lợi và Cục dự báo KTTV (Trung tâm quốc gia dự báo KTTV) tiến hành. Các hội thảo về đề tài này đã được tiến hành tại Trung tâm quốc gia dự báo KTTV, Bộ thủy lợi, Công ty điện lực I. Cuối tháng 3 năm 1996, một hội thảo liên ngành (Thủy lợi, Khí tượng Thủy văn và Năng lượng) do Bộ KH-CN & MT chủ trì bàn về nâng cao mực nước trước lũ của hồ Hoà Bình. Trong đó đặc biệt nhấn mạnh đến tầm quan trọng của dự báo thủy văn nói chung và dự báo hạn vừa, hạn dài nói riêng phục vụ quản lý và khai thác hồ.

Cho đến nay vấn đề dự báo dòng chảy tháng đến hồ Hoà Bình còn đang khó khăn, chưa có phương án dự báo nào đạt độ chính xác cao, nhất là các tháng chuyển tiếp giữa các mùa. Để có thêm một phương án dự báo dòng chảy tháng đến hồ Hoà Bình, mô hình chuỗi thời gian ARIMA đã được chọn và đưa vào ứng dụng. Cơ sở vật lý - toán của mô hình trong mô phỏng và dự báo dòng chảy tháng đã được trình bày và đăng trong Tập san KTTV số 11(419)/1995 [1]. Phần mềm máy tính "STATISTICA" tương ứng với các thuật toán của mô hình ARIMA là do hãng MICROSOFT xây dựng [4].

Bài báo này chủ yếu trình bày về các kết quả nghiên cứu ứng dụng dự báo.

1. Số liệu nghiên cứu và phần mềm STATISTICA

Số liệu

Sông Đà là một trong các sông có số liệu quan trắc đồng bộ và sớm. Chuỗi dòng chảy tháng của sông Đà - tr. Hoà Bình có từ năm 1902, song không đồng bộ. Đặc biệt là thiếu quan trắc từ 1945 đến 1955. Chuỗi số liệu tháng (Q_{max} , $Q_{tr.Bình}$ và Q_{min}) đồng bộ hơn cả là từ 1956 đến nay. Sau năm 1986 đến nay do ảnh hưởng trực tiếp của hồ Hoà Bình, dòng chảy tự nhiên tại Hoà Bình không còn đo được, mà phải phục hồi lại. Dòng chảy phục hồi tại Cục dự báo KTTV thực chất có khác đôi chút với dòng chảy tự nhiên, do tổn thất nước khi có hồ (bay hơi và thấm). Song kết quả nghiên cứu tính đồng nhất [2] cho rằng có thể coi các chuỗi đặc trưng dòng chảy tự nhiên tháng sông Đà của giai đoạn trước (1986) và dòng chảy tháng phục hồi (1/1986-3/1995) tại Trung tâm quốc gia dự báo KTTV là đồng nhất.

Phần mềm STATISTICA Các tính toán được thực hiện bằng phần mềm "STATISTICA" thuộc bản quyền của hãng Microsoft [4]. STATISTICA bao gồm hàng loạt các bài toán thống kê chuẩn, trong đó có mô hình ARIMA. Phần mềm thống kê STATISTICA làm việc trong môi trường WINDOWS và phải được cài đặt trong đĩa cứng với dung lượng đĩa dành cho nó tối thiểu là 13,5 mb.

Qua thời gian nghiên cứu ứng dụng STATISTICA vào dự báo thủy văn, chúng tôi đã chọn dùng mô hình ARIMA để dự báo dòng chảy tháng (lưu lượng nước, m³/s) đến hồ Hoà Bình. Tư tưởng ứng dụng ARIMA vào dự báo thủy văn có từ lâu và đã được trình bày tương đối kỹ trong [1]. Đặc biệt mô hình rút gọn dạng AR(P) có cải biên đôi chút được ứng dụng nhiều. Ở nước ta dạng đầy đủ của ARIMA(p, d, q) chưa được ứng dụng bao giờ và đây là thử nghiệm ứng dụng mô hình ARIMA đầu tiên trong dự báo dòng chảy tháng ở Trung tâm quốc gia dự báo KTTV.

2. Phân tích chuỗi dòng chảy tháng và định dạng mô hình

Chuỗi dòng chảy tháng được sắp xếp tuần tự theo các tháng trong năm và kế tiếp hết năm này năm sang năm khác. Số đầu tiên của chuỗi nghiên cứu là lưu lượng trung bình tháng 1 năm 1956 và số cuối cùng là tháng 3 năm 1995.

Chuỗi số liệu được chọn với những tiêu chuẩn sau:

- Màng tính liên tục,
- Đảm bảo độ chính xác và tính đồng nhất,
- Dung lượng chuỗi đủ dài.

Các thành phần chính của chuỗi lưu lượng trung bình tháng (LLBT):

- Thành phần mùa với chu kỳ 12 tháng,
- Thành phần tự tương quan (quán tính) và
- Nhiều động.

Tỷ trọng 3 thành phần trên của chuỗi LLBT biến đổi từ tháng này sang tháng khác và từ năm này sang năm khác. Thành phần chu kỳ xác định quá trình nền theo trung bình nhiều năm, thành phần quán tính xác định ảnh hưởng của lượng trữ nước lưu vực. Những biến đổi lượng mưa lệch quy luật trung bình nhiều năm là một trong các nguyên nhân gây ra nhiều động chuỗi LLBT.

- Trong những tháng mùa kiệt, do mưa ít, các nhiễu động không đáng kể, trong chuỗi, tính quán tính hoạt động trội lên. Dòng chảy của tháng chủ yếu phụ thuộc vào lượng trữ trong sông, mà chỉ số của nó là lưu lượng tháng trước.

- Trong các tháng mùa lũ cả 3 thành phần (quán tính, chu kỳ và nhiễu) đều hoạt động.

Như vậy, cần thiết có các biến đổi chuẩn hoá chuỗi số liệu ban đầu phù hợp với bản chất và đặc tính trên. Mô hình dự báo sẽ có cả hai thành phần là tự tương quan và trung bình trượt cùng với sai phân hoá và xử lý quy luật mùa. Mô hình được chọn là ARIMA(p,d,q),(P,D,Q).

Chuỗi dòng chảy tháng (Qt) sau khi được logarit hoá $Lx(Qt)$ được xem như các dao động điều hòa với các biên độ và chu kỳ khác nhau. Đã tiến hành khảo sát hàm tự tương quan R và thấy rằng (chúng thuộc dạng 1, như vậy có sự khác không [1]). Hệ số tự tương quan biến đổi từ 0,693 (ở 1) đến 0,566 (ở 12) và quay lại 0,807 (ở 12) với các sai số tính toán nhỏ hơn 0,045.

Đã tiến hành sai phân hoá với trễ 12 tháng $D(-12)$, logarit cơ số tự nhiên $Lx(Q_t)$, và sai phân với trễ 1 tháng $D(-1)$. Chuỗi thời gian lại được vẽ và phân tích; các dao động chu kỳ biến mất, còn lại là những dao động răng cưa phức tạp. Phân tích tiếp hàm tự tương quan của chuỗi mới thấy tính phụ thuộc của các thành phần còn lớn. Bằng phương pháp thử dần, đã sai phân tiếp với trễ 1 tháng $Lx(Q_t), D(-12), D(-1)$ và nhận được chuỗi có tính độc lập cao.

3. Xác định tham số mô hình và tổng hợp dòng chảy

Quy trình xác định tham số được tiến hành tuân theo các bước sau:

- Chọn chuỗi dự báo (Q_t),
- Chỉ ra dạng mô hình và số thành phần không mang tính mùa p, q, d ; mang tính mùa P, D, Q .

- Chỉ ra các kiểu biến đổi chuỗi,
- Chọn phương pháp xác định tham số.

Qua nhiều lần thử dần với mục đích là chọn mô hình tối ưu, số tham số ít và ổn định, mức bảo đảm cao, đã chọn:

- Mô hình $ARIMA(1,1,1)(1,1,1)$;
- Ba bước biến đổi: 1- $Ln(Q)$; 2- $D(-12)$; 3- $D(-1)$
- Phương pháp thích hợp tối đa.

Để loại bớt sai số, đã cập nhật số liệu sau mỗi tháng và lặp lại các bước quy trình trên cho mỗi lần dự báo.

Bảng 1. Tổng hợp kết quả xác định tham số

Các đặc trưng thống kê	p (1)	q (1)	Ps (1)	Qs (1)
Giá trị các hệ số	0,44802	0,93856	-0,1371	0,74205
Sai số tính toán	0,05017	0,02007	0,06031	0,04265
Chỉ tiêu Student	8,9031	46,756	-2,273	17,400
Ngưỡng dưới 95%	0,34942	0,89911	-0,2556	0,65824
Ngưỡng trên 95%	0,54661	0,97801	-0,0186	0,82586

Trong bảng: $p(1)$ - hệ số tự hồi quy;

$Ps(1)$ - hệ số tự hồi quy thành phần mùa;

$q(1)$ - hệ số trung bình trượt;

$Qs(1)$ - hệ số trung bình trượt thành phần mùa.

4. Kết quả tính toán và dự báo

Sau các bước định dạng mô hình, đã ghi nhận và tiến hành mô phỏng (dự báo kiểm tra) lại chuỗi lưu lượng nước đến hồ, tiến hành phân tích phần dư (chuỗi sai số). Khả năng mô phỏng của mô hình tương đối tốt. Hệ số tương quan giữa chuỗi mô phỏng và thực đo tương đối lớn, đa số lớn hơn 0,8, có hai tháng (VI, X) của thời kỳ chuyển tiếp mùa (kiệt $\leftrightarrow$ lũ) có giá trị nhỏ ($R_6 = 0,733$; $R_7 = 0,715$).

Phân tích các chỉ tiêu thống kê, thấy rằng đa số các hệ số của mô hình ARIMA tương đối ổn định, sai số tính toán nằm trong giới hạn cho phép. Riêng thành phần trung bình trượt mùa bất ổn định hơn cả, có chỉ tiêu Student bằng -2,273.

Trong dự báo, mô hình ARIMA cho ta 3 trị số là kỳ vọng, ngưỡng trên và ngưỡng dưới ứng với mức đảm bảo $P=95\%$. Phụ thuộc vào số tháng ngoại suy (thời gian dự kiến), các ngưỡng sẽ bị mở rộng dần và sai số tuyệt đối có thể cũng lớn theo. Trong trường hợp dự báo tháng (với thời gian dự kiến là 1 tháng), sai số tuyệt đối của ngưỡng 95% thường nhỏ hơn 20% giá trị kỳ vọng.

Đã đưa ra bộ tham số cùng các số đặc trưng thống kê của chúng ứng với chuỗi dài nhất (I/1956 - III/1995). Để xác định được mức bảo đảm, đã tiến hành dự báo trượt dần cho 25 tháng cuối của chuỗi LLTBT (I/1993 - III/1995). Mức bảo đảm đạt được là 77%. Sai số lớn nhất xảy ra vào tháng VIII/1993 là $1500 \text{ m}^3/\text{s}$.

Trong quá trình nghiên cứu, đã tiến hành dự báo thử nghiệm lưu lượng đến hồ Hoà Bình các tháng trong năm 1995 và cho kết quả tốt, trong 10 trị số dự báo có 8 trị số đúng. Hiện nay mô hình đang được thử nghiệm tiếp.

Tài liệu tham khảo

1. Bùi Văn Đức. Cơ sở vật lý - toán của mô hình ARIMA trong mô phỏng và dự báo lưu lượng tháng. Tập san KHKH Tổng cục KTTV, số 11(419)/1995.
3. Nguyễn Viết Thi. Dự báo dòng chảy sông lớn - Đề tài nghiên cứu của Tổng cục KTTV, 1993.
2. George E.P. Box, and Gwilym M. Jenkin. Time series analysis forecasting and control. University of Lancaster, 1971.
4. STATISTICA, StatSoft, 1994.