

ỨNG DỤNG HỒI QUI TUYẾN TÍNH NHIỀU CHIỀU TRONG DỰ BÁO MỨC NƯỚC HẠN VỪA VÀ DÀI MÙA CẠN TRÊN SÔNG LÔ TẠI THỊ XÃ HÀ GIANG

KS. Nguyễn Thị Nguyệt Hoà

Trung tâm dự báo KTTV tỉnh Hà Giang

I. Đặt vấn đề

Những năm trước đây do chưa được trang bị máy vi tính, các phương án dự báo thủy văn được xây dựng tương đối đơn giản. Các phương án cũ đã dùng để dự báo mực nước trên sông Lô tại thị xã Hà Giang hạn vừa và dài là tra biểu đồ, lập phương trình tương quan đơn giản giữa 2 biến.... Do phương án được xây dựng đơn giản, chưa mô tả được sự ảnh hưởng rất phức tạp của các nhân tố khí tượng thủy văn tới yếu tố cần dự báo. Chính vì vậy phương án có nhiều hạn chế, mang nặng tính chủ quan, sai số dự báo còn lớn.... Đặc biệt trong thời kỳ chuyển tiếp (từ mùa lũ sang mùa cạn và từ mùa cạn sang mùa lũ), mực nước dao động mạnh, chất lượng dự báo thường thấp chưa đạt yêu cầu.

Đến nay Trung tâm dự báo đã được trang bị máy vi tính, việc xây dựng các phương án dự báo mang tính khách quan trở nên dễ dàng hơn bằng cách xây dựng các phương trình dự báo phức tạp, trong đó mô tả rõ sự tác động của từng nhân tố khí tượng thủy văn tới yếu tố dự báo. Phương trình hồi qui tuyến tính nhiều chiều đã đáp ứng được một số các tiêu chuẩn trên. Do vậy, để dự báo mực nước trung bình, cao nhất, thấp nhất tuần và tháng cho sông Lô tại thị xã Hà Giang, chúng tôi đã sử dụng phương trình hồi qui tuyến tính nhiều chiều.

Trong thực tế, các yếu tố khí tượng thủy văn nói chung thường có những tác động qua lại ảnh hưởng lẫn nhau, bởi vậy khái niệm biến độc lập chỉ mang ý nghĩa hình thức. Do vậy có thể xảy ra trường hợp các biến độc lập được chọn để xây dựng phương trình dự báo có tương quan tốt với yếu tố cần dự báo nhưng lại có tương quan tốt với nhau. Điều này dẫn đến hậu quả là mặc dù phương trình hồi qui đã phức tạp do sự có mặt của nhiều biến độc lập nhưng độ chính xác lại giảm hơn do sai số quan trắc, sai số tính toán.... Chính vì vậy khi xây dựng phương trình dự báo cần lựa chọn kỹ từng nhân tố khí tượng thủy văn, xác định được nhân tố nào cần có mặt trong phương trình dự báo và nhân tố nào không cần thiết có mặt. Để đạt được điều này khi xây dựng phương trình dự báo tác giả đã dùng phương trình hồi qui từng bước.

Qua nghiên cứu và tính toán cho thấy mực nước tại trạm thủy văn thị xã Hà Giang không những có mối quan hệ tốt với mực nước các tuần (hoặc tháng liên kế) mà còn có liên quan khá tốt với lượng mưa, nhiệt độ của trạm khí tượng thị xã Hà Giang. Nói cách khác, mực nước cần dự báo là một hàm của các biến: lượng mưa, nhiệt độ, mực nước... của tuần (hoặc tháng) trước đó.

II. Nội dung

Phương trình hồi qui tuyến tính nhiều chiều để dự báo mực nước có dạng

$$Y = a_0 + a_1 X_1 + a_2 X_2 + \dots + a_m X_m$$

trong đó:

Y- yếu tố thủy văn cần dự báo,

$a_0, a_1 \dots a_m$ - các hệ số của phương trình hồi qui,

$X_1, X_2 \dots X_m$ - các nhân tố khí tượng thủy văn có tương quan với yếu

tố cần dự báo.

Với từng yếu tố mực nước cần dự báo (mực nước đặc trưng tuần, tháng, mực nước thấp nhất mùa cạn...), tùy theo mức độ tương quan của chúng với các nhân tố khí tượng thủy văn mà quyết định số chiều hồi qui của phương trình dự báo.

Để xây dựng một phương trình dự báo mực nước bằng hồi qui tuyến tính nhiều chiều theo phương pháp hồi qui từng bước với mỗi yếu tố dự báo gồm các bước sau:

Bước 1

Thống kê chuỗi số liệu mực nước, lượng mưa, nhiệt độ... tuần, tháng của trạm khí tượng và thủy văn thị xã Hà Giang có mối quan hệ với yếu tố mực nước cần dự báo.

Riêng với yếu tố cần dự báo: tính các đặc trưng thống kê và sai số cho phép (S_{cf}) của chuỗi số liệu đã thống kê.

Bước 2

Tính các hệ số tương quan giữa yếu tố mực nước cần dự báo với các nhân tố khí tượng thủy văn và giữa các nhân tố khí tượng thủy văn với nhau đã thống kê ở bước 1.

Bước 3

- Chọn hệ số tương quan giữa yếu tố dự báo và nhân tố dự báo có giá trị tuyệt đối lớn nhất. Giả sử nhân tố X_1 thỏa mãn điều kiện trên. Khi đó X_1 là nhân tố có tác động chính lên yếu tố cần dự báo Y . Ta lập phương trình hồi qui:

$$Y_{(1)} = a_{0(1)} + a_{1(1)} X_1$$

Tính sai số tương quan của phương trình dự báo $S_{(1)}$.

- Chọn tiếp hệ số tương quan có giá trị tuyệt đối lớn nhất trong các nhân tố còn lại. Nếu có 2 hệ số xấp xỉ bằng nhau thì ưu tiên chọn nhân tố có hệ số tương quan (với nhân tố đã được chọn trước) nhỏ hơn để tăng tính độc lập của nhân tố được chọn. Giả sử X_2 được chọn. Lập phương trình hồi qui dạng $Y(2) = a_0(2) + a_1(2) X_1 + a_2(2) X_2$. Tính sai số tương ứng của phương trình $S_{(2)}$. Lúc này ta có phương trình hồi qui 2 biến mà độ chính xác của nó được đánh giá bằng $S_{(2)}$.

- So sánh $S_{(1)}$ với $S_{(2)}$.

Nếu $\left| \frac{S^{(2)} - S^{(1)}}{S_{(2)}} \right| < \epsilon$ thì X_2 sẽ bị bỏ qua và một nhân tố khác trong các nhân tố còn lại sẽ được lựa chọn để tính như bắt đầu bước tính toán chọn X_2 .

Nếu $\left| \frac{S^{(2)} - S^{(1)}}{S_{(2)}} \right| \geq \epsilon$ thì X_2 sẽ được chọn. Quá trình tính toán được tiếp tục: chọn nhân tố $X_3 \dots$ cho đến khi tất cả các nhân tố đã được chọn vào phương trình (hoặc bị loại) như đã trình bày ở trên.

Như vậy, ở bước thứ K ta có sai số $S^{(k)}$ tương ứng với phương trình hồi qui

$$Y^{(k)} = a_0^{(k)} + a_1 X_1 + \dots + a_k^{(k)} X_k$$

và điều kiện để lựa chọn là:
$$\left| \frac{S^{(k)} - S^{(k-1)}}{S^{(k)}} \right| < \varepsilon.$$

Ở đây ε là một số dương tùy ý để đánh giá mức độ đóng góp thông tin của mỗi nhân tố mới (hay mỗi biến) đối với sai số. Nếu mức độ giảm sai số nhỏ hơn ε thì có thể bỏ qua nó.

- Trong thực tế khi tính toán ε được chọn phù hợp với từng yếu tố dự báo. Với mỗi yếu tố cần dự báo tương ứng có một sai số cho phép được tính theo công thức:

$$Scf = 0,674 \sigma_2 = 0,674 \sqrt{\frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{n-1}}$$

Các phương trình hồi qui được xây dựng cần phải thỏa mãn các điều kiện sau: sai số phương trình hồi qui phải nhỏ hơn sai số dự báo cho phép và mức đảm bảo của phương án $P \geq 75\%$.

III. Kết quả

1. Phương trình dự báo mực nước tháng

Với cách tính toán như trên chúng tôi đã xây dựng được 6 phương trình dự báo mực nước thấp nhất tháng mùa cạn như sau:

$$H_{min11db} = 613,2 + 0,415H_{tb10} - 0,0017R_6 + 0,00041R_{8,9,10}$$

$$H_{min12db} = 261,1 + 0,733H_{min11} - 0,00013R_6 + 0,00105R_{8,9,10}$$

$$H_{min1db} = 267,3 + 0,75H_{min12} - 0,00058R_{5,6,7} - 0,00089R_8$$

$$H_{min2db} = 115,4 + 0,888H_{min1} - 0,00056R_{5,6,7}$$

$$H_{min3db} = 221,3 + 0,784H_{min2} - 0,00025R_{5,6,7}$$

$$H_{min4db} = 295,2 + 0,71H_{min3} - 0,00147R_7 + 0,0021R_8$$

Trong đó:

$H_{min11db}, H_{min12db}, \dots, H_{min4db}$ - mực nước thấp nhất tháng 11, 12, ... 4 dự báo.

$H_{min11}, H_{min12}, \dots, H_{min4}$ - mực nước thấp nhất tháng 11, 12, ... 4.

H_{tb10} - mực nước trung bình tháng 10.

R_6, R_7, R_8 - lượng mưa tháng 6, 7, 8.

$R_{6,7,8}$ và $R_{8,9,10}$ là tổng lượng mưa của các tháng 6, 7, 8 và 8, 9, 10.

Tương tự, có 6 phương trình hồi qui dự báo mực nước trung bình tháng các tháng mùa cạn như sau:

$$H_{tb11db} = 658,3 + 0,407H_{tb10} - 0,0078R_6 + 0,00315R_{8,9,10}$$

$$H_{tb12db} = 674,8 + 0,362H_{tb10} - 0,0045R_6 + 0,00112R_{8,9,10}$$

$$H_{tb1db} = 256,8 + 0,743H_{tb12} + 0,00049R_{8,9,10}$$

$$H_{tb2db} = 90,7 + 0,90H_{tb1} + 0,00039R_{5,6,7}$$

$$H_{tb3db} = 64,4 + 0,93H_{tb2} - 0,00058R_{5,6,7} + 0,00255R_8$$

$$H_{tb4db} = -95,7 + 1,10H_{tb3} - 0,0024R_7 + 0,0025R_8$$

Trong đó:

$H_{tb11db}, H_{tb12db}, \dots, H_{tb4db}$ - mực nước trung bình tháng 11, 12, ... 4 dự báo.

$H_{tb11}, H_{tb12}, \dots, H_{tb4}$ - mực nước trung bình tháng 11, 12, ... 4.

Đặc biệt có thể sử dụng phương trình hồi qui tuyến tính để nhận định mực nước thấp nhất của tất cả các tháng mùa cạn tới ngay sau khi mùa lũ kết thúc (đầu tháng 11) cho kết quả khá tốt. Qua đó ta dễ dàng nhận định được mực nước thấp nhất mùa cạn và tháng xảy ra trị số kiệt nhất. Để nhận định mực nước tháng thấp nhất cho mùa cạn chúng tôi đã sử dụng 5 phương trình sau:

$$H_{12ND} = 683,1 + 0,3292H_{tb10} - 0,0018R_6 + 0,0038R_8 - 0,00135R_9.$$

$$H_{1ND} = 787,8 + 0,2426H_{tb10} - 0,0012R_6 - 0,0014R_7 + 0,002R_8 - 0,0016R_{9,10}.$$

$$H_{2ND} = 859,2 + 0,0599H_{tb9} + 0,1159H_{tb10} - 0,0011R_6 - 0,0025R_7 - 0,0007R_8$$

$$H_{3ND} = 433,3 + 0,1253H_{10} + 1,142T_6 + 0,561T_9 + 0,072T_{10} - 0,0036R_7.$$

$$H_{4ND} = 541,6 + 0,062H_{tb10} + 1,116T_6 + 0,515T_9 - 0,031T_{10} - 0,004R_7 + 0,0025R_6.$$

$H_{12ND}, H_{1ND}, \dots, H_{4ND}$ - mực nước thấp nhất tháng 12, 1, ..., 4 nhận định.

2. Phương trình dự báo mực nước tuần mùa cạn

Để dự báo mực nước cao nhất và thấp nhất tuần trong 6 tháng mùa cạn, tác giả đã xây dựng được 36 phương trình hồi qui tuyến tính nhiều chiều (18 phương trình dự báo cao nhất và 18 phương trình cho mực nước thấp nhất).

Các phương trình dự báo mực nước cao nhất tuần:

$$H_{\max XI(1)db} = 580,1 + 0,439H_{\max X(3)} + 0,0161R_8 - 0,0082R_9 + 0,0055R_{10},$$

$$H_{\max XI(2)db} = 1030,6 - 0,0322R_6 + 0,0282R_9 - 2,022T_6 + 2,77T_8,$$

$$H_{\max XI(3)db} = 225,2 + 0,426H_{\max XI(2)} - 0,006R_6 + 0,0012R_9 + 2,457T_6 - 0,932T_8.$$

$$H_{\max XII(1)db} = 215,6 + 0,2413H_{\max XI(3)} - 0,018R_5 - 0,00144R + 1,160T_6 + 1,304T_9.$$

$$H_{\max XII(2)db} = 1006,1 + 0,296H_{\max XII(1)} - 0,0007R_6 + 0,0046R_7 + 1,667T_6 - 2,575T_7.$$

$$H_{\max XII(3)db} = -49,1 + 0,688H_{\max XII(2)} + 1,692T_6 + 0,365T_{10},$$

$$H_{\max I(1)db} = 472,5 + 0,575H_{\max XII(3)} - 0,0044R_6 + 0,010R_{12}$$

$$H_{\max I(2)db} = 361,7 + 0,67H_{\max I(1)} - 0,0013R_6$$

$$H_{\max I(3)db} = 147,2 + 0,858H_{\max I(1)} + 0,0027R_{12}$$

$$H_{\max II(1)} = -53,9 + 0,859 H_{\max I(3)} + 0,619T_6 - 0,008T_{10} + 0,204T_{12}$$

$$H_{\max II(2)} = 283,1 + 0,425 H_{\max II(1)} + 0,019T_{12} - 0,949T_5 + 0,292T_6$$

$$H_{\max II(3)} = -210,9 + 1,030H_{\max II(2)} + 0,405T_6 + 0,241T_{10}$$

$$H_{\max III(1)db} = 163,1 + 0,834H_{\max II(3)} - 0,265T_6 + 0,366T_{10}$$

$$H_{\max III(2)db} = -12,0 + 0,956H_{\max III(1)} - 0,146T_6 + 0,487T_{11}$$

$$H_{\max III(3)db} = -349,6 + 0,88H_{\max III(2)} + 1,822T_6 - 0,038T_{10}$$

$$H_{\max IV(1)db} = 561,5 + 0,419H_{\max III(2)} - 0,0078R_7 + 0,037R_3 - 0,126T_6 + 0,632T_{10}$$

$$H_{\max IV(2)db} = 1049,4 + 0,378H_{\max IV(1)} + 0,0091R_8 + 0,024R_3 - 1,514T_9$$

$$H_{\max IV(3)db} = -417,6 + 0,367H_{\max IV(2)} + 0,0091R_8 + 0,016R_3 + 3,236T_{10} + 1,592T_{11}.$$

Trong đó:

$H_{\max XI(1)}, \dots, H_{\max IV(3)}$ - mực nước cao nhất tuần 1 tháng XI, ..., tuần 3 tháng IV (chữ số La Mã chỉ tháng, chữ số Ả-rập trong ngoặc đơn chỉ tuần).

$H_{\max XII(1)}, \dots, H_{\max IV(3)}$ - mực nước cao nhất dự báo tuần 1 tháng XII, ..., tuần 3 tháng IV.

Tương tự, đã xây dựng được 18 phương trình dự báo mực nước thấp nhất tuần mùa cạn (không kể ra ở đây).

Ghi chú: Đơn vị tính trong tất cả các phương trình: lượng mưa được tính đến 0,1mm; nhiệt độ: 0,1°C; mực nước: cm.

Mực nước được tính theo cao độ giả định trước năm 1998 (như ghi trong sổ quan trắc).

3. Mức bảo đảm của phương án

Mức bảo đảm của phương án $P = n/N$.

n - số lần dự báo đúng (có sai số dự báo nhỏ hơn hoặc bằng sai số cho phép),

N - tổng số lần dự báo.

Mức đảm bảo của phương án dự báo tháng (%)

Tháng Yếu tố	Tháng 11	Tháng 12	Tháng 1	Tháng 2	Tháng 3	Tháng 4
Hmin	82,8	91,7	91,4	100	94,3	91,2
Htb	76,7	77,4	96,6	96,7	86,7	86,2

Mức bảo đảm của phương án nhận định mực nước thấp nhất mùa cạn (%)

Tháng Yếu tố	Tháng 12	Tháng 1	Tháng 2	Tháng 3	Tháng 4
Hmin	78,4	73,0	77,1	73,0	78,4

Mức bảo đảm của phương án dự báo tuân theo tính toán đều đạt trên 75% (không kể ra ở đây).

IV. Kết luận

Xây dựng phương án dự báo mực nước hạn vừa và dài bằng phương trình hồi qui tuyến tính nhiều biến qua nhiều bước tính toán nên tương đối phức tạp và mất khá nhiều thời gian. Tuy nhiên, với sự trợ giúp của máy tính chúng ta hoàn toàn có thể thực hiện được.

Phương án dự báo này mang tính khách quan cao, chất lượng dự báo khá tốt. Sau khi đã có phương trình dự báo cho từng yếu tố việc dự báo trở nên rất dễ dàng, tiện lợi, nhanh chóng. Trong những tháng gần đây trong dự báo hạn vừa và hạn dài, tác giả đã dùng phương trình hồi qui tuyến tính nhiều chiều, chất lượng dự báo thủy văn đã được nâng lên đáng kể.

Tuy nhiên, một số tuần, tháng mực nước có biến động quá mạnh so với TBNN của cùng thời kỳ thì sai số dự báo còn lớn.

Tài liệu tham khảo

1. Bùi Văn Đức. Các phương pháp thống kê trong dự báo KTTV.- Hướng dẫn dự báo khí tượng thủy văn ở các trạm tỉnh (tài liệu cho lớp bồi dưỡng dự báo viên KTTV địa phương, 1994).
2. Phan Văn Tân. Phương pháp thống kê trong khí hậu.- Trường Đại học khoa học tự nhiên - Đại học quốc gia, Hà Nội (11-1997).