CƠ SỞ VẬT LÝ - TOÁN CỦA MÔ HÌNH ARIMA TRONG MÔ PHỎNG VÀ DỰ BÁO LƯU LƯỢNG THÁNG

PTS. Bùi Văn Đức
Trung tâm quốc gia dự báo KTTV

Trong dự báo thủy văn hạn vừa và hạn dài, các yếu tố dự báo được hình thành dưới sự ảnh hưởng phức tạp của nhiều nhân tố khác nhau. Trong điều kiện hạn chế về các thông tin dự báo hiện nay, thì các mô hình thống kê nói chung và đặc biệt là các phương pháp phân tích chuỗi thời gian (PTCTG) đang chiếm ưu thế. Bài này giới thiệu mô hình ARIMA – một trong các mô hình thuộc nhóm PTCTG có nhiều khả quan trong dự báo khí tượng thủy văn.

1. GIỚI THIỆU CHUNG

Cơ sở ứng dụng của mô hình ARIMA (Autoregressive – Integrated – Moving – Average) dựa trên giả thiết: ảnh hưởng của các nhân tố chủ yếu xác định xu thế biến đổi của chuỗi thời gian vẫn được duy trì trong thời kỳ dự báo. Mô hình ARIMA sử dụng các thông tin chứa trong các thành phần đã biết của chuỗi quá khứ để dự báo quá trình này trong tương lai. Tác giả của mô hình ARIMA là Box và Jenkin. Những công trình ứng dụng đầu tiên của mô hình ARIMA được công bố vào những năm 1970.

*** Phương trình tổng quát**

Mô hình ARIMA (p, d, q) có thể viết dưới dạng tổng quát sau:

$$Y_t^d = \sum_{i=1}^{p} a_i Y_{t-1}^d + \sum_{J=1}^{q} b_i \varepsilon_{t-j} + \varepsilon_t, \qquad (1)$$

Trong đó:

* p, d, q là các số nguyên, dương.

* a_i và b_i là các hằng số

* $Y_t^d = Y_t^{d-1} - Y_{t-1}^{d-1}$ ($Y_t^1 = Y_t - Y_{t-1}$; $Y_t^2 = DY_t - DY_{t-1}$; ....).

Tổng thứ nhất AR(P) là thành phần tự hồi quy, tổng thứ 2 MA (q) – thành phần trung bình trượt và ε_t – sai số ngẫu nhiên.

2. CƠ SỞ VẬT LÝ

2.1. Mô hình AR (1)

Mô hình tự tương quan AR(P) mô tả quan hệ giữa giá trị phụ thuộc Xt với các biến độc lập Y_{t-1}, Y_{t-2},...Y_{t-p} hoặc chuỗi chuẩn hóa Z_{t-1}, Z_{t-2},...Z_{t-p} và được viết dưới dạng tổng quát sau:

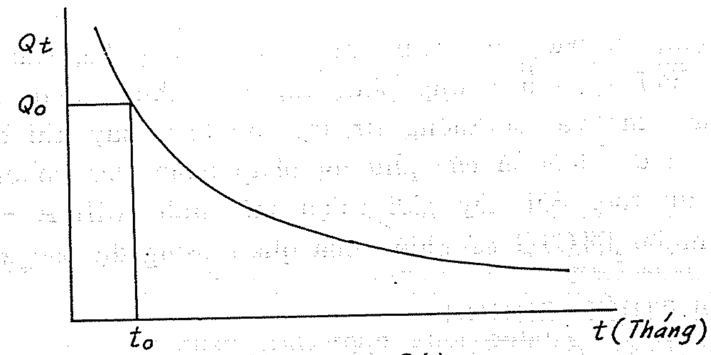
$$Z_t = \sum_{j=1}^{p} \phi_j z_{t-j} + \varepsilon_t \qquad (2)$$

trong đó ϕ là hệ số tự hồi quy hay hệ số tương quan riêng phần: εt là sai số ngẫu nhiên (White noise); p – số thành phần tự hồi quy.

Trong những tháng mùa khô, dòng chảy tháng tiếp theo Q_t chỉ phụ thuộc vào lượng trữ nước trong lưu vực, thông qua chỉ số lượng trữ là dòng chảy tháng đầu mùa khô Q_o. Quan hệ này trong thủy văn được mô phỏng bằng đường nước rút (hình 1) hoặc hàm

$$Q_t = Q_o * EXP(-Kt) \qquad (3)$$

Trong đó K là tham số: t là thứ tự tháng tính từ đầu mùa.



Hình 1. Đường nước rút Q(t)

Nếu chuỗi dòng chảy tháng Q_i được mô phỏng bằng mô hình AR(1)

$$Q_t - Q_{tb} = \phi(Q_t - 1 - Q_{tb}) + \varepsilon_t$$

Khai triển và đổi vế phương trình trên ta được

$$Q_t = \phi * Qtb + \phi * Qt-1 - \phi * Qtb + \varepsilon t$$

Vậy ta có $$Q_t = \phi * Qt - 1 + \varepsilon t \qquad (4)$$

Như vậy, không khó khăn để nhận ra rằng, khi (t=1) ϕ, của (4) chính là EXP (–k).

$$\phi = EXP(-k)$$

2.2. Mô hình ARIMA (p,d,q)

Trong những tháng có lượng mưa đáng kể, dòng chảy tháng được hình thành từ hai thành phần (Ngầm và mặt) (Hình 2).

$$Zt = c\ St - 1 + d\ Xt \qquad (5)$$

Tương tự ta có

$$Zt - 1 = c\ St - 2 + d\ Xt - 1$$

Từ đây suy ra

$$St - 2 = \frac{Z_{t-1} - dX_{t-1}}{c} \qquad (6)$$

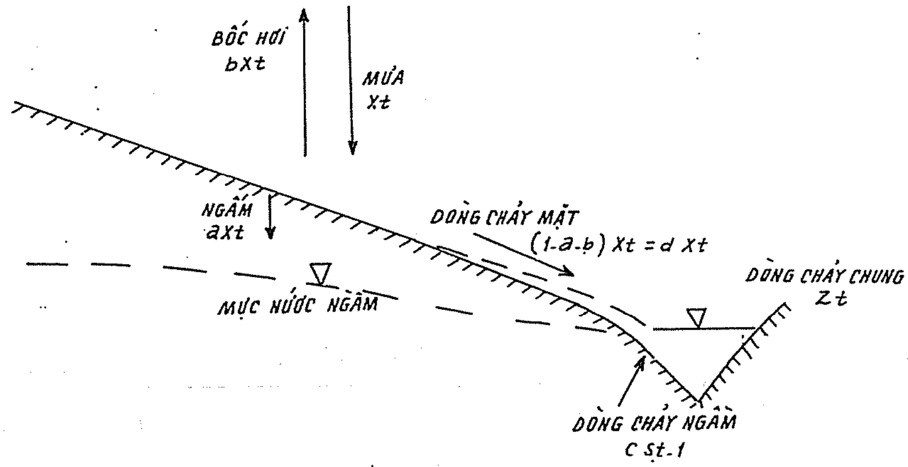
Phương trình cân bằng lượng trữ ngầm St được viết như sau:

$$St = St-1 + a\ Xt - c\ St-1$$

$$St = (1-c)\ St-1 + a\ Xt \quad (7)$$

Tương tự ta có

$$St-1 = (1-c)\ St-2 + a\ Xt-1 \quad (8)$$



Hình 2. Sơ đồ hình thành dòng chảy

Tổng hợp từ các công thức (5 – 8) ta có

$$Zt = (1-c)Zt-1 + d\ Xt-[d(1-c)-ac]\ Xt-1 \quad (9)$$

Đặt $\phi = (1-c)$ và $\Theta = [d(1-c)-ac]$ và d, Xt = εt ta có mô hình ARMA (1,1).

Cũng tương tự như vậy có thể gán ý nghĩa vật lý cho các hệ số của mô hình ARMA (p,q), và khi lấy sai phân (bậc d=2) với thời gian trễ 1 tháng, 12 tháng ta có mô hình ARIMA (p,2,q).

3. XÁC ĐỊNH THAM SỐ MÔ HÌNH

3.1. Chiến lược xác định p,d,q

Các tham số p, d, q đều là các số nguyên dương, chúng được xác định qua phân tích quá trình biến đổi hàm tự tương quan r_k và hàm tự tương quan riêng phần ϕK.

Xác định d đầu tiên và các bước tiến hành như sau:

– Tính và vẽ quá trình r_k k = 1,2, ..(Hình 3)

Tồn tại 3 loại quá trình r ứng với các hình 3a, 3b, 3c.

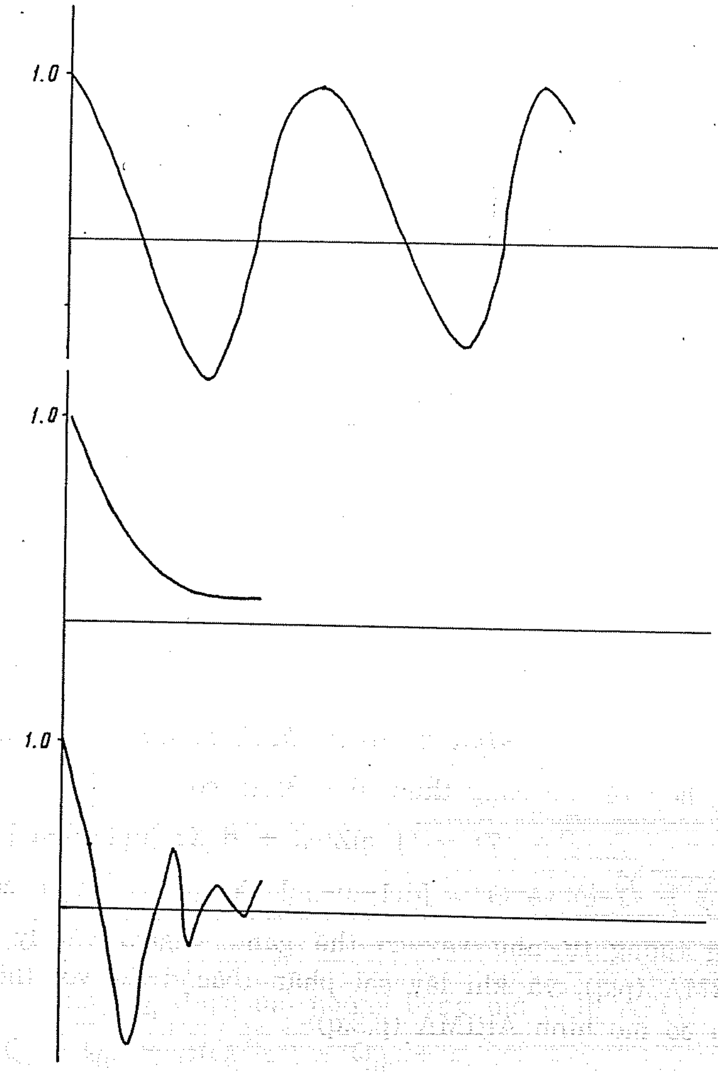
Khi rơi vào trường hợp b hoặc c ta lấy d=0. Rơi vào trường hợp a, sau sai phân bậc 1 với trễ k_1, ta lại tính và vẽ quá trình r mới của chuỗi sai phân bậc 1. Nếu lại có quá trình r mới giống dạng a, lại tiến hành sai phân lần nữa. Tiến hành sai phân tiếp tục tới khi quá trình r mới nhất không còn có dạng a.

Xác định p,q:

– Số thành phần tự hồi quy được xét theo quá trình r cuối cùng ở phần xác định d. Sẽ lấy tới p mà ở đó có r lớn hơn hoặc bằng sai số của nó.

– Số thành phần trung bình trượt sẽ lấy đến khi $r_q \neq 0$, $r_{q+1}, r_{q+2}, \ldots = 0$.

Trong thực tế mô phỏng và dự báo các chuỗi thủy văn, giá trị các hằng p, d, q thường không vượt quá 2 [6].



Hình 3. Các kiểu quá trình điển hình của r_k

3.2. Xác định hệ số tự hồi quy (hệ số tương quan riêng phần)

Xuất phát từ phương trình Yule – Walker [7]:

$$\varsigma_k = \phi_1 \varsigma_{k-1} + \phi_2 \varsigma_{k-2} + \ldots + \phi_p \varsigma_{k-p},\ k > 0 \text{ và } k \geq p.$$

Viết (10) tuần tự cho $\varsigma_1, \varsigma_2, \ldots \varsigma_k$, ta có hệ phương trình trong đó ẩn số là ϕj (j = 1,2,...k) và viết dưới dạng ma trận ta được:

$$\underbrace{\begin{bmatrix} 1 & \varsigma_1 & \varsigma_2 & \ldots & \varsigma_{k-1} \\ \varsigma_1 & 1 & \varsigma_1 & \ldots & \varsigma_{k-2} \\ \varsigma_2 & \varsigma_1 & 1 & \ldots & \varsigma_{k-3} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \varsigma_{k-1} & \varsigma_{k-2} & \varsigma_{k-3} & \ldots & 1 \end{bmatrix}}_{P} \times \underbrace{\begin{bmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \\ \ldots \\ \phi_k \end{bmatrix}}_{\varphi} = \underbrace{\begin{bmatrix} \varsigma_1 \\ \varsigma_2 \\ \varsigma_3 \\ \ldots \\ \varsigma_k \end{bmatrix}}_{\varsigma} \qquad (11)$$

Hoặc dạng rút gọn

$$P\phi = \varsigma$$

$$\phi = P^{-1}\varsigma$$

3.3. Xác định hệ số trung bình trượt (Θ)

3.3. Xác định hệ số trung bình trượt (Θ)

Sau khi ước lượng các tham số ϕ ở bước trên, nếu Z không có trung bình bằng 0, thì cần xác định Θ_{oo}

$$\Theta_{oo} = Z\ (1- \sum_{j=1}^{p}\phi_j)$$

Các hệ số trung bình trượt Θ được xác định bằng phương pháp lặp theo các công thức của Yule – Walker với ước lượng ban đầu $\Theta_j = 0$; j = 1, ..., q.

$$\sigma_e^2 = \frac{Co}{1+\Theta_1^{-2}+...+\Theta_q^{-2}} \tag{12}$$

$$\overline{\Theta}i = - \left[\frac{C'_j}{\sigma_e^{-2}} - \overline{\Theta}_1\overline{\Theta}_{j-1} - \overline{\Theta}_2\overline{\Theta}_{j-2} - ... - \overline{\Theta}_{q-1}\overline{\Theta}_{j-q}\right] \tag{13}$$

Ở đây

$$C'_j = \sum_{i=o}^{p}\phi_i^2 C_j + \sum_{i=1}^{p}(\phi_o\phi_i + \phi_1\phi_{i+1} + \phi_{p-1}\phi_p)xd_j \tag{14}$$

Trong đó $d_j = C_{j+i} + C_{j-i}$

Phần dư $\varepsilon_{p+j} = \Theta_{oo} - \sum_{i=1}^{p}\phi_i Z_{p+j-i} + \sum_{i=1}^{q}\Theta_i\varepsilon_{p+j-i}$

Chỉ tiêu chất lượng $S = \sum_{t=i}^{N}\varepsilon_t^2$

Qúa trình lặp sẽ dừng khi S đạt giá trị mininum.

4. MÔ PHỎNG CHUỖI BẰNG MÔ HÌNH ARIMA

Với các tham số đã được xác định ở các bước trên, chuỗi Zt được mô phỏng lại theo công thức sau:

$$Z_t = \Theta_{oo} + \sum_{i=1}^{p}\phi_i Z_{p+j-i} + \varepsilon_t + \sum_{i=1}^{q}\Theta_{ip+j-i}$$

Chuỗi số liệu gốc Yt sẽ được tính ngược từ chuỗi Zt và các tham số hằng số Y_{tb} và σ_y.

– Nếu Z_t là chuỗi lệch chuẩn

$$Zt = Yt - Ytb,$$

thì $$Yt = Ytb + Zt$$

– Nếu Zt là chuỗi chuẩn (0,1)

$$Zt = \frac{Y_t - Y_{tb}}{\sigma_y}$$

thì $$Y_t = Y_{tb} + \sigma_y Z_t$$

– Nếu chuỗi Z_t được biến đổi từ chuỗi Y_t mang chu kỳ năm (là = 12 tháng)

$$Z_t = \frac{Y_{i,j} - Y_{tbj}}{\sigma_j}$$

thì $$Y_t = Y_{tbj} + \sigma_j Z_t \ (t = (n - i)^* 12 + j).$$

Ở đây i là số năm: j – chỉ số tháng; n là số năm.

Vùng tin cậy

– Phương sai của ước lượng trung bình Ytb được tính theo công thức sau:

$$Var(Y_{tb}) = \frac{\sigma^{-2}}{N} \left[1 + \frac{2}{N} \sum_{k=1}^{N-1} (N-k)\varsigma_k\right]$$

$$S(Ytb) = [Var(ytb)]$$

Vùng tin cậy của Ytb sẽ là:

$$[Ytbb - t(N-1)_{1-}\alpha/2\ S(Ytb),\ Ytb + t(N-1)_{1-}\alpha/2\ S(Ytb)]$$

Ở đây $t(N-1)_{1-\alpha/2}$ là 1 – $\alpha/2$ "quantile" của phân bố t với N–1 bậc tự do.

$$V(\phi) = (N-p)^{-1} \ (1-\sum_{j=1}^{p}\varphi_{jr}) \begin{bmatrix} 1 & \varsigma_1 & \varsigma_2 & \dots & \varsigma_{p-1} \\ \varsigma_1 & 1 & \varsigma_1 & \dots & \varsigma_{p-2} \\ \varsigma_2 & \varsigma_1 & 1 & \dots & \varsigma_{p-3} \\ \dots & \dots & \dots & \dots & \dots \\ \varsigma_{p-1} & \varsigma_{p-2} & \varsigma_{p-3} & \dots & 1 \end{bmatrix}$$

Vùng tin cậy của ϕ_i sẽ là:

$$[\phi j - U_{1-}\alpha/2S(\phi), \qquad \phi j - U_{1-}\alpha/2S(\phi_j)]$$

Ở đây $U_{1-\alpha/2}$ là 1 – $\alpha/2$ quantile của phân bố chuẩn.

5. VÀI NÉT VỀ KHẢ NĂNG ỨNG DỤNG

Mô hình ARIMA thích hợp hơn cả trong dự báo các chuỗi tự nhiên mang tính vĩ mô về mặt thời gian và không gian, thể hiện mạnh mẽ tính chu kỳ và quán tính. Cho tới nay mô hình ARIMA đã được ứng dụng trong nhiều lĩnh vực khác nhau, như dự báo số hành khách đi máy bay, biến đổi giá cả một số mặt hàng, v.v... [3,6].

Mô hình ARIMA cũng được đặc biệt quan tâm trong dự báo một số yếu tố khí tượng thủy văn [1, 2, 4...7]. Gần đây nhất, nhóm tác giả Trung tâm quốc gia dự báo KTTV và Khoa địa lý địa chất Trường Đại học Tổng hợp Hà Nội, đã tiến hành nghiên cứu ứng dụng mô hình ARIMA cho dự báo lưu lượng nước trung bình tháng đến hồ Hòa Bình.

TÀI LIỆU THAM KHẢO

1. Alièkhin U.M. Dự báo bằng phương pháp động lực thống kê. Tuyển tập báo cáo tại hội nghị thủy văn toàn Liên Xô. Tập 7, Dự báo thủy văn, NXB KTTV, Leningrat 1976.

2. Đặng Văn Bảng. Mô hình toán thủy văn và phân tích hệ thống. Tập bài giảng chuyên đề sau đại học. Trường Đại học thủy lợi. Tháng VI/1991.

3. George E.P. Box, and Gwilym M. Jenkin. Time series analysis forecasting and control. University of lancaster, 1971.

4. Nguyễn Viết Thi. Dự báo dòng chảy sông lớn – đề tài nghiên cứu của Tổng cục KTTV, 1993.

5. STATISTICA, StatSoft, 1994.

6. O Connell, P.E. ARIMA Models in Synthetic Hydrology. In Mathematical Models for Surface Hydrology, T. A. Ciriani, U. Maione and J. R. Wallis, Editors, Wiley, New York. 1977.

7. Sales J.D., Delleur J. W., Yevjevich V. and Lane W. L. Applied Modelling of Hydrological Time Series.