



PHÂN TÍCH THỐNG KÊ ĐỈNH LŨ TẠI TRẠM ĐO CHÂU ĐỐC TRÊN SÔNG HẬU

Võ Văn Tài¹

¹ Khoa Khoa học Tự nhiên, Trường Đại học Cần Thơ

Thông tin chung:

Ngày nhận: 14/08/2014

Ngày chấp nhận: 30/10/2014

Title:

Statistical analysis crest of a flood at Chau Doc's measuring station on Hau river

Từ khóa:

Chuỗi thời gian, chuỗi thời gian mờ, dự báo, chu kỳ, tiêu chuẩn AIC

Keywords:

Time series, model, forecast, cycle, fuzzy

ABSTRACT

Using historical data crest of flood at Chau Doc's measuring station on Hau river (CFCDHR) and basing on the autocorrelation coefficient and spectrum analysis, this study considered its cycle. Different models of time series and fuzzy time series were built from the original data, smooth data and fuzzy data of CFCDHR. By using statistical criterions in order to find the most appropriate model, the CFCDHR to 2020 was then forecasted.

TÓM TẮT

Sử dụng số liệu quá khứ đỉnh lũ tại trạm đo Châu Đốc trên sông Hậu (CFCDHR), dựa vào hệ số tự tương quan, phân tích phổ, bài viết xem xét chu kỳ của nó. Từ số liệu gốc, số liệu làm trơn, số liệu mờ hóa của CFCDHR, bài viết xây dựng các mô hình khác nhau của chuỗi thời gian và chuỗi thời gian mờ. Sử dụng các tiêu chuẩn thống kê để tìm mô hình thích hợp nhất, từ đó chúng tôi tiến hành dự báo CFCDHR đến năm 2020.

1 GIỚI THIỆU

Đồng bằng sông Cửu Long (ĐBSCL) là vựa lúa của cả nước, vùng sản xuất trái cây quan trọng bậc nhất của quốc gia. Với địa hình bằng phẳng, nhiều vùng trũng, có nhiều sông ngòi chằng chịt, hằng năm ĐBSCL luôn có một mùa lũ từ tháng 7 đến tháng 11. Lũ mang đến nhiều lợi ích cũng như gây ra nhiều tác hại về người và của cho vùng. Có thể nói lịch sử phát triển kinh tế xã hội của vùng ĐBSCL gắn liền với lịch sử phòng chống lũ lụt của người dân. Ngày nay, việc đối phó với lũ gắn liền với các cụm từ: “kiểm soát lũ”, “sống chung với lũ”, tuy nhiên phương cách để thực hiện các phương châm đó còn nhiều bản cãi khác nhau. Để đối phó với lũ theo cách nào, chúng ta đều cần phải dự báo được chế độ thủy văn hai con sông chính trong vùng: Sông Hậu và sông Tiền.

Phân tích thống kê đỉnh lũ tại trạm đầu nguồn sông Hậu là cơ sở quan trọng để đưa ra những phân

tích tại các vị trí khác trong vùng. Trong phân tích thống kê việc dự báo đỉnh lũ, cũng như xác định chu kỳ của nó là những công việc quan trọng nhất. Nó là cơ sở khoa học cần phải có trong qui hoạch xây dựng và phát triển hệ thống thủy lợi, sản xuất nông nghiệp. Nó cũng là cơ sở để mỗi địa phương làm căn cứ trong xây dựng cụm tuyến dân cư, phát triển đô thị,... Chúng ta đã chứng kiến sự lãng phí và hậu quả vô cùng tai hại của những công trình xây dựng do không dự báo được đỉnh lũ của các địa phương trong vùng các năm qua. Nghiên cứu về đỉnh lũ ĐBSCL từ lâu đã được Phân viện khảo sát huy quách Nam Bộ, Đài khí tượng thủy văn khu vực Nam Bộ và nhiều nhà khoa học thực hiện trong các giai đoạn 1996 – 2000, 2000 – 2004, các báo cáo về lũ lớn 1984, 1991, 1996, 1997, 2000,... và rất nhiều báo cáo thường niên khác. Các nghiên cứu này được tổng kết trong [13]. Những nghiên cứu trên cũng đã thực hiện dự báo đỉnh lũ, tuy nhiên thực tế cho thấy các kết quả này cũng chỉ là

thông tin mang tính chất tham khảo. Việc dự báo đỉnh lũ cũng như xác định chu kỳ của nó vẫn là bài toán khó chưa có lời giải cuối cùng. Các phương pháp thống kê hiện đại trong xác định chu kỳ, dự báo chuỗi số liệu thời gian đã được áp dụng tương đối tốt cho các bộ số liệu khác nhau trong các nghiên cứu [2], [3], [4], ...[9] nhưng chưa được áp dụng ở nước ta.

ĐBSCL là một trong những nơi chịu ảnh hưởng nặng nề nhất của biến đổi khí hậu. Cùng với các nhà khoa học trên thế giới, các nhà khoa học Việt Nam đang gấp rút xây dựng những phương pháp để ứng phó với sự biến đổi của khí hậu. Bài viết kết hợp mô hình chuỗi thời gian mờ và không mờ, cùng với những phương pháp xử lý số liệu ban đầu để thực hiện dự báo cụ thể đỉnh lũ, xác định chu kỳ của nó tại trạm đầu nguồn Châu Đốc. Đây là những

nghiên cứu quan trọng để nghiên cứu tổng thể thủy văn ĐBSCL phục vụ cho các quy hoạch ngắn hạn cũng như dài hạn trong xây dựng phát triển kinh tế xã hội toàn vùng.

2 PHƯƠNG PHÁP XÁC ĐỊNH CHU KỲ VÀ CÁC MÔ HÌNH DỰ BÁO ĐỈNH LŨ

2.1 Phương pháp xác định chu kỳ đỉnh lũ

Xét chuỗi thời gian đỉnh lũ theo năm $\{x_t, t = 1..n\}$, chúng ta cần tìm chu kỳ k của đỉnh lũ này. Theo [6], trong thống kê ta có 2 phương pháp chính để thực hiện: Sử dụng hàm tự tương quan và phương pháp phân tích phổ.

2.1.1 Sử dụng hệ số tự tương quan

Hệ số tự tương quan của chuỗi thời gian đỉnh lũ được xác định bởi công thức:

$$r(k) = \frac{1}{n-k} \sum_{t=1}^{n-k} \frac{(x_t - \bar{x}_0)(x_{t+k} - \bar{x}_k)}{s_0 s_k} = \frac{1}{n-k} \sum_{t=1}^{n-k} \frac{x_t x_{t+k} - \bar{x}_0 \bar{x}_k}{s_0 s_k} \quad (1)$$

$$\text{với } \bar{x}_0 = \frac{1}{n-k} \sum_{t=1}^{n-k} x_t; \bar{x}_k = \frac{1}{n-k} \sum_{t=k+1}^n x_t, s_0 = \sqrt{\frac{1}{n-k} \sum_{t=1}^{n-k} (x_t - \bar{x}_0)^2},$$

$$s_k = \sqrt{\frac{1}{n-k} \sum_{t=k+1}^n (x_t - \bar{x}_k)^2} \quad k = 1, 2, \dots, m \text{ (đơn vị tính là năm).}$$

Giá trị k ứng với $r(k)$ khá lớn hoặc gần bằng 1 sẽ được xem là các chu kỳ của chuỗi đỉnh lũ.

2.1.2 Sử dụng phương pháp phân tích phổ

Mật độ phổ được xác định như sau:

$$S(\lambda_k) = \frac{1}{\pi} \sum_{\tau=0}^m R(\tau) \cos \frac{\pi k}{m} \tau \quad (2)$$

Trong đó τ là bước trượt của hàm tự tương quan ($\tau = 1, 2, 3, \dots, m$); m là số giá trị hàm tự

tương quan cần tính; $m = \frac{\tau_{max}}{\Delta t}$ với τ_{max} là

bước trượt lớn nhất và Δt là bước thời gian của chuỗi x_t , λ_k là tần số mật độ phổ có công thức

$$\lambda_k = \frac{2\pi k}{\tau_{max}} = \frac{2\pi k}{m\Delta t}.$$

Phần tính toán từ số liệu thực có thể có những sai lệch, nên người ta dùng hàm cửa sổ để làm tron

và giảm độ sai lệch này. Có nhiều hàm cửa sổ được sử dụng như hàm Bartlette, hàm Hamming, ... Trong bài viết này, chúng tôi sử dụng cửa sổ là hàm Tukey – Hanning:

$$w(\tau) = \begin{cases} \frac{1}{2} \left(1 - \cos \frac{\pi \tau}{\tau_m}\right) & \text{khi } \tau < \tau_m \\ \tau_m & \\ 0 & \text{khi } \tau > \tau_m \end{cases}$$

Đưa hàm cửa sổ này vào (2), khi đó ta được:

$$S(\lambda_k) = 2 \left[R(0) + \sum_{\tau=1}^{\tau_m} w(\tau) R(\tau) \cos \frac{\pi k}{\tau_m} \tau \right] \quad (3)$$

Phổ chuẩn hóa được xác định bởi công thức:

$$s(\lambda_k) = \frac{S(\lambda_k)}{\sigma^2} = 2 \left[1 + \sum_{\tau=1}^{\tau_m} w(\tau) r(\tau) \cos \frac{\pi k}{\tau_m} \tau \right] \quad (4)$$

Sử dụng tiêu chuẩn Tukey để kiểm tra chu kỳ k có ý nghĩa thống kê không. Trong kiểm định này giá trị tới hạn được xác định là

$$t_\alpha(s_\lambda) = s_\lambda \frac{\chi_\alpha^2(m-1)}{v} \quad (s_\lambda \text{ là trung bình của } m$$

giá trị mật độ thực nghiệm; v là bậc tự do: $v = \frac{2n-0,5m}{m}$. Nếu $s(\lambda_k) < t_\alpha(s_\lambda)$ thì phổ

không có ý nghĩa hay chuỗi thời gian đỉnh lũ không có chu kỳ k .

2.2 Dự báo bằng mô hình chuỗi thời gian

Với số liệu theo thời gian, trong thống kê có hai phương pháp cơ bản thực hiện việc dự báo: Sử dụng mô hình hồi qui và chuỗi thời gian. Với số liệu biến đổi phức tạp như đỉnh lũ sông Hậu, việc dự báo theo phương pháp hồi qui không cho kết quả tốt, vì vậy việc dự báo ở đây chỉ thực hiện theo các mô hình chuỗi thời gian mờ và không mờ, trong đó chú trọng việc xử lý số liệu ban đầu (làm trơn, mờ hóa) trước khi dự báo.

$$y_t = \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + u_t + \beta_1 u_{t-1} + \beta_2 u_{t-2} + \dots + \beta_q u_{1-q} \quad (7)$$

Một quá trình ARMA(p,q) sẽ có quá trình tự hồi qui bậc p và quá trình trung bình di động bậc q .

$$y_t = \delta + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \dots + \alpha_p y_{t-p} + \beta_1 \varepsilon_{t-1} + \dots + \beta_q \varepsilon_{t-q} + e_t \quad (8)$$

trong đó $\alpha_i, i=1,2,\dots,p$ là tham số tự hồi qui; $\varepsilon_{t-j}, j=1,2,\dots,q$ là tham số trung bình di động; $\delta = \mu(\beta_1 + \beta_2 + \dots + \beta_q)$; μ là giá trị trung bình của chuỗi thời gian; e_t là sai số dự báo ($e_t = \hat{y}_t - y_t$ = số liệu dự báo - số liệu thực tế).

2.2.2 Mô hình chuỗi thời gian mờ Abbasov-Mamedova

Mô hình chuỗi thời gian mờ này gồm 6 bước như sau:

$$A^t = \left[\mu_{A_t}(u_i) / u_i \right], u_i \in U, \mu_{A_t} \in [0,1], \mu_{A_t}(u_i) = \frac{1}{1 + \left[C \times (U - u_m^i) \right]^2} \quad (9)$$

trong đó A^t là mờ hóa các biến của năm t ; C là hằng số tự chọn sao cho $\mu_{A_t}(u_i) \in [0,1]$; U là các biến đổi của từng năm, hoặc là giá trị trung bình; u_m^i là giá trị trung bình của từng đoạn thứ i .

Bước 4: Mờ hóa các dữ liệu đầu vào hoặc

2.2.1 Mô hình chuỗi thời gian không mờ

Mô hình tự hồi qui bậc p (AR(p)):

$$y_t = \phi_0 + \sum_{i=1}^p \phi_i y_{t-i} + u_t \quad (5)$$

trong đó ϕ_i là các hệ số ước lượng của mô hình, u_t là số hạng đảm bảo tính ổn trắng.

Mô hình trung bình di động bậc q (MA(q)):

$$y_t = \beta_0 + \sum_{i=1}^q \beta_i u_{t-i} \quad (6)$$

trong đó β_i cũng là các hệ số ước lượng của mô hình và u_i giống như trong (6).

Mô hình tự hồi qui và trung bình di động (ARMA(p,q)):

Mô hình trung bình di động tổng hợp với tự hồi qui ARIMA(p,d,q):

Bước 1: Xác định tập nền U chứa đoạn thời gian giữa các biến đổi nhỏ nhất và lớn nhất trong chuỗi dữ liệu.

Bước 2: Chia tập U thành n đoạn thời gian có độ dài bằng nhau chứa các giá trị biến đổi tương ứng với tỷ lệ tăng trưởng khác nhau của đỉnh lũ. Đồng thời tính các giá trị trung bình của từng đoạn ($u_m^i, i = 1, \dots, n$).

Bước 3: Mô tả chất lượng của các giá trị biến đổi đỉnh lũ như là một biến ngôn ngữ, xác định các giá trị tương ứng của biến ngôn ngữ hoặc thiết lập các tập mờ $F(t)$:

chuyển đổi các giá trị số vào các giá trị mờ. Hoạt động này cho phép phản ánh sự tương ứng giá trị định lượng hay định tính của tỷ lệ phát triển đỉnh lũ tiêu biểu trong giá trị của hàm quan hệ.

Bước 5: Lựa chọn tham số w ($1 < w < n$); n là số năm của dữ liệu ban đầu tương ứng với đoạn

thời gian trước khi sang năm có liên quan, tính toán các mối quan hệ mờ của ma trận $P^w(T)$.

$$R(t)[i, j] = O^w[i, j] \cap K(t)[j]$$

Hay

$$F(t) = \left[\max(R_{11}, R_{21}, \dots, R_{i1}) \max(R_{12}, R_{22}, \dots, R_{i2}) \dots \max(R_{1j}, R_{2j}, \dots, R_{ij}) \right]$$

trong đó $i = 1, \dots, w; j = 1, \dots, n$.

Bước 6: Giải mờ kết quả thu được hoặc chuyển đổi các giá trị mờ vào các giá trị định tính. Dự báo cho năm tới

$$V(t): V(t) = \frac{\sum_{i=1}^w \mu_t(u_i) \times u_m^i}{\sum_{i=1}^w \mu_t(u_i)} \quad (10)$$

Kết quả dự báo cho năm thứ t được tính theo công thức: $N(t) = N(t-1) + V(t)$ (11)

trong đó $N(t)$ là đỉnh lũ của năm t , $V(t)$ là đỉnh lũ thay đổi từ năm $t-1$ đến năm t .

3 XỬ LÝ SỐ LIỆU BAN ĐẦU TRONG DỰ BÁO

3.1 Làm trơn số liệu

Làm trơn là phương pháp cổ điển để xử lý số liệu ban đầu trong dự báo, tuy nhiên cho đến nay nó cũng được áp dụng phổ biến trong thống kê, đặc biệt cho các số liệu phức tạp của thủy văn. Theo [5] và [6] các phương pháp làm trơn chủ yếu như sau:

3.1.1 Phương pháp trung bình trượt

a. Trung bình trượt đơn

Công thức:

$$M_t = \frac{1}{N} (x_t + x_{t-1} + \dots + x_{t-N+1}) \quad (12)$$

Trong đó M_t là trung bình trượt của M thời kỳ; $x_t, x_{t-1}, \dots, x_{t-N+1}$, các giá trị của chuỗi tại N thời kỳ về phía trước.

b. Trung bình trượt kép

Trung bình trượt kép là trung bình trượt của trung bình trượt đơn, nghĩa là từ trung bình trượt đơn tính lấy trung bình trượt một lần nữa. Khi

$$x'_{t+1} = s_t = \alpha t + (1-\alpha)x_{t-1} + \alpha(1-\alpha)^2 x_{t-2} + \dots + \alpha(1-\alpha)^{t-1} x_t \quad (15)$$

$$R(t) = O^w(t) \otimes K(t) = \begin{bmatrix} R_{11} & R_{12} & \dots & R_{1j} \\ R_{21} & R_{22} & \dots & R_{2j} \\ \dots & \dots & \dots & \dots \\ R_{i1} & R_{i2} & \dots & R_{ij} \end{bmatrix}$$

đó ta có:

$$M_t^{[2]} = \frac{1}{N} (M_t + M_{t-1} + \dots + M_{t-N+1}) \quad (13)$$

Trong đó $M_t^{[2]}$ là trung bình trượt kép.

c. Trung bình trượt trung tâm

Trung bình trượt trung tâm của thời kỳ hiện tại t được lấy với cả thời kỳ trước và sau, đối xứng qua t .

– Nếu L lẻ thì trung bình trượt trung tâm được tính bởi:

$$M_{t1}^{[1]} = \frac{1}{L} (x_{t-\frac{L+1}{2}} + \dots + x_t + \dots + x_{t-\frac{L-1}{2}}) \quad (14)$$

Trong đó x_t là điểm giữa khoảng L các quan trắc.

– Nếu L chẵn thì tính theo hai bước:

Tính 2 trung bình trượt bao quanh khoảng L :

$$M_{t1}^{[1]} = \frac{1}{L} (x_{t-\frac{L}{2}+1} + \dots + x_t + \dots + x_{t+\frac{L}{2}}),$$

$$M_{t2}^{[1]} = \frac{1}{L} (x_{t-\frac{L}{2}+2} + \dots + x_t + \dots + x_{t+\frac{L}{2}+1})$$

Trong đó $M_{t1}^{[1]}, M_{t2}^{[1]}$ là các trung bình trượt đơn bao quanh khoảng L .

Khi đó trung bình trượt trung tâm của chuỗi đỉnh lũ x_t là $\frac{1}{2} (M_{t1}^{[1]} + M_{t2}^{[1]})$.

3.1.2 Phương pháp hàm mũ

Chuỗi thời gian đỉnh lũ x_{t+1} được xác định bởi công thức:

Trong đó α là hằng số làm tròn, $\alpha \in (0;1]$. Thông thường được lấy từ 0.01 đến 1.00.

a. Hàm mũ đơn

Thực hiện công thức (15) ở dạng truy hồi. Xuất phát từ giá trị ban đầu s_0 , ta cần xác định hằng số làm tròn α . Giá trị α được xác định sao cho tổng bình phương sai số $\sum (x_t - x'_t)^2$ hay $\sum (x_t - s_{t-1})^2$ nhỏ nhất. Kết quả của ràng buộc này cho ta giá trị của α là $\alpha = \frac{2}{L+1}$ với L là độ

dài thời kỳ làm tròn. Ước lượng ban đầu s_0 thường được lấy bằng trung bình số học của cả chuỗi.

b. Hàm mũ kép (Phương pháp Brown)

Đối với chuỗi có xu thế tuyến tính, giá trị thực hiện theo phương trình

$$x_t^{(2)} = ax_t + (1-\alpha)x_{t-1}^{(2)} \quad (16)$$

Trong đó $x_t^{(2)}$ là giá trị làm tròn hàm mũ kép; x_t là giá trị làm tròn hàm mũ đơn, được tính theo (16).

Giá trị ban đầu s_0 được tính bởi

$$x_0 = a - \frac{1-\alpha}{\alpha}b, x_t^{(2)} = a - \frac{2(1-\alpha)}{\alpha}b$$

Trong đó a, b là các hệ số xác định theo hồi qui tuyến tính giữa các giá trị của chuỗi và của thời gian t : $x_t = a + bt$.

Đối với chuỗi có xu thế đường cong bậc hai thì $x_t^{(3)} = ax_t^{(2)} + (1-\alpha)x_{t-1}^{(2)}$, trong đó $x_t^{(3)}$ là giá trị làm tròn hàm mũ kép, $x_t^{(2)}$ và x_t là các giá trị làm tròn tính theo (15) và (16).

Các ước lượng ban đầu tính như sau:

$$x_0 = a - \frac{1-\alpha}{\alpha}b_1 + \frac{(1-\alpha)(2-\alpha)}{2\alpha^2}2b_2,$$

$$x_0^{(2)} = a - \frac{2(1-\alpha)}{\alpha}b_1 + \frac{2(1-\alpha)(3-2\alpha)}{2\alpha^2}2b_2,$$

$$x_0^{(3)} = a - \frac{3(1-\alpha)}{\alpha}b_1 + \frac{3(1-\alpha)(4-3\alpha)}{2\alpha^2}2b_2,$$

Trong đó a, b_1 và b_2 là các hệ số hồi qui bậc hai của chuỗi số liệu với thời gian t :

$$x_t = a + b_1t + b_2t^2.$$

3.2 Mờ hóa số liệu

Hiện tại có nhiều mô hình mờ hóa số liệu, trong thực nghiệm ta thường sử dụng mô hình của Huarng [4], Chen [9], Chen-Hsu [10] và Singh [12]. Các mô hình này đều gồm 4 bước thực hiện, trong đó có 3 bước đầu giống nhau chỉ khác nhau ở bước cuối cùng: *mờ hóa dữ liệu*. Ba bước chung của các mô hình được đề nghị như sau:

Bước 1: Xác định tập nền U trên các giá trị lịch sử của chuỗi thời gian: $U = [D_{\min} - D_1; D_{\max} + D_2]$, trong đó D_{\min} , D_{\max} lần lượt là giá trị lớn nhất và nhỏ nhất của chuỗi dữ liệu, D_1, D_2 là các số dương thích hợp được chọn.

Bước 2: Chia tập U thành từng đoạn thích hợp và đều nhau U_1, U_2, \dots, U_n . Xác định các tập mờ A_i tương ứng với U_i . Nếu A_i là giá trị mờ hóa tại thời điểm t và A_j là giá trị mờ hóa tại thời điểm $t + 1$ thì ta có mối quan hệ mờ $A_i \rightarrow A_j$ ($i, j = 1, 2, \dots$).

Bước 3: Xác định các nhóm quan hệ mờ.

Bước cuối cùng của từng mô hình được đề nghị cụ thể sau:

3.2.1 Mô hình của Chen

Nguyên tắc 1: Nếu A_i là giá trị mờ hóa tại thời điểm t và chỉ có mối quan hệ mờ duy nhất là $A_i \rightarrow A_j$ thì giá trị dự báo tại thời điểm $t + 1$ là m_j (m_j là điểm giữa của đoạn U_j).

Nguyên tắc 2: Nếu A_i là giá trị mờ hóa tại thời điểm t và có nhóm mối quan hệ mờ là $A_i \rightarrow A_j, A_k, A_l, \dots$ thì giá trị dự báo tại thời điểm $t + 1$ là trung bình cộng của m_j, m_k, m_l, \dots

(m_j, m_k, m_l, \dots là điểm giữa của đoạn U_j, U_k, U_l, \dots).

Nguyên tắc 3: Nếu A_i là giá trị mờ hóa tại thời điểm t và không tồn tại mối quan hệ mờ nào thì giá trị dự báo tại thời điểm $t + 1$ là m_i .

3.2.2 Mô hình của Singh

Với $k = \overline{3, n}$, mối quan hệ mờ của phần tử k và

$k + 1$ là $A_i \rightarrow A_j$.

Với $R = 0, S = 0$, tính các giá trị sau:

$$D_i = \left\| E_i - E_{i-1} \right\| - \left\| E_{i-1} - E_{i-2} \right\|, X_i = E_i + \frac{D_i}{2}, XX_i = E_i - \frac{D_i}{2}, Y_i = E_i + D_i, YY_i = E_i - D_i, P_i = E_i + \frac{D_i}{4}, PP_i = E_i - \frac{D_i}{4}$$

$$Q_i = E_i + 2 \times D_i, QQ_i = E_i - 2 \times D_i, G_i = E_i + \frac{D_i}{6}, GG_i = E_i - \frac{D_i}{6}, H_i = E_i + 3 \times D_i, HH_i = E_i - 3 \times D_i; F_j = \frac{R + M(A_j)}{S + 1}$$

Trong đó E_i, E_{i-1}, E_{i-2} lần lượt là giá trị tại thời điểm $t, t - 1, t - 2$; A_i, A_j lần lượt là giá trị mờ tại thời điểm $t, t + 1$; F_j là giá trị dự báo tại thời

điểm $t + 1$.

Khi đó ta có các nguyên tắc mờ hóa dữ liệu như sau:

$$X_i \in U_j \Rightarrow R = R + X_i, S = S + 1; XX_i \in U_j \Rightarrow R = R + XX_i, S = S + 1; Y_i \in U_j \Rightarrow R = R + Y_i, S = S + 1;$$

$$YY_i \in U_j \Rightarrow R = R + YY_i, S = S + 1; P_i \in U_j \Rightarrow R = R + P_i, S = S + 1; PP_i \in U_j \Rightarrow R = R + PP_i, S = S + 1;$$

$$Q_i \in U_j \Rightarrow R = R + Q_i, S = S + 1; QQ_i \in U_j \Rightarrow R = R + QQ_i, S = S + 1; G_i \in U_j \Rightarrow R = R + G_i, S = S + 1;$$

$$GG_i \in U_j \Rightarrow R = R + GG_i, S = S + 1; H_i \in U_j \Rightarrow R = R + H_i, S = S + 1; HH_i \in U_j \Rightarrow R = R + HH_i, S = S + 1$$

3.2.3 Mô hình Heuristic

Ta có giá trị mờ $F(t)$ có nhóm quan hệ mờ $A_j \rightarrow A_p, A_q, A_r, A_s, \dots$ và hàm Heuristic

$$h(x; A_p, A_q, A_r, A_s, \dots) = A_{p1}, A_{p2}, \dots, A_{pk} \text{ với}$$

$x = X(t) - X(t - 1)$. Nếu $x > 0$ thì

$p_1, p_2, \dots, p_k \geq j$, ngược lại nếu $x < 0$ thì

$p_1, p_2, \dots, p_k \leq j$. Khi đó, nếu $x > 0$ thì

$$A_j \rightarrow A_{p1}, A_{p2}, \dots, A_{pk} \text{ với } p_1, p_2, \dots, p_k \geq j,$$

nếu $x < 0$ thì $A_j \rightarrow A_{p1}, A_{p2}, \dots, A_{pk}$ với

$p_1, p_2, \dots, p_k \leq j$. Nguyên tắc mờ hóa dữ liệu

tương tự mô hình của Chen.

4 TỔNG QUAN VIỆC THỰC HIỆN

4.1 Mục tiêu

Sử dụng số liệu của quá khứ về mực nước cao nhất hàng năm trên trạm đầu nguồn của sông Hậu (trạm Châu Đốc), xác định chu kỳ đỉnh lũ, tìm mô hình chuỗi thời gian dự báo thích hợp nhất, từ đó

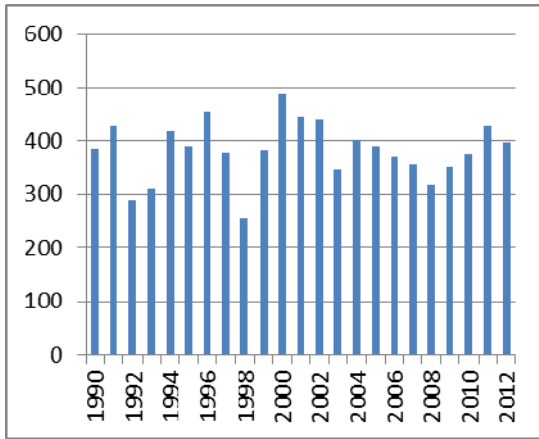
dự báo được mực nước cao nhất tại trạm đo trên cho những năm tiếp theo làm cơ sở cho các qui hoạch trong xây dựng, trong phát triển kinh tế vĩ mô của vùng.

4.2 Số liệu

Số liệu được cung cấp từ Trung tâm dự báo khí tượng khu vực Nam Bộ đặt tại Thành phố Cần Thơ. Cụ thể tại trạm đo Châu Đốc, chúng ta sử dụng số liệu mực nước cao nhất của 23 năm (giai đoạn 1990 – 2012) để thực hiện việc nghiên cứu.

Bảng 1: CFCDHR giai đoạn 1990 – 2012

Năm	Mực nước	Năm	Mực nước
1990	386	2002	442
1991	427	2003	347
1992	289	2004	402
1993	309	2005	390
1994	419	2006	370
1995	391	2007	356
1996	454	2008	320
1997	379	2009	352
1998	255	2010	377
1999	384	2011	427
2000	490	2012	398
2001	447		



Hình 1: Biểu đồ cột CFCDHR giai đoạn 1990 – 2012

4.3 Phương pháp thực hiện

Với số liệu Bảng 1, nghiên cứu lần lượt thực hiện những công việc sau:

- a) Kiểm tra dữ liệu bất thường theo nguyên tắc 3 xichma.
- b) Xác định chu kỳ đỉnh lũ bằng phương pháp hàm tự tương quan và phân tích phổ với mức ý nghĩa 5%.
- c) Tìm mô hình tối ưu nhất cho dự báo và tiến hành dự báo đỉnh lũ đến năm 2020. Cụ thể các bước sau được thực hiện:

i) Làm trơn dữ liệu gốc bằng phương pháp trung bình trượt đơn, trung bình trượt kép, trung bình trượt trung tâm và phương pháp hàm mũ đơn, hàm mũ kép. Mờ hóa dữ liệu gốc theo mô hình của

Chen, Singh, Chen-Hsu và mô hình Heuristic. Sử dụng tiêu chuẩn sai số trung bình bé nhất để chọn bộ số liệu tối ưu.

ii) Xây dựng mô hình của chuỗi thời gian từ bộ số liệu gốc, từ các bộ số liệu của đã thiết lập trong i). Sử dụng tiêu chuẩn AIC [3] để lựa chọn mô hình tốt nhất cho dự báo.

iii) Sử dụng mô hình chuỗi thời gian mờ Abbasov-Mamedova để dự báo.

iv) Kết hợp ii) và iii) để chọn mô hình tốt nhất, từ đó dự báo đỉnh lũ đến năm 2020.

Việc xây dựng các mô hình chuỗi thời gian cụ thể từ số liệu, cũng như việc tính chỉ số AIC cho mỗi mô hình được thực hiện bằng phần mềm thống kê R, phiên bản 2.0.

5 KẾT QUẢ THỰC HIỆN

5.1 Kiểm tra dữ liệu bất thường

Sắp xếp chuỗi số liệu theo thứ tự tăng dần, qui tắc 3 xichma cho ta giá trị của dữ liệu trong khoảng (216.53; 549.65). Số liệu quá khứ của đỉnh lũ sông Hậu đều nằm trong khoảng này, như vậy ta không có dữ liệu bất thường.

5.2 Xác định chu kỳ dữ liệu

5.2.1 Phương pháp hàm tự tương quan

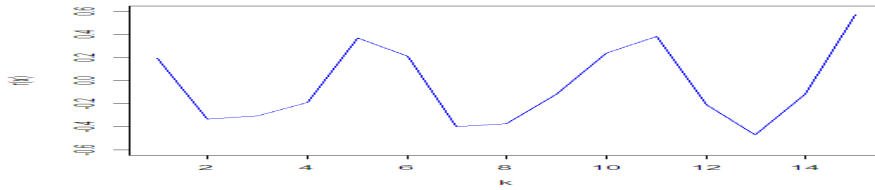
Từ số liệu ta tính được $\bar{x} = 383.09$, $Var(x) = 3082.54$. Với $\tau_{max} = 15$, $m = \frac{\tau_{max}}{\Delta t} = 15$, $\Delta t = 1$ năm, hệ số tự tương quan $r_x(\tau)$ theo (1) được tính cụ thể trong bảng sau:

Bảng 2: Hệ số tự tương quan CFCDHR với $\tau_{max} = 15$

τ	1	2	3	4	5	6	7
$r_x(\tau)$	0.1945	-0.331	-0.3043	-0.1888	0.3712	0.2109	-0.3994
τ	9	10	11	12	13	14	15
$r_x(\tau)$	-0.1174	0.2397	0.3826	-0.2084	-0.4692	-0.118	0.5718

Bảng 2 cho ta thấy hệ số tự tương quan cao hơn ở các chu kỳ 5 năm, 11 năm và 15 năm, trong đó cao nhất là chu kỳ tương ứng $\tau = 15$ năm với hệ

số tương quan 0.5718. Tuy nhiên, hệ số tương quan này cũng nhỏ hơn nhiều so với 1, nên không có ý nghĩa thống kê.



Hình 2: Đồ thị tự tương quan của CFCDHR với $\tau_{max} = 15$

5.2.2 Phương pháp phân tích phổ

kết quả tổng hợp sau:

Sử dụng các công thức trong phần 2.1.2, ta có

Bảng 3: Phổ chuẩn hóa và chu kỳ dao động của CFCDHR với $\tau_{max} = 15$

k	λ_k	$s(\lambda_k)$	t_k	k	λ_k	$s(\lambda_k)$	t_k
1	0.273	1.4200	23.0000	9	2.4574	1.2723	2.5556
2	0.5461	2.0476	11.5000	10	2.7304	1.9822	2.3000
3	0.8191	1.9294	7.6667	11	3.0035	2.209	2.0909
4	1.0922	1.9905	5.7500	12	3.2765	1.7906	1.9167
5	1.3652	3.6503	4.6000	13	3.5496	1.1935	1.7692
6	1.6383	4.7230	3.8333	14	3.8226	1.2512	1.6429
7	1.9110	2.1272	3.2857	15	4.0957	1.0238	1.5333
8	2.1843	1.3851	2.875				

Ta có:

$$v = \frac{2n - 0.5m}{m} = \frac{2 \times 23 - 0.5 \times 15}{15} = 2.57,$$

$$s\lambda = 1.9997, \chi_{0,05}^2(22) = 33.9245$$

Giá trị ngưỡng trong kiểm định Tukey là

$$t_{\alpha}(s_{\lambda}) = s_{\lambda} \frac{\chi^2}{v} = 26.4.$$

Các giá trị phổ chuẩn hóa $s(\lambda_k)$ trong Bảng 2 đều nhỏ hơn ngưỡng của kiểm định Tukey. Điều đó có nghĩa là các giá trị phổ tính được không có ý nghĩa thống kê.

Tóm lại: Cả hai phương pháp thực hiện ở trên đều cho ta kết luận không tồn tại chu kỳ rõ ràng của đỉnh lũ sông Hậu.

5.3 Mô hình tối ưu dự báo đỉnh lũ

a. Mô hình chuỗi thời gian từ số liệu gốc

Từ số liệu Bảng 1, kiểm tra ta có số liệu dừng ở sai phân bậc 3. Qua đồ thị hàm ACF và đồ thị PACF ta xác định được độ trễ $p = 1, 2, 3; q = 0, 1, 2, 3$, do đó các mô hình có thể cho việc dự báo là ARIMA(0,3,1); ARIMA(0,3,2); ARIMA(0,3,3);

ARIMA(1,3,0); ARIMA(1,3,1); ARIMA(1,3,2); ARIMA(1,3,3); ARIMA(2,3,0); ARIMA(2,3,1); ARIMA(2,3,2); ARIMA(2,3,3); ARIMA(3,3,0); ARIMA(3,3,1); ARIMA(3,3,2); ARIMA(3,3,3).
 Tính chỉ số AIC của các mô hình trên ta có bảng tổng hợp sau:

Bảng 4: Chỉ số AIC các mô hình dự báo của CFCDHR từ số liệu gốc

Mô hình	AIC	Mô hình	AIC
ARIMA(0,3,1)	252.39	ARIMA(0,3,3)	245.49
ARIMA(1,3,1)	251.44	ARIMA(1,3,3)	245.77
ARIMA(2,3,1)	249.14	ARIMA(2,3,3)	246.65
ARIMA(3,3,1)	250.9	ARIMA(3,3,3)	248.27
ARIMA(0,3,2)	243.67	ARIMA(1,3,0)	262.37
ARIMA(1,3,2)	245.61	ARIMA(2,3,0)	255.59
ARIMA(2,3,2)	250.81	ARIMA(3,3,0)	257.42
ARIMA(3,3,2)	252.22	ARIMA(0,3,3)	245.49

Bảng 4 cho ta mô hình ARIMA(0,3,2) có AIC nhỏ nhất, nghĩa là khi sử dụng dữ liệu gốc mô hình này tối ưu nhất.

b. Mô hình chuỗi thời gian từ số liệu làm trơn

Từ số liệu gốc, làm trơn theo phương pháp trung bình trượt và hàm mũ bởi các công thức trong 3.1, ta tính được bảng tổng hợp sau:

Bảng 5: Số liệu CFCDHR qua các phương pháp làm trơn

Số liệu gốc		Trượt đơn		Trượt kép		Trượt trung tâm		Hàm mũ
Năm	Q(cm)	4 năm	7 năm	4 năm	7 năm	4 năm	7 năm	
1990	386							383.09
1991	427							405.00
1992	289							347.00
1993	309	352.75				356.88	382.14	328.00
1994	419	361.00				356.50	381.14	373.50
1995	391	352.00				372.63	356.57	382.25
1996	454	393.25	382.14	364.75		402.00	370.14	418.13
1997	379	410.75	381.14	379.25		390.25	396.00	398.56
1998	255	369.75	356.57	381.44		368.88	400.00	326.78
1999	384	368.00	370.14	385.44		372.50	407.29	355.39
2000	490	377.00	396.00	381.38		385.50	392.00	422.70
2001	447	394.00	400.00	377.19		417.38	395.29	434.85
2002	442	440.75	407.00	394.94	384.76	436.13	414.57	438.42
2003	347	431.50	392.00	410.81	386.16	420.50	412.57	392.71
2004	402	409.50	395.29	418.94	388.18	402.38	393.43	397.36
2005	390	395.25	414.57	419.25	396.47	386.25	375.29	393.68
2006	370	377.25	412.57	403.38	402.53	378.38	362.43	381.84
2007	356	379.50	393.43	390.38	402.16	369.25	366.71	368.92
2008	320	359.00	375.29	377.75	398.80	354.25	370.29	344.46
2009	352	349.50	362.43	366.31	392.22	350.38	371.43	348.23
2010	377	351.25	366.71	351.25	388.60	360.13		362.61
2011	427	369.00	370.29	369.00	385.04	378.75		394.81
2012	398	388.50	371.00	388.5	378.88			396.40
	ME	39.70	40.08	46.23	35.19	34.61	45.21	23.9

Thực hiện các bước như dữ liệu gốc, ta có các mô hình có thể và chỉ số AIC tương ứng được cho bởi bảng tổng hợp sau:

Bảng 6: Chỉ số AIC các mô hình dự báo CFCDHR từ số liệu làm trơn theo phương pháp hàm mũ

Mô hình	AIC	Mô hình	AIC
ARIMA _T (0,3,1)	216.81	ARIMA _T (0,3,3)	212.98
ARIMA _T (1,3,1)	218.25	ARIMA _T (1,3,3)	213.41
ARIMA _T (2,3,1)	216.74	ARIMA _T (2,3,3)	215.14
ARIMA _T (3,3,1)	217.87	ARIMA _T (3,3,3)	216.39
ARIMA_T(0,3,2)	213.22	ARIMA _T (1,3,0)	226.76
ARIMA _T (1,3,2)	219.44	ARIMA _T (2,3,0)	223.71
ARIMA _T (2,3,2)	218.75	ARIMA _T (3,3,0)	225.25
ARIMA _T (3,3,2)	219.94		

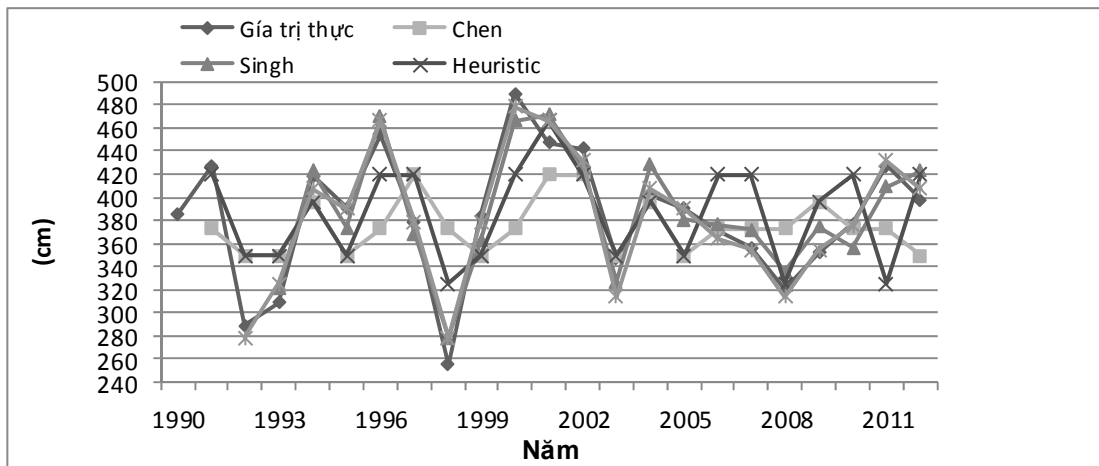
Bảng tổng hợp này cho ta mô hình tối ưu là ARIMA_T(0,3,2) vì có chỉ số AIC nhỏ nhất.

c. Mô hình chuỗi thời gian từ số liệu mờ hóa

Mờ hóa từ dữ liệu gốc theo mô hình của Chen, Singh, Heuristic và Chen-Hsu ta có bảng tổng hợp tính toán sau:

Bảng 7: Kết quả mờ hóa dữ liệu bằng mô hình Chen, Singh, Heuristic và Chen-Hsu giai đoạn 1990 - 2012 CFCDHR

Năm	Giá trị Thực tế	Chen	Singh	Heuristic	Chen-Hsu
1990	386				
1991	427	372.5		419.5	
1992	289	349.0		349.0	278.50
1993	309	349.0	320.35	349.0	325.50
1994	419	396.0	423.25	396.0	407.75
1995	391	349.0	373.25	349.0	390.13
1996	454	372.5	469.75	419.5	466.50
1997	379	419.5	368.50	419.5	378.38
1998	255	372.5	278.50	325.5	278.50
1999	384	349.0	362.75	349.0	378.38
2000	490	372.5	466.50	419.5	478.25
2001	447	419.5	471.07	466.5	466.50
2002	442	419.5	425.69	419.5	431.25
2003	347	349.0	326.75	349.0	313.75
2004	402	396.0	428.25	396.0	407.75
2005	390	349.0	380.77	349.0	390.13
2006	370	372.5	375.77	419.5	363.69
2007	356	372.5	372.21	419.5	354.88
2008	320	372.5	335.83	325.5	313.75
2009	352	396.0	374.17	396.0	354.88
2010	377	372.5	356.45	419.5	375.44
2011	427	372.5	408.75	325.5	431.25
2012	398	349	422.94	419.5	407.75
ME		42.45	17.38	38.30	9.27



Hình 3: Hình vẽ cho số liệu mờ hóa CFCDHR giai đoạn 1990 – 2011 theo các mô hình của Chen, Singh, Chen-Hsu và Heuristic

Vì số liệu có được từ mô hình Chen-Hsu cho ta ME nhỏ nhất, nên chúng ta sử dụng bộ dữ liệu này

cho dự báo. Cũng thực hiện như số liệu gốc, ta được các mô hình có thể và chỉ số AIC như sau:

Bảng 8: AIC các mô hình dự báo CFCDHR từ số liệu mờ hóa của Chen-Hsu

Mô hình	AIC	Mô hình	AIC
ARIMA _{CH} (0,2,1)	221.7	ARIMA _{CH} (2,2,1)	223.9
ARIMA _{CH} (0,2,2)	221.9	ARIMA _{CH} (2,2,2)	221.49
ARIMA_{CH}(0,2,3)	220.76	ARIMA _{CH} (2,2,3)	222.76
ARIMA _{CH} (1,2,0)	229.53	ARIMA _{CH} (3,2,0)	229.57
ARIMA _{CH} (1,2,1)	223.67	ARIMA _{CH} (3,2,1)	222.51
ARIMA _{CH} (1,2,2)	225.5	ARIMA _{CH} (3,2,2)	222.14
ARIMA _{CH} (1,2,3)	222.75	ARIMA _{CH} (3,2,3)	221.72
ARIMA _{CH} (2,2,0)	229.53		

So sánh AIC của các mô hình trên ta thấy mô hình ARIMA_{CH}(0,2,3) có chỉ số AIC bằng 220.76 là nhỏ nhất.

d. Mô hình chuỗi thời gian mờ

Thực hiện theo 6 bước của mô hình chuỗi thời gian mờ của Abbasov-Mamedova như đã trình bày trong 2.2.2 với dữ liệu Bảng 1, ta có bảng tính toán sau cùng như sau:

Bảng 9: Kết quả nội suy mô hình chuỗi thời gian mờ Abbasov-Mamedova

Năm	Đỉnh lũ thực tế	Đỉnh lũ nội suy
1998	255	371.64
1999	384	245.77
2000	490	384.87
2001	447	489.82
2002	442	440.83
2003	347	437.52
2004	402	338.69
2005	390	399.98
2006	370	385.18
2007	356	364.83
2008	320	351.08
2009	352	314.17
2010	377	349.07
2011	427	373.83
2012	398	424.60
AIC		28.08

Bảng 10: Dự báo CFCDHR giai đoạn 2013 – 2020 theo các mô hình tối ưu

Năm	ARIMA(0,3,2)	ARIMA _{T-H} (0,3,3)	ARIMA _{C-H} (0,2,3)	Chuỗi thời gian mờ
2013	396.2	386.31	337.62	419.03
2014	395.08	383.39	353.73	414.48
2015	394.63	381.02	348.31	409.98
2016	394.84	379.18	342.88	404.50
2017	395.73	377.89	337.46	400.00
2018	397.29	377.14	332.04	395.50
2019	399.52	376.93	326.61	391.00
2020	402.42	377.27	321.19	386.50

Giá trị AIC theo mô hình Abbasov-Mamedova nhỏ hơn so với các mô hình tối ưu đã thiết lập khi sử dụng dữ liệu gốc, dữ liệu làm trơn và dữ liệu mờ hóa.

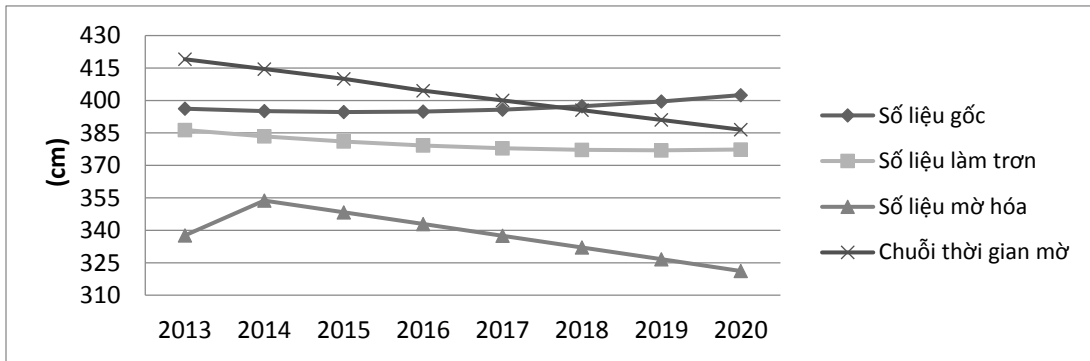
5.4 Dự báo

Từ các mô hình tối ưu đã lựa chọn trong 5.3a, 5.3b, 5.3c và 5.3d, tiến hành dự báo đỉnh lũ tại trạm đo Châu Đốc đến năm 2020, ta có bảng tổng hợp (Bảng 10).

Ta có những nhận xét sau từ việc xây dựng mô hình tối ưu và kết quả dự báo đỉnh lũ của các mô hình này như sau:

- Chỉ số AIC của các mô hình tối ưu xây dựng từ dữ liệu làm trơn, dữ liệu mờ hóa, chuỗi thời gian mờ nhỏ hơn so với mô hình xây dựng từ số liệu gốc. Cụ thể AIC tăng dần theo thứ tự: Chuỗi thời gian mờ (mô hình Abbasov-Mamedova), dữ liệu làm trơn, dữ liệu mờ hóa, dữ liệu gốc.

- Kết quả dự báo đỉnh lũ giai đoạn 2013 – 2014 từ các mô hình tối ưu có sự khác biệt. Đối với mô hình chuỗi thời gian mờ và mô hình chuỗi thời gian xây dựng từ số liệu mờ hóa thì đỉnh lũ có chiều đi xuống. Hai mô hình chuỗi thời gian được xây dựng từ số liệu gốc và số liệu làm trơn dự báo đỉnh lũ có khuynh hướng tăng nhưng tốc độ tăng không nhiều.



Hình 4: Hình vẽ dự báo CFCDHR giai đoạn 2013-2020 từ các mô hình tối ưu

6 KẾT LUẬN

Mặc dù, hệ số tương quan cũng như phổ của các chu kỳ 5 năm, 11 năm và 15 năm cao hơn ở các chu kỳ khác nhưng chúng cũng không có ý nghĩa thống kê. Kết luận chu kỳ đỉnh lũ sông Hậu không có chu kỳ phù hợp với thực tế các năm qua.

Dựa vào mô hình chuỗi thời gian mờ và không mờ, với dữ liệu gốc, dữ liệu làm tròn, dữ liệu mờ hóa, bài viết đã xây dựng mô hình tối ưu đỉnh lũ sông Hậu cho từng trường hợp. Trong các mô hình xây dựng, chỉ số *AIC* của mô hình chuỗi thời gian mờ nhỏ nhất. Dự báo thủy văn nói chung và dự báo đỉnh lũ nói riêng là bài toán phức tạp, sự chính xác của các mô hình cần phải có thời gian để kiểm chứng, tuy nhiên kết quả dự báo từ các phương pháp thống kê hiện đại trong bài viết có thể được sử dụng như một kênh tham khảo quan trọng.

Các mô hình và cách làm như đã thực hiện cho dự báo đỉnh lũ sông Hậu trong bài viết này, có thể được thực hiện theo cách tương tự cho nhiều ứng dụng khác của thực tế.

TÀI LIỆU THAM KHẢO

1. A.M. Abbasov *et al*, 2002. Fuzzy relational model for knowledge processing and decision making. *Advances in Mathematics*. 1: 1991-223.
2. A.M. Abbasov and M.H. Mamedova, 2003. Application of fuzzy time series to population forecasting, *Vienna University of Technology*. 12: 545-552.
3. H.Bozdogan, 2000. Akaike's information criterion and recent developments in information complexity. *Journal of mathematical psychology*. 44: 62-91.
4. K. Huarng, 2001. Huarng models of fuzzy time series for forecasting. *Fuzzy Sets and Systems*. 123: 369-386.
5. Nguyễn Thanh Sơn, 2003. *Tính toán thủy văn*. Nhà xuất bản Đại học Quốc gia Hà Nội. Hà Nội. 187 trang.
6. Phan Văn Tân, 2005. *Các phương pháp thống kê trong khí hậu*. Nhà xuất bản Đại học Quốc gia Hà Nội. Hà Nội. 162 trang.
7. Q. Song and B.S. Chisom, 1993. Forecasting enrollments with fuzzy time series (Part I), *Fuzzy Sets and Systems*. 54: 1-9.
8. Q. Song and B.S. Chisom, 1994. Forecasting enrollments with fuzzy time series (Part II), *Fuzzy Sets and Systems*. 62: 1-8.
9. S.M.Chen, 1996. Forecasting enrollments based on fuzzy time series. *Fuzzy Sets and Systems*. 81: 311-319.
10. S.M. Chen and C.C.Hsu, 2004. A New method to forecast enrollments using fuzzy time series. *International Journal of Applied Science and Engineering*, 12: 234-244.
11. S.R. Singh, 2008. A computational method of forecasting based on fuzzy time series. *Mathematics and Computers in Simulation*. 79: 539-554.
12. S.R. Singh, 2009. A computational method of forecasting based on high-order fuzzy time series. *Expert Systems with Applications*. 36:10551-10559.
13. Tô Văn Trường, 2005. *Phương pháp và công nghệ dự báo lũ Đồng bằng sông Cửu Long đề tài cấp nhà nước*, MS: MS:KT-08-14. Phân viện khảo sát qui hoạch thủy lợi. 112 trang.