

NGHIÊN CỨU, SỬ DỤNG MẠNG NƠ-RON NHÂN TẠO TRONG DỰ BÁO LƯU LƯỢNG NƯỚC ĐẾN HỒ HOÀ BÌNH TRƯỚC MƯỜI NGÀY

Phạm Thị Hoàng Nhung *, **Hà Quang Thụy ****

* *Khoa Công nghệ thông tin - Đại học Thủy lợi*

** *Đại học Công nghệ - Đại học Quốc gia Hà nội*

Tóm tắt: Dự báo thủy văn nói chung, dự báo lưu lượng nước nói riêng đóng vai trò hết sức quan trọng trong việc điều tiết, phòng chống thiên tai ở Việt Nam [2]. Sử dụng mô hình mạng nơ-ron nhân tạo trong dự báo lưu lượng nước đến các hồ chứa là một hướng tiếp cận được sự quan tâm đặc biệt của một số nhóm nghiên cứu trên thế giới [3-6,8-11]. Áp dụng các giải pháp được trình bày trong [4,6,9,11] vào thực tế dự báo thủy văn Việt Nam, báo cáo này trình bày một mô hình áp dụng mạng nơron nhân tạo dự báo lưu lượng nước đến hồ Hoà Bình với thời gian dự kiến 10 ngày. Kết quả thực nghiệm được thực hiện trên phần mềm tương ứng cho thấy mô hình dự báo được đề xuất là có tính khả thi cao.

Từ khoá: chỉ số hiệu quả (*Coefficient of Efficiency*), chỉ số xác định (*Coefficient of Determination*), dữ liệu học (training set), dữ liệu kiểm tra (test set), dự báo lưu lượng (flow prediction), mạng nơ-ron nhân tạo (Artificial Neural Networks), sai số quân phương (*Mean Square Error*), sai số căn quân phương (*Root Mean Square Error*), sai số tuyệt đối (*Mean Absolute Error*).

1. GIỚI THIỆU

Hiện nay có rất nhiều bài toán được đặt ra trong dự báo thủy văn như dự báo mực nước, dự báo lũ và các hiện tượng khác trên sông ngòi và hồ [1-11].

Cùng với sự phát triển của dự báo thủy văn, dự báo lưu lượng nước đến hồ chứa (nói riêng là hồ Hoà Bình) đã và đang trở nên vô cùng quan trọng vì ý nghĩa thực tiễn của nó. Dự báo chính xác lưu lượng nước đến hồ có ý nghĩa đặc biệt quan trọng trong hoạch định kế hoạch đảm bảo tốt các chức năng mà công trình phải thực hiện. Đối với hồ Hoà Bình, đó là phòng lũ cho hạ du và cho chính công trình, phòng tránh hạn hán và ổn định năng lượng điện cho cả nước¹.

Dự báo lưu lượng chịu tác động của rất nhiều nhân tố như mưa, nhiệt độ, độ ẩm, khí hậu, địa chất thổ nhưỡng, thảm phủ thực vật và hoạt động của con người trên lưu vực... Trong những nhân tố đó, nhân tố địa chất, thổ nhưỡng, thảm phủ thực vật ít thay đổi. Sự thay đổi của chế độ dòng chảy, nguồn nước theo thời gian chủ yếu do các nhân tố khí hậu quyết định, đặc biệt là nhân tố mưa, sau đó là tác động của nguồn nước dự trữ trong các tầng chứa nước trong lưu vực.

Thời gian dự báo của công trình được yêu cầu là trước 1-2 ngày (dự báo ngắn hạn), 5-10 ngày (dự báo trung hạn), 1 tháng (dự báo dài hạn) và 1 mùa (dự báo siêu dài hạn). Hiện nay, dự báo thủy văn ngắn hạn đã đạt được những thành công đáng kể cả về phương pháp dự báo lẫn kết quả dự báo. Phương pháp dự báo dài hạn cho kết quả tin cậy là vấn đề nghiên cứu vẫn nhận được sự quan tâm đặc biệt trên thế giới².

Trong những năm gần đây sử dụng mô hình mạng nơ-ron nhân tạo (Artificial Neural Network - ANN) trong dự báo lưu lượng nước đến các hồ chứa là một hướng tiếp cận được sự quan tâm đặc biệt của một số nhóm nghiên cứu trên thế giới [3-6,8-11]. ANN được coi là một công cụ mạnh để giải quyết các bài toán có tính phi tuyến, phức tạp và đặc biệt trong các trường hợp mà mối quan hệ giữa các quá trình không dễ thiết lập một cách tường minh. Khai

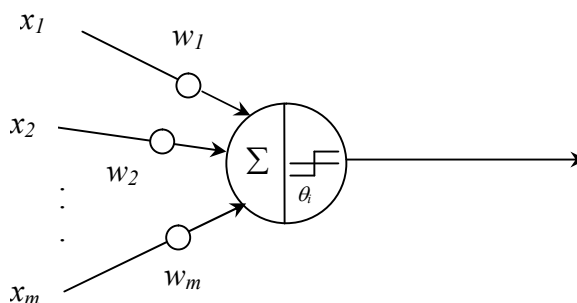
¹ <http://dut.ud.edu.vn/xdtl/modules.php?name=News&file=save&sid=8> 12/10/2006. *Hiệu quả các công trình thủy điện với công tác phòng chống lũ.*

² *Proceeding of the AGU Hydrology Days 2007*, March 19 - March 21, 2007, Colorado State University, Fort Collins, Colorado, USA. http://hydrologydays.colostate.edu/Proceeding_2007.htm

thác các giải pháp được trình bày trong [4,6,9,11], chúng tôi đề xuất các phương án ứng dụng mô hình ANN vào việc dự báo lưu lượng nước đến hồ Hoà Bình với thời gian dự kiến 10 ngày.

2. GIỚI THIỆU VỀ MẠNG NƠ-RON NHÂN TẠO

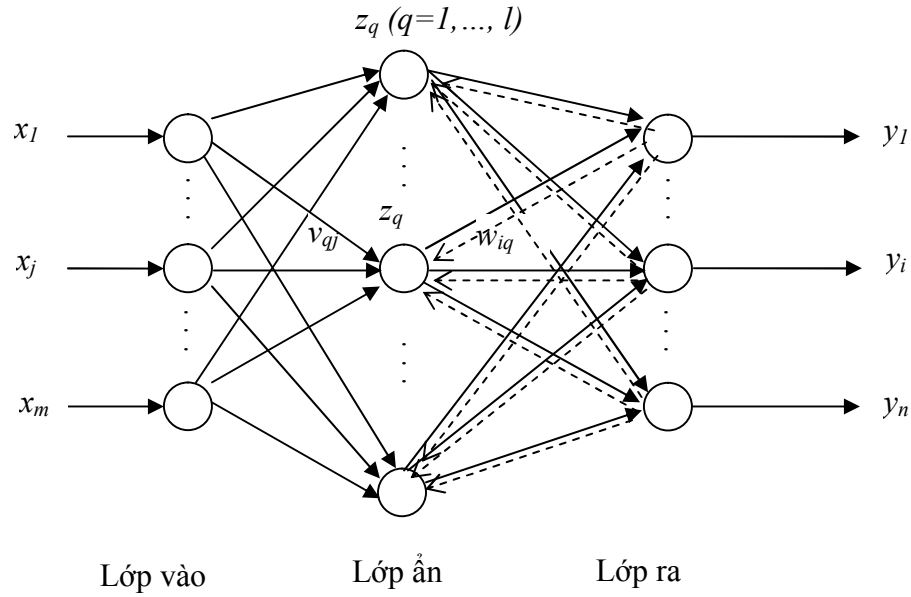
Mạng nơ-ron nhân tạo là sự mô phỏng toán học của mạng nơ-ron sinh học. Một mạng nơ-ron nhân tạo được xây dựng từ những thành phần cơ sở là những nơ-ron nhân tạo gồm nhiều đầu vào và một đầu ra (hình 1). Các đầu vào tiếp nhận kích thích từ đầu ra của những nơ-ron khác hoặc từ môi trường. Mỗi nơ-ron vào có một bộ trọng số nhằm khuếch đại tín hiệu kích thích sau đó tất cả sẽ được cộng lại. Tín hiệu sau đó sẽ được tiếp tục biến đổi nhờ một hàm phi tuyến, thường gọi là hàm kích hoạt. Và cuối cùng tín hiệu sẽ được đưa đến đầu ra của nơ-ron để lại trở thành đầu vào của các nơ-ron khác hoặc trở thành tín hiệu ra của toàn bộ mạng.



Hình 1. Kiến trúc một nơ-ron nhân tạo

Khi kết hợp các nơ-ron lại với nhau ta có một mạng nơ-ron nhân tạo. Tùy theo cách thức liên kết giữa các nơ-ron mà ta có các loại mạng khác nhau như: mạng truyền thẳng (Hình 2), mạng phản hồi, ... Ta có thể xem như mạng nơ-ron nhân tạo là một mô hình toán $Y=F(X)$ với X là véctơ số liệu đầu vào và Y là véctơ số liệu đầu ra. Ưu điểm của một mạng nơ-ron nhân tạo là nó cho phép xây dựng một mô hình tính toán có khả năng học dữ liệu rất cao. Có thể coi mạng nơ-ron nhân tạo là một hộp đen có nhiều đầu vào và nhiều đầu ra có khả năng học được mối quan hệ giữa đầu ra và đầu vào dựa trên dữ liệu được học.

Chin-Teng Lin và C.S. George Lee [10] cho rằng chỉ cần đưa vào mạng một tập mẫu dữ liệu trong quá trình học mạng sẽ phát hiện được các ràng buộc dữ liệu hữu ích mà không cần phải có thêm các tri thức về miền ứng dụng và vì vậy rất thuận lợi khi xây dựng mô hình dữ liệu. Chính vì lý do đó, mạng nơ-ron có thể được ứng dụng trong các mô hình dự báo. Theo các tác giả, cần đưa thêm hằng số quán tính để tăng tốc độ học và tránh cực tiểu cục bộ



Hình 2. Mạng nơ-ron lan truyền thẳng

Demetris F. Lekkas [8] giới thiệu hai tiếp cận mô hình mạng nơ-ron nhân tạo dự báo lưu lượng là Iterative Forecasting Procedure (Sequential Forecasting Procedure) và Simultaneous Forecasting Procedure (Direc multi-step Forecasting Procedure).

Các chỉ số dự báo được sử dụng để đánh giá chất lượng của phương pháp dự báo. Dưới đây là một số chỉ số dự báo thông dụng nhất [1]:

Sai số quân phương (*Mean Square Error*):
$$MSE = \frac{1}{n} \sum_{i=1}^n (Q_i - \hat{Q}_i)^2 \quad (1)$$

Sai số căn quân phương (*Root Mean Square Error*):
$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Q_i - \hat{Q}_i)^2} \quad (2)$$

Sai số tuyệt đối (*Mean Absolute Error*):
$$MAE = \frac{1}{n} \sum_{i=1}^n |Q_i - \hat{Q}_i| \quad (3)$$

Trong đó:

- \hat{Q}_i : Giá trị tính toán tại thời điểm i
- Q_i : Giá trị thực đo tại thời điểm i

Tuy các chỉ số (1)-(3) là trực quan và được tính toán dễ dàng song trong nhiều trường hợp khi dung lượng dữ liệu lớn hay dữ liệu có độ biến động cao thì các chỉ số này không đánh giá tốt được các kết quả dự báo. D.R. Legates và G.J. McCabe Jr. [7] trình bày về chỉ số hiệu quả - E (*Coefficient of Efficiency*) và chỉ số xác định - R^2 (*Coefficient of Determination*) tuy có độ phức tạp tính toán cao hơn song khắc phục được các hạn chế của các chỉ số (1)-(3):

$$E = 1 - \frac{\sum_{i=1}^n (Q_i - \hat{Q}_i)^2}{\sum_{i=1}^n (Q_i - \bar{Q})^2} \quad (4)$$

$$R^2 = \frac{\sum_{i=1}^n (Q_i - \bar{Q})(\hat{Q}_i - \bar{\hat{Q}})}{\sqrt{\sum_{i=1}^n (Q_i - \bar{Q})^2 \sum_{i=1}^n (\hat{Q}_i - \bar{\hat{Q}})^2}} \quad (5)$$

Các chỉ số E và R^2 có thể được dùng theo cách kết hợp hoặc riêng rẽ. Phương pháp tốt là phương pháp nhận được giá trị của các chỉ số này lớn.

Lekkas D.F. và Onof C [9] trình bày một số phương pháp cập nhật mạng nơ-ron dự báo lưu lượng để nâng cao hiệu quả dự báo. Ibrahim Can và các đồng tác giả [4] trình bày việc ứng dụng hai mô hình mạng nơ-ron mô hình one-hidden layer feed forward backpropagation (FFBP) và mô hình Generalized Regression Neural Networks (GRNN) vào dự báo tháng lưu lượng nước sông Karasu ở Thổ Nhĩ Kỳ. Các tác giả chỉ ra rằng mô hình GRNN cho kết quả rất tốt (chỉ số R^2 đạt tới 0.8572).

Bài báo ứng dụng mạng nơ-ron nhiều lớp lan truyền thẳng làm công cụ cho dự báo lưu lượng nước đến hồ Hoà Bình với giải thuật học lan truyền ngược sai số (Hình 2). Chỉ số sai số căn quân phương - RMSE và chỉ số xác định R^2 được sử dụng để đánh giá độ chính xác của dự báo.

3. ỨNG DỤNG MẠNG NƠ-RON NHÂN TẠO DỰ BÁO LƯU LƯỢNG NƯỚC ĐẾN HỒ HOÀ BÌNH TRƯỚC 10 NGÀY

3.1 Số liệu sử dụng

Chúng tôi sử dụng số liệu từ năm 1964 đến năm 2002 tại trạm đo Tạ Bú trên sông Đà, trạm đo lưu lượng gần hồ Hoà Bình nhất, bao gồm số liệu về lưu lượng và lượng mưa trong quá khứ và tại thời điểm hiện tại. Số liệu này được đo trong mùa cạn từ tháng 12 năm trước đến tháng 5 năm sau. Bộ dữ liệu được chia làm hai phần:

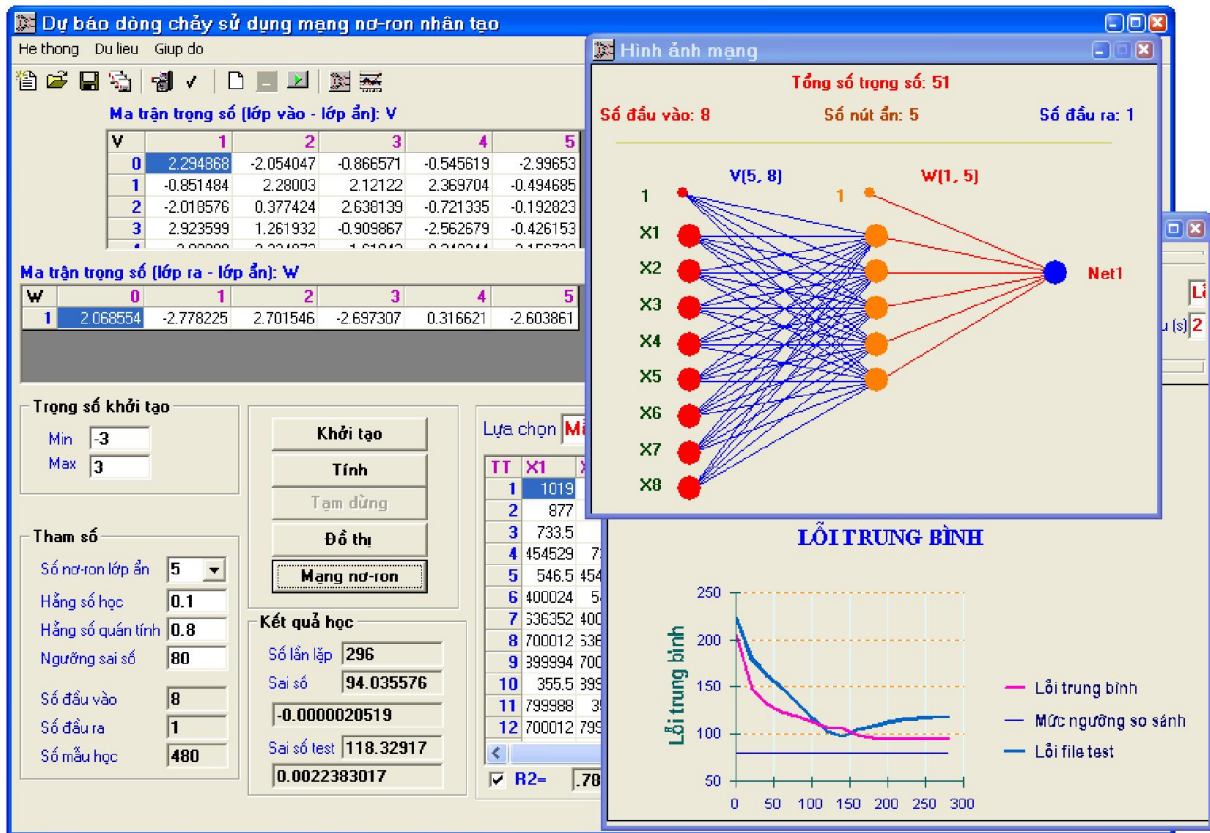
- Phần dữ liệu học (training set): Từ cuối năm 1964 đến đầu năm 1995.
- Phần dữ liệu kiểm tra (test set): Từ cuối năm 1995 đến đầu năm 2002.

3.2 Các tham số của mạng

- Kiến trúc mạng: chúng tôi sử dụng mạng nơ-ron lan truyền thẳng ba lớp: lớp vào, lớp ẩn và lớp ra. Trong đó lớp vào là đầu vào của các tham số phục vụ cho việc dự báo. Lớp ra chỉ có một nơ-ron, giá trị đầu ra của nơ-ron này chính là lưu lượng nước cần dự báo.
- Hàm kích hoạt được sử dụng là hàm sigmoid.
- Giá trị trọng số được khởi tạo ngẫu nhiên trong khoảng (-3, 3).
- Chúng tôi sử dụng thêm hằng số quán tính nhằm tăng tốc độ học và tránh cực tiểu cục bộ [10].

3.3 Các phương án thử nghiệm

Chúng tôi đã tiến hành thử nghiệm một số phương án dự báo để tìm ra phương án tối ưu. Để thực hiện điều này chúng tôi đã xây dựng một phần mềm mô phỏng mạng nơ-ron truyền thẳng sử dụng thuật toán học lan truyền ngược sai số. Phần mềm được phát triển riêng cho việc dự báo dòng chảy do đó trong quá trình chạy các tham số dự báo luôn được tính toán và hiển thị trực quan giúp cho người dự báo dễ dàng lựa chọn các phương án (hình 3).



Hình 3. Giao diện phần mềm dự báo

3.3.1 Phương án 1

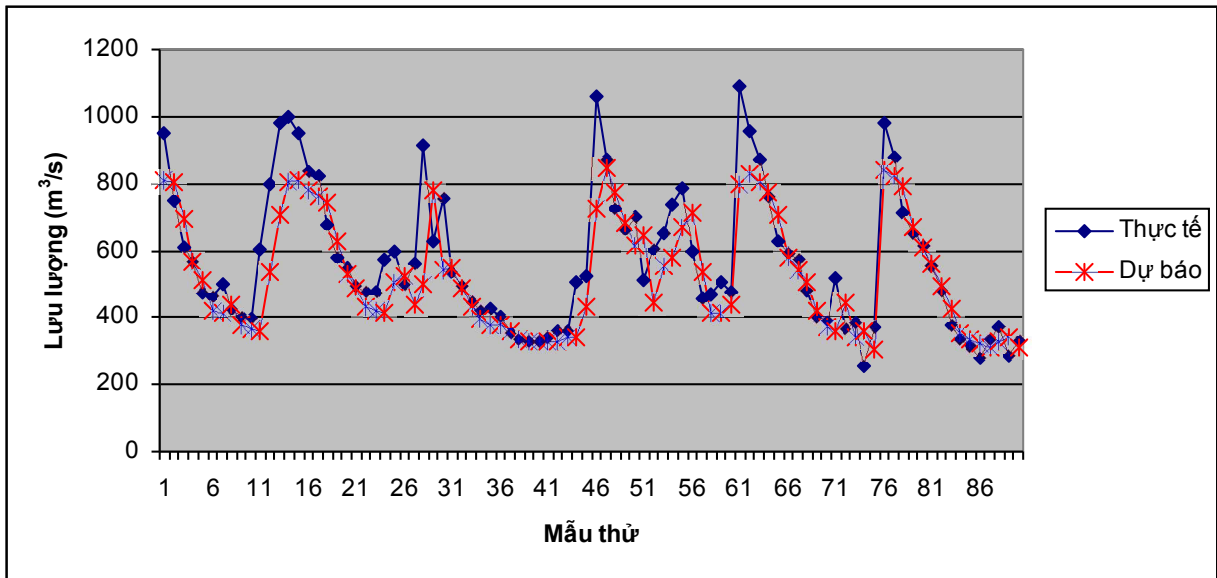
Trong phương án này việc dự báo lưu lượng nước tương lai trước 10 ngày $Q(t+10)$ dựa vào các lưu lượng nước tại thời điểm hiện tại và quá khứ. Chúng tôi sử dụng ba giá trị lưu lượng làm đầu vào của mạng gồm:

- Lưu lượng nước hiện tại: $Q(t)$
- Lưu lượng nước trung bình của 10 ngày trước đó: $Q(t-10)$
- Lưu lượng nước trung bình của 20 ngày trước đó: $Q(t-20)$

$$Q(t+10) = f(Q(t), Q(t-10), Q(t-20)) \quad (6)$$

Ta sẽ dùng mạng nơ-ron để học được mối quan hệ này. Kết quả sau khi học, các chỉ số dự báo với dữ liệu kiểm tra (test file):

- Sai số căn quân phương $RSME = 110.49 \text{ m}^3/\text{s}$
- Chỉ số $R^2 = 0.7509$



Hình 4. Kết quả dự báo với dữ liệu kiểm tra theo phương án 1

3.3.2 Phương án 2

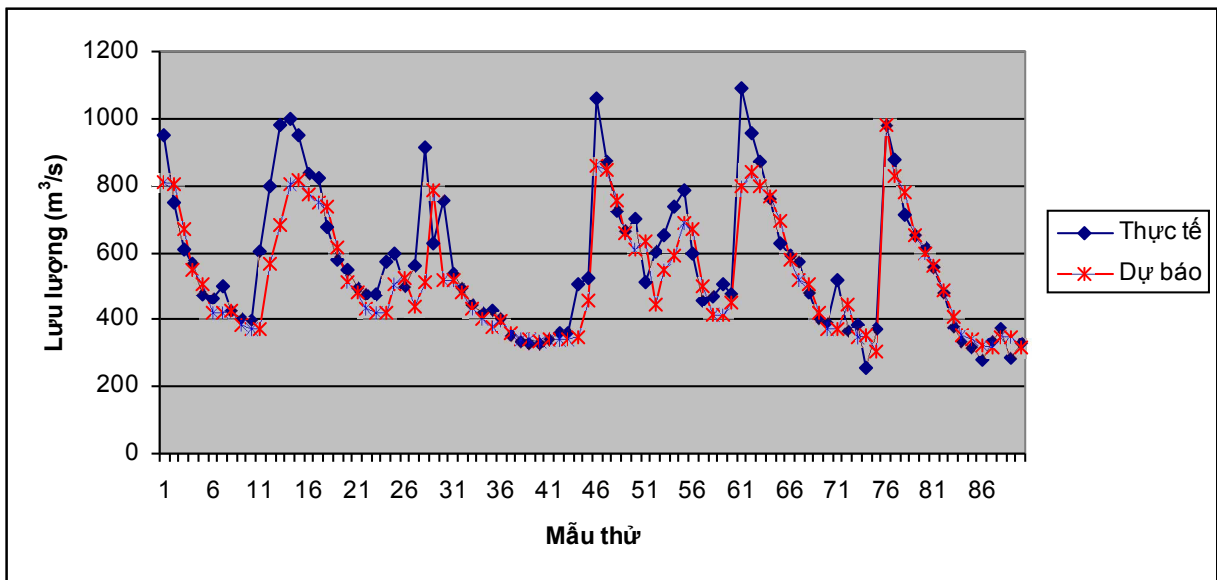
Theo các nghiên cứu thủy văn, dữ liệu về mưa trên lưu vực cũng ảnh hưởng rất lớn đến lưu lượng nước trong tương lai. Việc dự báo cần cả thông số về lượng mưa tại thời điểm hiện tại và lượng mưa trong quá khứ (do ảnh hưởng đến trữ lượng nước ngầm). Trong phương án này, dự báo lưu lượng nước tương lai trước 10 ngày $Q(t+10)$ không những chỉ dựa vào các lưu lượng nước quá khứ và hiện tại ($Q(t), Q(t-10), Q(t-20)$) như phương án 1 mà còn phụ thuộc vào lượng mưa trong quá khứ và hiện tại tại lưu vực đó ($X(t), X(t-10), X(t-20)$).

$$Q(t+10) = f(Q(t), Q(t-10), Q(t-20), X(t), X(t-10), X(t-20)) \quad (7)$$

Chúng tôi đã tiến hành cho mạng nơ-ron học mối quan hệ này và cho kết quả dự báo chính xác hơn phương án 1.

Kết quả với dữ liệu kiểm tra:

- Sai số căn quân phương $RSME = 103.22 \text{ m}^3/\text{s}$
- Chỉ số $R^2 = 0.7866$



Hình 5. Kết quả dự báo với dữ liệu kiểm tra theo phương án 2

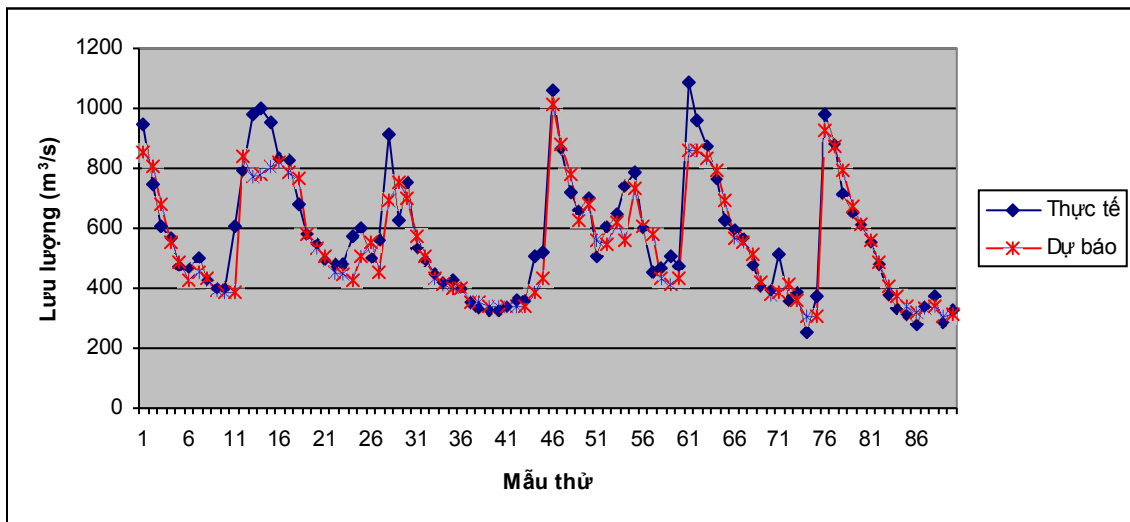
3.3.3 Phương án 3

Chúng tôi nhận thấy ngoài các yếu tố như đã nêu trong phương án 2, lưu lượng nước trong 10 ngày tới còn phụ thuộc vào lưu lượng nước của ngày hiện tại $Q_{ng}(t)$ và lượng mưa tại ngày hiện tại $X_{ng}(t)$ (khác với $Q(t)$ và $X(t)$ là giá trị lưu lượng và lượng mưa trung bình trong 10 ngày). Do đó:

$$Q(t+10) = f(Q(t), Q(t-10), Q(t-20), Q_{ng}(t), X(t), X(t-10), X(t-20), X_{ng}(t)) \quad (8)$$

Kết quả với dữ liệu kiểm tra:

- Sai số căn quân phương RSME = 76.10 m³/s
- Chỉ số $R^2 = 0.8737$



Hình 6. Kết quả dự báo với dữ liệu kiểm tra theo phương án 3

Độ chính xác của phương án dự báo này cao hơn nhiều so với hai phương án trên.

4. KẾT LUẬN

Mạng nơ-ron nhân tạo được ứng dụng trong rất nhiều lĩnh vực: nhận dạng, điều khiển, tối ưu hoá,... Trong nghiên cứu này, chúng tôi đã ứng dụng mạng nơ-ron nhân tạo vào việc dự báo lưu lượng nước đến hồ Hoà Bình trước 10 ngày. Kết quả thu được khá khả quan với chỉ số R^2 đạt 0.8737 (phương án 3), một chỉ số khá cao đối với dự báo lưu lượng nước trước 10 ngày. Trong thời gian tới, tiếp tục phát triển thêm kết quả nghiên cứu của Ibrahim Can và các đồng tác giả [4], chúng tôi hy vọng mô hình sẽ cho độ chính xác dự báo cao hơn và mở rộng thời gian dự báo. Hơn nữa, chúng tôi cũng quan tâm cải tiến mô hình theo hướng tăng tốc độ thực hiện chương trình.

LỜI CẢM ƠN. Chúng tôi xin chân thành cảm ơn GS. TS. Hà Văn Khôi, Trường Đại học Thủy lợi đã hỗ trợ cung cấp tài liệu và số liệu thử nghiệm.

TÀI LIỆU THAM KHẢO

- [1]. Lê Văn Nghinh, Bùi Công Quang, Hoàng Thanh Tùng (2006). Mô hình toán thủy văn, *Nhà xuất bản xây dựng*.
- [2]. Viện Khí tượng Thủy văn. Danh sách các đề tài, dự án nghiên cứu khoa học công nghệ. http://www.imh.ac.vn/c_tt_chuyen_nganh/ce_detai_duan/.
- [3]. Paulin Coulibaly, Francois Anctil, and Bernard Bobee (2001). Multivariate Reservoir Inflow Forecasting Using Temporal Neural Networks. *Journal of Hydrologic Engineering*, **6**(5), 367-376.
- [4]. Ibrahim Can, Cahit Yerdelen, Ercan Kahya1 (2007). Stochastic modeling of Karasu River (Turkey) using the methods of Artificial Neural Networks, *Proceeding of the AGU Hydrology Days 2007*, March 19 - March 21, 2007, Colorado State University, Fort Collins, Colorado, USA, 138-144.
- [5]. Oscar R. Dolling, Eduardo A. Varas (2002). Artificial neural networks for stream flow prediction. *Journal of Hydraulics research*, **40**(5), 547-554.
- [6]. Ismail Kiliş, Kerem Çiğizouğlu (2005). Reservoir Management Using Artificial Neural Networks, *14th. Reg. Directorate of DSI (State Hydraulic Works)*, Istanbul, Turkey, (http://balwois.mpl.ird.fr/balwois/administration/full_paper/ffp-471.pdf).
- [7]. D.R. Legates, G.J. McCabe Jr. (1998). Evaluating the Use of "Goodness-of-Fit" Measures in Hydrologic and Hydroclimatic Model Validation; *Water Resour. Res.* 1998WR900018, **35** (1): 233.
- [8]. Demetris F. Likkas (2002). Development and Comparison of Data-Based Flow Forecasting Methods, *PhD Thesis*, Department of Civil and Environmental Engineering, Imperial College of Science, Technology and Medicine, London, March 2002.
- [9]. Likkas D.F., Onof C (2005). Improved flow forecasting using artificial neural networks. *9th International Conference on Environmental and Technology*, Rhodes Island, Greece, 1-3 September 2005, 877-884.
- [10]. Chin-Teng Lin, C.S. George Lee (1996). Neural fuzzy systems: a neuro-fuzzy synergism to intelligent systems, *Prentice-Hall Inc*.
- [11]. Cristiane Medina Finzi Quintao, Walmir Matos Caminhas, Selenio Rocha Silva, Bruno Rabelo Versiani (2004). Neo Fuzzy Neuron and its applications to Prediction flood and wind in Brazil. *Workshop on Modelling and Control for Participatory Planning and Managing Water Systems*, September 29-October 1, 2004 (Poster), Venice, Italia.