

Mô phỏng dòng chảy lưu vực sông Sêrêpôk với mạng nơ-ron nhân tạo

• **Đào Nguyên Khôi**

Trường Đại học Khoa học Tự nhiên, ĐHQG-HCM

• **Huỳnh Ái Phương**

Trung Tâm Quản Lý Nước và Biến đổi Khí hậu, ĐHQG-HCM

(Bài nhận ngày 11 tháng 12 năm 2015, nhận đăng ngày 06 tháng 05 năm 2016)

TÓM TẮT

Trong nghiên cứu này, mô hình ANN được sử dụng để mô phỏng lưu lượng dòng chảy cho lưu vực sông Sêrêpôk. Phân tích tương quan về thời gian của chuỗi số liệu lượng mưa và lưu lượng được sử dụng để xác định đầu vào cho mô hình ANN. Kết quả phân tích cho thấy các giá trị lượng mưa với thời gian trễ là 2 ngày và lưu lượng với thời gian trễ là 1 ngày được chọn làm đầu vào cho mô hình ANN. Tương ứng với các số liệu đầu vào này, 3 mô hình ANN: ANN1, ANN2, và ANN3 được phát triển. Chuỗi số liệu mưa và lưu lượng năm 2002–2004 trên lưu vực được lựa

Từ khóa: mô hình ANN, mô hình thủy văn, lưu lượng dòng chảy, lưu vực sông Sêrêpôk

chọn để hiệu chỉnh và kiểm định các mô hình ANN này. Kết quả và hiệu chỉnh cho thấy mô hình ANN2 với 3 thông số đầu vào: $P(t)$, $P(t-1)$, và $Q(t-1)$ cho kết quả mô phỏng lưu lượng dòng chảy tại trạm Bản Đôn tốt nhất ($NSE = 0,95$ cho giai đoạn hiệu chỉnh và $NSE = 0,96$ cho giai đoạn kiểm định) so với 2 mô hình ANN còn lại. Bên cạnh đó, kết quả so sánh các mô hình ANN với các thông số đầu vào khác nhau cũng cho thấy rằng việc tăng các thông số đầu vào không phải lúc nào cũng cho kết quả tốt hơn.

MỞ ĐẦU

Mối quan hệ giữa lượng mưa – dòng chảy là một mối quan hệ phức tạp do sự thay đổi theo không gian và thời gian của các đặc tính lưu vực và lượng mưa. Từ những năm 1930 việc sử dụng mô hình mưa – dòng chảy để mô phỏng và dự báo dòng chảy được thực hiện khá phổ biến. Các quá trình thủy văn được mô hình hóa thành các phương trình toán học với một lượng lớn các thông số. Thí dụ, mô hình thủy văn SWAT cần đến hơn 100 thông số đầu vào để phục vụ cho bài toán mô phỏng dòng chảy. Sự tương tác của các thông số này là rất phức tạp và tối ưu hóa các thông số mô hình thường được thực hiện bằng phương pháp thử và sai [4].

Một phương pháp tiếp cận thay thế là dựa vào dữ liệu (data-based approach) để mô hình hóa mưa – dòng chảy là sử dụng mạng nơ-ron

nhân tạo (Artificial Neural Network, ANN). Một đặc tính nổi bật của ANN là chúng có khả năng điều chỉnh như cấu trúc não người. ANN bao gồm các nút, các mũi tên, và các hàm toán học để truyền tải các thông tin với nhau trong hệ thống để nhận dạng mối quan hệ giữa đầu ra và đầu vào [2]. Ứng dụng mô hình ANN trong mô phỏng và dự báo dòng chảy đã được nghiên cứu ở nhiều nơi trên thế giới và cho kết quả rất tốt. Thí dụ, Zadeh và cộng sự [4] sử dụng mô hình ANN dự báo dòng chảy ngày cho lưu vực Khosrow Shirin ở phía tây bắc Iran với hai hàm kích hoạt là logistic sigmoid và tangent sigmoid. Kết quả cho thấy hàm tangent sigmoid cho kết quả mô phỏng tốt hơn. Rezaeianzadeh và cộng sự [3] so sánh kết quả mô phỏng của mô hình ANN và mô hình HEC-HMS trong dự báo dòng chảy ở lưu vực

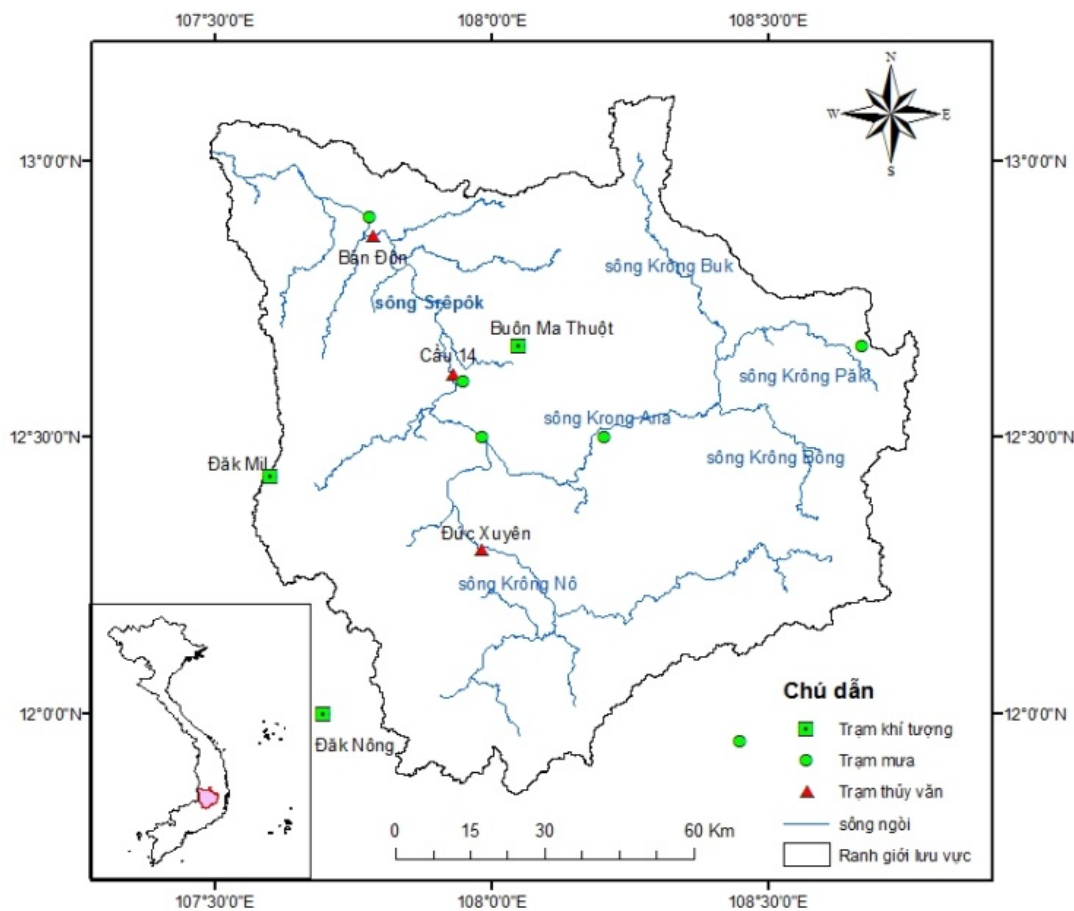
Khosrow Shirin. Kết quả nghiên cứu cho thấy mô hình ANN với hàm kích hoạt tangent sigmoid cho kết quả mô phỏng lưu lượng dòng chảy tốt hơn kết quả mô phỏng bằng mô hình HEC-HMS. Trong nghiên cứu này, mô hình ANN (thực hiện trên phần mềm WinNN32) với hàm kích hoạt tangent sigmoid được áp dụng trong mô phỏng lưu lượng dòng chảy trong lưu vực sông Sêrêpôk. Đầu vào cho mô hình ANN được xác định dựa vào phân tích tương quan chuỗi thời gian của lượng mưa và lưu lượng dòng chảy.

PHƯƠNG PHÁP

Giới thiệu khu vực nghiên cứu

Lưu vực sông Sêrêpôk trải dài trên địa bàn hai tỉnh Đắk Lắk và Đắk Nông, nằm ở khoảng

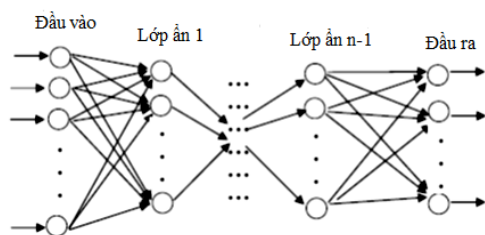
$11^{\circ}45' - 13^{\circ}15'$ vĩ độ Bắc và $107^{\circ}15' - 109^{\circ}$ kinh độ Đông (Hình 1). Sông Sêrêpôk được hình thành từ hai nhánh sông chính là Krông Nô và Krông Ana. Tổng diện tích lưu vực là 12.000 km^2 với tổng số dân khoảng 2,3 triệu người (2014). Đặc điểm khí hậu của lưu vực này là có độ ẩm cao (khoảng 78–83 %) và có hai mùa khô và mưa rõ rệt. Mùa mưa kéo dài từ tháng 5 đến tháng 10 (với đỉnh lũ thường vào khoảng tháng 9 và 10) và lượng mưa chiếm khoảng 75–95 % tổng lượng mưa năm của lưu vực. Trong lưu vực này, có hai loại đất chính là đất xám và đất bazan nâu đỏ. Các loại đất này rất màu mỡ, phù hợp cho phát triển nông nghiệp. Do đó, nông nghiệp cũng là hoạt động kinh tế chính của lưu vực.



Hình 1. Khu vực nghiên cứu và vị trí các trạm khí tượng thủy văn

Mô hình ANN

Mô hình nơ ron nhân tạo được thiết lập dựa trên cơ chế hoạt động của não người ở đó có một số lượng lớn nơ ron gắn kết để xử lý thông tin. ANN có khả năng học dữ liệu thông qua số lần lặp để điều chỉnh trọng số (hiệu chỉnh) cho phù hợp với dữ liệu quan trắc, sau đó lưu trữ tự động để dự đoán cho một giai đoạn khác (kiểm định).



Hình 2. Mạng perceptron đa lớp tổng quát tổng quát

Một cấu trúc mô hình ANN bao gồm ba phần: đầu vào, lớp ẩn và đầu ra (Hình 2). Mỗi mô hình có thể có một hoặc nhiều đầu vào, lớp ẩn và đầu ra. Các dữ liệu đầu vào sẽ được truyền vào lớp ẩn thông qua sự kết hợp với các trọng số, mỗi đầu vào sẽ có một trọng số riêng tương ứng với từng nơ-ron trên lớp ẩn. Sau khi vào lớp ẩn các giá trị sẽ được tính toán bởi các thuật toán để cho ra kết quả đầu ra, các kết quả này được so sánh với các giá trị tính toán trước nếu sai số lớn thì tiến hành điều chỉnh trọng số và tiếp tục được tính toán lại trên các lớp ẩn cho đến khi sai số được giảm tối thiểu thì cho ra kết quả đầu ra cuối cùng. Số lượng đầu ra phụ thuộc vào mục đích của người sử dụng.

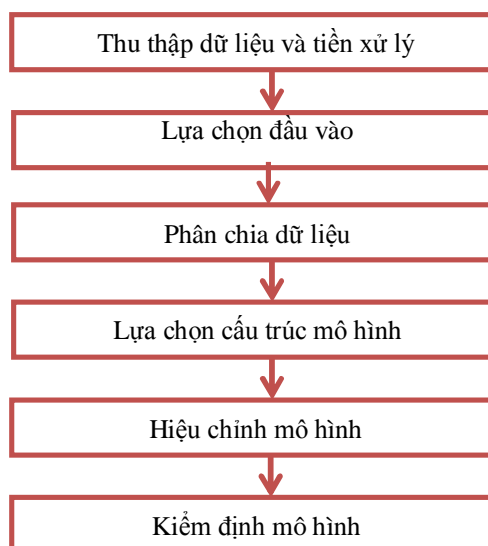
Các bước thiết lập mô hình ANN

Hình 3 mô tả các bước thiết lập mô hình ANN trong mô phỏng dòng chảy. Các bước thiết lập được mô tả như sau:

Bước 1: Thu thập dữ liệu quan trắc từ 9 trạm đo mưa trên lưu vực sông Sêrêpôk và 1 trạm thủy văn tại Bản Đôn trong giai đoạn 2002–2004. 9 trạm mưa trên lưu vực được thu thập bao gồm: trạm Đức Xuyên, Đà Lạt, Madrăk, Buôn Ma

Thuột, Buôn Hồ, Giang Sơn, Bản Đôn, Cầu 14, và Đăk Nông. Lượng mưa lưu vực được tính bằng phương pháp đa giác Thiessen với số liệu lượng mưa từ 9 trạm đo mưa.

Bước 2: Để xác định đầu vào cho mô hình ANN, phân tích tương quan riêng phần (partial autocorrelation) chuỗi số liệu lượng mưa với lưu lượng dòng chảy nhằm để xác (thực hiện trên phần mềm Mathlap) tại khu vực nghiên cứu. Điều này nhằm mục đích xác định độ lệch pha giữa lượng mưa của 9 trạm nêu trên so với lưu lượng tại trạm Bản Đôn, khoảng thời gian lệch pha được gọi là thời gian trễ. Sau đó lưu lượng trễ và lượng mưa lần lượt được kết hợp với nhau để lựa chọn đầu vào thích hợp nhất với mô hình ANN với hiệu quả mô phỏng tối ưu nhất.



Hình 3. Các bước thiết lập mô hình ANN

Bước 3: Dữ liệu lượng mưa và lưu lượng (2002–2004) trong nghiên cứu được chia thành hai phần: số liệu năm 2002–2003 được sử dụng cho hiệu chỉnh mô hình để xác định các thông số mô hình và số liệu năm 2004 được sử dụng cho kiểm định mô hình.

Bước 4: Trong bước này sẽ tiến hành lựa chọn số đầu vào, lớp ẩn, đầu ra. Mô hình được áp

dụng là mô hình một lớp ẩn vì chúng thích hợp với bất cứ một hàm liên tục nào và chúng linh động dễ thích nghi khi số nơ-ron ẩn của chúng thay đổi, trên lớp ẩn cần xác định số nơ-ron ẩn số này có thể dao động trong khoảng từ $N/3-N*4$ nơ-ron ẩn. Áp dụng thuật toán tangent sigmoid cho lớp ẩn và đầu ra, còn đầu vào sử dụng các phép tính tuyến tính ngoài ra còn sử dụng thuật toán truyền ngược để sai số được tự động giảm thiểu sau một lần lặp.

Bước 5 và 6: Hiệu chỉnh và kiểm định mô hình sẽ biểu diễn kết quả mô phỏng và kết quả tính toán có nghĩa là giữa kết quả lưu lượng được mô phỏng bởi mô hình và lưu lượng thực đo tại trạm Bản Đôn.

Đánh giá hiệu quả mô hình

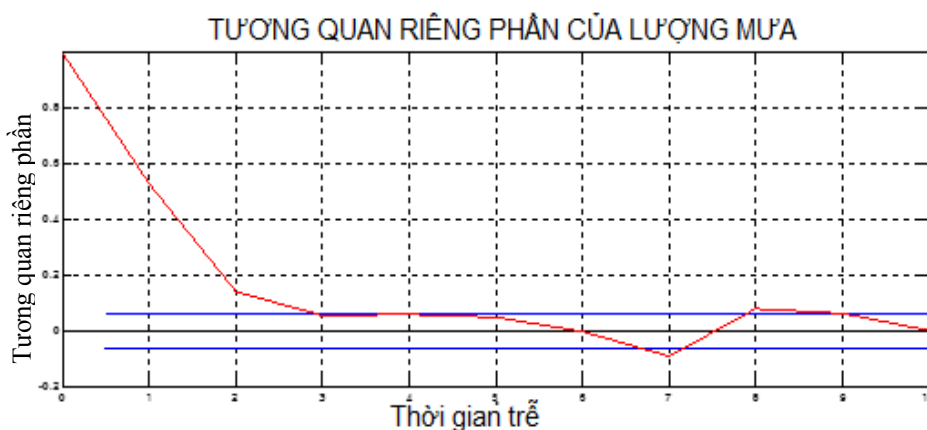
Hiệu quả mô phỏng của mô hình được đánh giá bằng phương pháp đồ thị và phương pháp thống kê để so sánh chất lượng và độ tin cậy của kết quả mô phỏng với số liệu thực đo. Trong nghiên cứu này, hai phương pháp thống kê đánh giá kết quả mô hình bao gồm: hệ số tương quan (R^2) và chỉ số hiệu quả Nash-Sutcliffe (NSE). Mô hình được xem là tốt khi giá trị R^2 và NSE lớn hơn 0,75, thỏa mãn khi R^2 và NSE thuộc khoảng

0,36 và 0,75, và không thỏa mãn khi R^2 và NSE nhỏ hơn 0,36 (Krause và cộng sự, 2007).

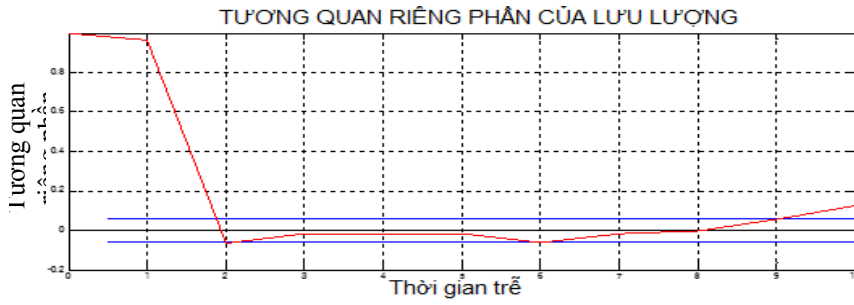
KẾT QUẢ VÀ THẢO LUẬN

Phân tích tương quan của chuỗi thời gian để đánh giá ảnh hưởng của lượng mưa và lưu lượng trữ. Đại lượng thống kê tương quan riêng phần và khoảng tin cậy 95 % từ lag 0 đến lag 10 được ước tính từ số liệu lượng mưa (Hình 4) và lưu lượng (Hình 5). Theo kết quả trên Hình 4 và 5, hàm tương quan riêng phần chỉ rõ sự dao động đáng kể đến lag 2 đối với chuỗi số liệu lượng mưa và đến lag 1 đối với chuỗi số liệu lưu lượng, sau đó thì chúng nằm trong khoảng giới hạn tin cậy 95 %.

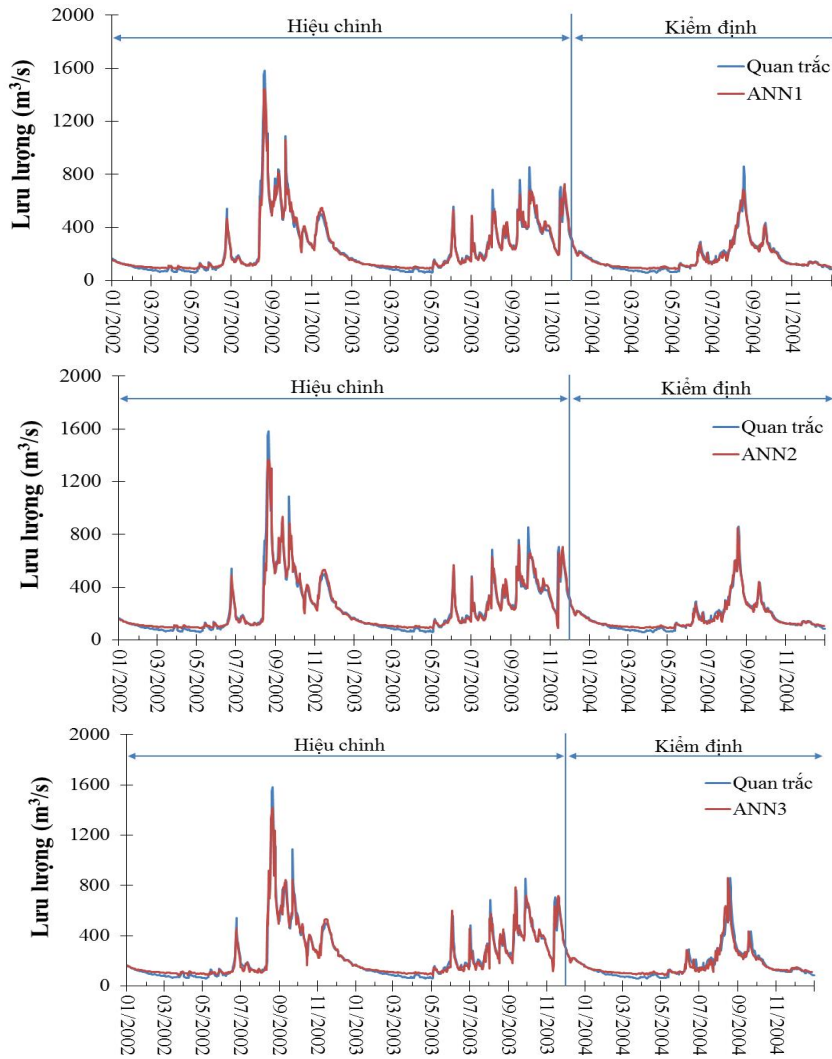
Từ kết quả phân tích tương quan, số liệu lượng mưa với thời gian trễ từ 0 đến 2 ngày (lag 0 đến 2) và số liệu lưu lượng với thời gian trễ là 1 ngày (lag 1) được chọn là thông số đầu vào cho mô hình ANN. Các thông số đầu vào này được kết hợp với nhau để tạo ra 3 mô hình ANN với các đầu vào khác nhau (Bảng 1). Tất cả các mô hình ANN này được hiệu chỉnh đều sử dụng mô hình 3 lớp: đầu vào, lớp ẩn và đầu ra với hàm kích hoạt tangent sigmoid.



Hình 4. Kết quả hàm tương quan riêng phần của chuỗi số liệu lượng mưa



Hình 5. Kết quả hàm tương quan riêng phần của chuỗi số liệu lưu lượng



Hình 6. Đồ thị so sánh đường quá trình lưu lượng mô phỏng và thực đo tại trạm Bản Đôn (2002–2004)

Trong nghiên cứu này, quá trình hiệu chỉnh (2002–2003) và kiểm định (2004) mô hình ANN được tiến hành dựa vào số liệu quan trắc tại trạm thủy văn chính của lưu vực (trạm Bản Đôn). Kết quả đường so sánh quá trình lưu lượng mô phỏng từ 3 mô hình ANN với lưu lượng thực đo tại trạm Bản Đôn được thể hiện trên Hình 6. Nhìn chung kết quả mô phỏng và kết quả thực đo là khá giống nhau và điều này cũng được thể hiện qua các chỉ tiêu đánh giá thống kê (NSE và R^2) của 3 mô hình ANN trong Bảng 1. Kết quả cho thấy hiệu quả mô phỏng của mô hình ANN1 với chỉ số $NSE = 0,94$; $R^2 = 0,97$ cho quá trình hiệu chỉnh và $NSE = 0,96$; $R^2 = 0,98$ cho quá trình kiểm định; hiệu quả mô phỏng của mô hình ANN2 với $NSE = 0,95$; $R^2 = 0,97$ cho quá trình hiệu chỉnh và $NSE = 0,96$; $R^2 = 0,98$ cho quá trình kiểm định; và hiệu quả mô phỏng của mô hình ANN3 là $NSE = 0,95$; $R^2 = 0,97$ cho quá trình hiệu chỉnh và $NSE = 0,95$; $R^2 = 0,97$ cho quá trình kiểm định. Mặc dù hiệu quả mô phỏng của 3 mô hình ANN là tốt, tuy nhiên tại một vài thời điểm đỉnh lũ của kết quả mô hình thấp hơn so với đỉnh lũ của kết quả quan trắc. Điều này có thể được lý giải thông qua

sự phân bố không đều của các trạm mưa trên lưu vực (Hình 1).

Với kết quả đánh giá hiệu quả mô phỏng của các mô hình ANN này dựa vào đồ thị so sánh và các chỉ tiêu đánh giá cho thấy mô hình ANN2 với 3 thông số đầu vào: $P(t)$, $P(t-1)$, và $Q(t-1)$ cho kết quả mô phỏng tốt hơn so với 2 mô hình ANN còn lại (ANN1 và ANN3). Bên cạnh đó kết quả so sánh các mô hình ANN với các thông số đầu vào cũng cho thấy rằng việc tăng các thông số đầu vào không phải lúc nào cũng cho kết quả tốt hơn. Cụ thể trong trường hợp này mô hình ANN3 với 3 thông số đầu vào cho kết quả tốt hơn mô hình ANN1 với 2 thông số đầu vào và mô hình ANN3 với 4 thông số đầu vào.

Kết quả hiệu chỉnh và kiểm định mô hình chỉ ra rằng các mô hình ANN (ANN1, ANN2, và ANN3) có thể mô phỏng tốt dòng chảy trong lưu vực Sêrêpôk trong giai đoạn hiện tại và có thể sử dụng các mô hình ANN này với bộ trọng số đã được hiệu chỉnh và kiểm định này cho mô phỏng dòng chảy dưới ảnh hưởng của các kịch bản biến đổi khí hậu.

Bảng 1. Bảng so sánh hiệu quả mô phỏng của các mô hình ANN với giá trị thực đo tại trạm Bản Đôn

Đầu vào	Tên mô hình	NSE		R^2	
		Hiệu chỉnh	Kiểm định	Hiệu chỉnh	Kiểm định
$P(t)$, $Q(t-1)$	ANN1	0,94	0,96	0,97	0,98
$P(t)$, $P(t-1)$, $Q(t-1)$	ANN2	0,95	0,96	0,97	0,98
$P(t)$, $P(t-1)$, $P(t-2)$, $Q(t-1)$	ANN3	0,95	0,95	0,97	0,97

KẾT LUẬN

Nghiên cứu này được thực hiện nhằm mục tiêu đánh giá khả năng mô phỏng dòng chảy bằng mô hình ANN – một trường hợp nghiên cứu cho lưu vực sông Sêrêpôk. Kết quả đạt được của nghiên cứu này có thể được tóm tắt như sau:

Kết quả phân tích tương quan của chuỗi số liệu lượng mưa và lưu lượng cho thấy chuỗi số liệu lượng mưa với thời gian trễ từ 0 đến 2 ngày (lag 0 đến lag 2) và chuỗi số liệu lưu lượng với thời gian trễ là 1 ngày (lag 1) là tương quan

mạnh. Các thông số này có thể được sử dụng làm thông số đầu vào của mô hình ANN cho các nghiên cứu tương tự tiếp theo trên lưu vực này.

Kết quả hiệu chỉnh và kiểm định của các mô hình ANN cho thấy các mô hình này có thể mô phỏng tốt lưu lượng dòng chảy cho khu vực nghiên cứu và mô hình ANN2 với 3 thông số đầu vào ($P(t)$, $P(t-1)$, và $Q(t-1)$) cho kết quả mô phỏng tốt nhất.

Kết quả nghiên cứu cũng cho thấy việc tăng các thông số đầu vào không phải lúc nào cũng cho kết quả tốt hơn. Do đó, việc lựa chọn các thông số đầu vào cho mô hình ANN có ảnh hưởng nhất định đến độ chính xác của kết quả mô phỏng. Nghiên cứu này mới chỉ sử dụng số liệu của một lưu vực sông (cụ thể là lưu vực sông Sêrêpôk) nên để tổng quát hóa được các kết quả

này thì các nghiên cứu tiếp theo với sự phong phú về số liệu (ở các lưu vực khác) cần được tiến hành.

Lời cảm ơn: Nghiên cứu này được tài trợ bởi Quỹ phát triển Khoa học và Công nghệ Quốc gia (NAFOSTED) trong đề tài mã số “105.06-2013.09”.

Using artificial neural network in simulating of the streamflow in the Srepok watershed

- **Dao Nguyen Khoi**
University of Science, VNU-HCM
- **Huynh Ai Phuong**
Center of Water Management and Climate Change, VNU-HCM

ABSTRACT

In this study, artificial neural network (ANN) model was used to simulate the streamflow in the Srepok watershed, Vietnam. Correlation analysis of time series for precipitation and streamflow was employed to determine input data for the ANN model. This result indicated a significant correlation up to 2 day time lag and 1 day time lag for the precipitation and streamflow series data, respectively. According to the correlation analysis, three ANN models including ANN1, ANN2, and ANN3 were investigated. A 3-year

data record for the precipitation and streamflow was used for ANN training and testing. The result of ANN training and testing showed that the ANN2 with 3 input data ($P(t)$, $P(t-1)$, and $Q(t-1)$) gave the best simulation ($NSE = 0.95$ for training period and $NSE = 0.96$ for testing period) comparing to those of ANN1 and ANN3. In addition, the comparison of ANNs showed that the increase of the input data did not offer the better result.

Keywords: ANN model, hydrological model, streamflow, Srepok watershed

TÀI LIỆU THAM KHẢO

- [1]. P. Krause, D.P. Boyle, F. Base, Comparison of different efficiency criteria for hydrological model assessment, *Advances in Geosciences*, 5, 89-97 (2005).
- [2]. E. Mutlu, I. Chaubey, H. Hexmoor, S.G. Bajwa, Comparison of artificial neural network models for hydrologic predictions at multiple gauging stations in an agricultural watershed, *Hydrological Processes*, 22, 26, 5097-5106 (2008).
- [3]. M. Rezaeianzadeh, A. Stein, H. Tabari, H. Abghari, N. Jalalkamali, E.Z. Hosseinipour, V.P. Singh, Assessment of a conceptual hydrological model and artificial neural networks for daily outflows forecasting, *International Journal of Environmental Science and Technology*, 10, 1181-1192 (2013).
- [4]. M.R. Zadeh, S. Amin, D. Khalili, V.P. Singh, Daily outflow prediction by multi-layer perceptron with logistic sigmoid and tangent sigmoid activation functions, *Water Resources Management*, 24, 2673-2688 (2010).