

# MÔ HÌNH BỘ NHỚ DÀI - NGẮN LSTM CHO MÔ PHỎNG DÒNG CHẢY LƯU VỰC SÔNG THU BỒN

Phạm Văn Chiến

Trường Đại học Thủy lợi

Nguyễn Hoàng Bách

Viện Khoa học Tài nguyên nước

**Tóm tắt:** Bài báo này trình bày các kết quả mô phỏng lưu lượng dòng chảy trên lưu vực sông Thu Bồn tính đến trạm thủy văn Nông Sơn sử dụng mô hình bộ nhớ dài – ngắn LSTM. Chuỗi số liệu mưa tại Nông Sơn, Hiệp Đức, Tiên Phước, Trà My và lưu lượng dòng chảy tại Nông Sơn từ 1/1/1977 đến 31/12/2020 đã được thu thập để phục vụ cho các mục đích tính toán. Phương pháp thử sai đã được sử dụng để xác định giá trị phù hợp của các thông số mô hình LSTM, trong khi bốn chỉ tiêu đánh giá sai số (sai số quân phương - RMSE, sai số trung bình tuyệt đối - MAE, hệ số tương quan -  $r$  và hệ số Nash–Sutcliffe efficiency - NSE) đã được áp dụng để đánh giá sự phù hợp giữa lưu lượng dòng chảy tính toán và thực đo. Kết quả tính toán thể hiện rằng mô hình LSTM đã tái hiện khá tốt chuỗi lưu lượng dòng chảy thực đo trong thời kỳ nhiều năm trên lưu vực nghiên cứu. Hệ số tương quan  $r$  thay đổi từ 0,87 đến 0,95, trong khi hệ số NSE dao động từ 0,73 đến 0,90. Giá trị RMSE và MAE của lưu lượng dòng chảy nhỏ hơn 3,52% biên độ lưu lượng dòng chảy thực đo ghi nhận tại trạm. Trong bốn trạm mưa thu thập trên lưu vực nghiên cứu, dữ liệu mưa tại trạm Trà My cho các kết quả mô phỏng chuỗi lưu lượng dòng chảy thực đo tại Nông Sơn tốt nhất khi sử dụng mô hình LSTM, với hệ số tương quan  $r$  lớn hơn 0,94 và hệ số NSE lớn hơn 0,88. Ảnh hưởng của dữ liệu mưa và khả năng nâng cao độ chính xác kết quả mô phỏng lưu lượng dòng chảy khi sử dụng mô hình LSTM cũng được thảo luận.

**Từ khóa:** Sông Thu Bồn, Mô hình LSTM, Dòng chảy.

**Summary:** This paper presents the results of streamflow simulation in the Thu Bon river basin up to Nong Son hydrological station using the long-short memory model LSTM. Rainfall data at Nong Son, Hiep Duc, Tien Phuoc, Tra My and streamflow at Nong Son from 1/1/1977 to 31/12/2020 were collected for different calculation purposes. The trial-and-error method was used to determine the appropriate values of modelling parameters in the LSTM model, while four criteria (including root mean square error - RMSE, mean absolute error - MAE, correlation coefficient -  $r$ , and Nash–Sutcliffe efficiency - NSE) were applied to quantitatively evaluate the agreement between calculated and measured streamflow. The results show that the LSTM model reproduced well the observed streamflow over the long-collected period in the studied basin. The correlation coefficient  $r$  varies from 0.87 to 0.95, while the NSE coefficient ranges from 0.73 to 0.90. The values of RMSE and MAE are less than 3.52% of the observed magnitude of streamflow recorded at the station. Among the four rainfall stations collected in the studied river basin, rainfall data at Tra My station showed the best simulation results of the observed streamflow at Nong Son when using the LSTM model, with the correlation coefficient  $r$  is greater than 0.94 and NSE coefficient is greater than 0.88. Finally, the effects of rainfall and the potential for enhancing the accuracy of streamflow simulation using the LSTM model were also discussed.

**Keywords:** Thu Bon river, LSTM model, Streamflow.

## 1. ĐẶT VẤN ĐỀ

Mô hình bộ nhớ dài – ngắn LSTM đang ngày càng được áp dụng phổ biến trong các nghiên

cứu về khoa học thủy văn [1-2]. Tuy nhiên, việc áp dụng các mô hình trên trong các nghiên cứu về tài nguyên nước, dòng chảy trên các lưu vực sông cụ thể ở nước ta còn chưa phổ biến như các lĩnh vực khác. Lý do là các mô hình dựa trên dữ liệu như LSTM thường yêu cầu một lượng dữ liệu quan trắc đủ dài để

Ngày nhận bài: 08/3/2024

Ngày thông qua phản biện: 29/3/2024

Ngày duyệt đăng: 02/4/2024

học được hiệu quả các đặc trưng dòng chảy. Mặt khác, các đặc trưng của dòng chảy trên lưu vực sông phụ thuộc vào vị trí địa lý và khí hậu, do đó, các lưu vực sông khác nhau sẽ có các đặc trưng dòng chảy khác nhau, đòi hỏi cần có sự xem xét cẩn thận trong quá trình tính toán, mô phỏng. Nghiên cứu, phát triển và ứng dụng mô hình LSTM cho mô phỏng, dự báo dòng chảy là một hướng tiếp cận mới nhiều tiềm năng [1], làm cơ sở cho việc mô hình hóa và dự báo dòng chảy một cách chính xác nhằm đáp ứng các yêu cầu sử dụng nước khác nhau của các ngành kinh tế xã hội.

Gần đây, một số nhà khoa học đã nghiên cứu và áp dụng mô hình LSTM cho mô phỏng, dự báo các đặc trưng dòng chảy tại một số lưu vực sông ở Việt Nam. Hien et al. [3] đã xây dựng mô hình mạng nơ-ron hồi quy dựa trên mã nguồn mở để dự báo lưu lượng dòng chảy lũ tại Lai Châu trên sông Đà sử dụng dữ liệu dòng chảy ngày thực đo tại trạm Lai Châu trong thời kỳ nhiều năm từ 1959-2004. Pham Van and Le [4] đã phát triển và ứng dụng mô hình LSTM cho mô phỏng dòng chảy trên lưu vực sông Hiếu sử dụng dữ liệu khí tượng thủy văn trong thời kỳ nhiều năm từ 1991-2020. Nguyen et al. [5] ứng dụng mô hình LSTM cho dự báo dòng chảy tại các vị trí khác nhau vùng Đồng bằng sông Cửu Long sử dụng chuỗi số liệu thủy văn trong thời kỳ từ 2014-2018. Kết quả nghiên cứu và ứng dụng mô hình LSTM cho thấy sự vượt trội của mô hình LSTM trong việc tính toán, mô phỏng lưu lượng dòng chảy so với các mô hình truyền thống (như mưa – dòng chảy NAM, phương pháp đường quan hệ mực nước – lưu lượng). Đồng thời, các ví dụ trên cũng chứng minh rằng mô hình LSTM hoàn toàn có thể được sử dụng để mô phỏng lưu lượng dòng chảy trên lưu vực sông Thu Bồn.

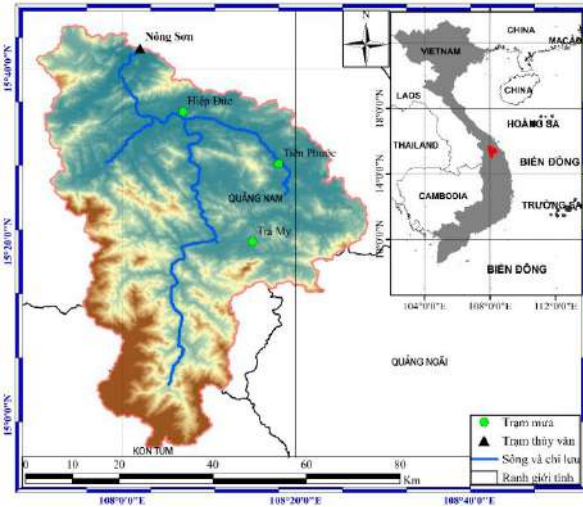
Mục tiêu chính của bài báo này là nghiên cứu ứng dụng mô hình LSTM mô phỏng lưu lượng dòng chảy trên lưu vực sông Thu Bồn

tính đến trạm thủy văn Nông Sơn. Mục tiêu cụ thể của nghiên cứu là (i) đánh giá định lượng ảnh hưởng của các dữ liệu đầu vào (mưa tại bốn trạm trên lưu vực) đến kết quả mô phỏng lưu lượng dòng chảy tại Nông Sơn và (ii) ước lượng giá trị phù hợp của các tham số khi ứng dụng mô hình LSTM. Chuỗi số liệu mưa ngày tại Nông Sơn, Hiệp Đức, Tiên Phước, Trà My và lưu lượng dòng chảy ngày tại trạm Nông Sơn trong thời kỳ từ 1/1/1977 đến 31/12/2020 đã thu thập và sử dụng cho đào tạo và kiểm định mô hình LSTM. Bốn chỉ tiêu đánh giá sai số khác nhau sẽ được sử dụng để đánh giá định lượng sự phù hợp giữa lưu lượng dòng chảy tính toán và giá trị thực đo.

## 2. LƯU VỰC NGHIÊN CỨU VÀ THU THẬP DỮ LIỆU

### 2.1. Giới thiệu về lưu vực nghiên cứu

Vu Gia - Thu Bồn là một trong những lưu vực sông lớn ở Việt Nam (Hình 1), với tổng diện tích lưu vực là 10.350 km<sup>2</sup> thuộc các tỉnh Kon Tum, Quảng Nam và Đà Nẵng [6]. Lượng mưa hàng năm trên lưu vực thay đổi từ 2.000 đến 4.000 mm. Trong đó, ở vùng núi cao như Trà My, Tiên Phước, Khâm Đức, lượng mưa hàng năm dao động trong khoảng từ 3.000 đến 4.000 mm. Ở vùng núi trung bình như Nông Sơn và Quế Sơn, lượng mưa hàng năm biến đổi từ 2.500 đến 3.000 mm. Ở vùng núi thấp và đồng bằng ven biển như Hội Khách, Ái Nghĩa, Giao Thủy, Hội An, Tam Kỳ, lượng mưa hàng năm từ 2.000 đến 2.500 mm. Lượng mưa phong phú như vậy là điều kiện tốt để thúc đẩy tăng trưởng về kinh tế - xã hội lưu vực sông Vu Gia – Thu Bồn. Tuy nhiên, lượng mưa phân bố không đều theo thời gian mà chủ yếu tập trung vào 3 – 4 tháng mùa lũ. Với thời kỳ khô hạn kéo dài tới 8 – 9 tháng, dẫn đến tình trạng thiếu nước trong các tháng mùa kiệt. Đặc biệt, trong những năm gần đây, tình trạng thiếu nước diễn ra thường xuyên với mức độ ngày càng nghiêm trọng.



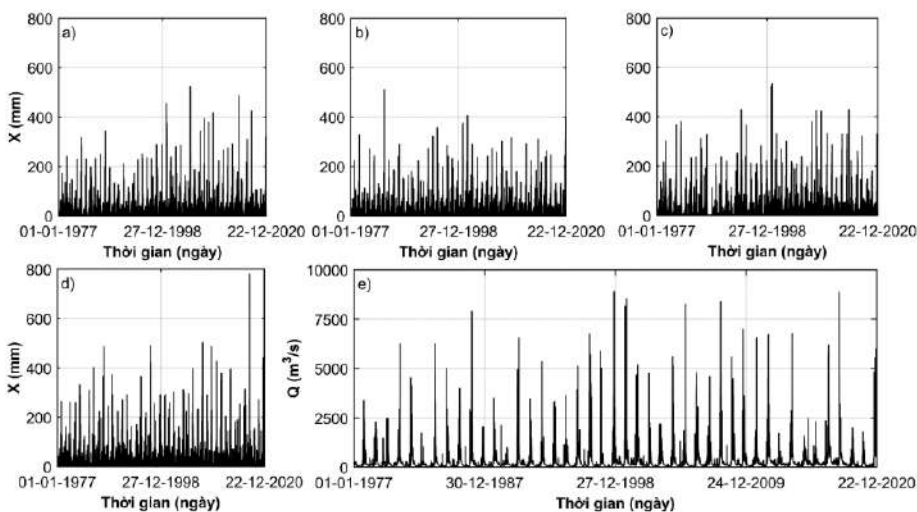
Hình 1: Bản đồ lưu vực sông Thu Bồn tính đến trạm thủy văn Nông Sơn

Do tác động của biến đổi khí hậu kết hợp ảnh hưởng của các hồ chứa thủy điện và nhiều hoạt động phát triển kinh tế xã hội, dòng chảy trên lưu vực đã có những thay đổi ngày càng khó đánh giá và dự báo. Hậu quả của sự thay đổi nguồn nước đã tác động mạnh mẽ đến kinh tế, xã hội và môi trường [6]. Cụ thể, về kinh tế, việc thiếu nước thường xuyên sẽ làm giảm năng suất, sản lượng cây trồng và từ đó thu nhập của nông dân cũng giảm đáng kể. Về xã hội, thiếu nước sẽ thể gây ra tranh chấp giữa các ngành, nghề, lĩnh vực sử dụng nhiều nước với các ngành, nghề, lĩnh vực sử dụng ít nước hay tạo ra sự bất bình

đẳng trong phân phối nguồn nước giữa các yêu cầu sử dụng nước. Về môi trường, thiếu nước trong thời gian dài sẽ dẫn đến tình trạng hạn hán, từ đó làm tăng khả năng xói mòn đất, nguy cơ suy thoái cảnh quan và cả rủi ro cháy rừng. Do đó, để đảm bảo quá trình phát triển kinh tế xã hội bền vững, tính toán định lượng dòng chảy và dự báo chính xác nguồn nước sử dụng các công cụ hiện đại (như trí tuệ nhân tạo, học máy, học sâu) là việc làm hết sức cần thiết cho xây dựng kế hoạch phân bổ, điều tiết, khai thác và sử dụng nguồn nước phù hợp.

## 2.2. Thu thập dữ liệu

Số liệu mưa ngày thực đo tại bốn trạm (Nông Sơn, Hiệp Đức, Tiên Phước và Trà My) và lưu lượng dòng chảy ngày tại Nông Sơn (Hình 1) trên lưu vực sông Thu Bồn trong giai đoạn từ 1977-2020 (Hình 2) đã được thu thập để phục vụ cho các mục đích tính toán của nghiên cứu này. Chi tiết về các đặc trưng thống kê của lượng mưa và lưu lượng dòng chảy tại các trạm nêu trên được tổng hợp trong Bảng 1. Lưu ý rằng chuỗi số liệu mưa tại các trạm sẽ được sử dụng là dữ liệu đầu vào trong mô hình bộ nhớ dài ngắn, trong khi số liệu dòng chảy thực đo tại trạm Nông Sơn sẽ được sử dụng để so sánh đánh giá kết quả đầu ra của mô hình.



Hình 2: Chuỗi số liệu mưa ngày tại: a) Hiệp Đức, b) Nông Sơn, c) Tiên Phước, d) Trà My và e) lưu lượng dòng chảy ngày trạm Nông Sơn

**Bảng 1: Dữ liệu các trạm mưa và dòng chảy thu thập trong lưu vực nghiên cứu**

Tên trạm	Dữ liệu	Thời kỳ thu thập	Giá trị trung bình	Khoảng biến động	Hệ số tương quan
Hiệp Đức	Mưa, X (mm)	1977-2020	8,27	0 ÷ 527	0,593
Nông Sơn			8,11	0 ÷ 513	0,549
Tiên Phước			8,63	0 ÷ 534	0,594
Trà My			11,32	0 ÷ 782	0,669
Nông Sơn	Lưu lượng, Q (m <sup>3</sup> /s)		291,33	4 ÷ 8920	1,0

### 3. PHƯƠNG PHÁP NGHIÊN CỨU

#### 3.1. Mô hình LSTM

Mô hình LSTM là cải tiến của mạng thần kinh hồi quy nhằm giải quyết vấn đề học, lưu trữ thông tin [2]. Mô hình LSTM cũng có cấu trúc mắt xích tương tự như mạng thần kinh hồi quy, nhưng các module lặp có cấu trúc khác. Cụ thể, thay vì chỉ có một lớp mạng nơ-ron thì LSTM có tới bốn lớp, tương tác với nhau theo một cấu trúc cụ thể. Ý tưởng cốt lõi của mô hình LSTM chính là trạng thái nhớ thông qua các nơ-ron. Trạng thái nhớ giống như băng chuyền, chạy xuyên suốt toàn bộ các mắt xích (hay còn gọi là các nút mạng) và chỉ có một số tương tác tuyến tính được thực hiện. Điều này giúp cho thông tin ít bị thay đổi xuyên suốt quá trình truyền.

Mô hình LSTM có khả năng thêm hoặc bớt thông tin vào trạng thái nhớ, được quy định một cách cẩn thận bởi các cấu trúc gọi là cổng (gate). Các cổng này là một cách tùy chọn để định nghĩa thông tin truyền qua thông qua các hàm *sigmoid* và một toán tử nhân *pointwise*. Hàm kích hoạt *sigmoid* có giá trị từ 0 đến 1, mô tả thông tin được phép truyền qua tại mỗi lớp mạng. Nếu thu được giá trị 0 điều này có nghĩa là không cho bất kỳ thông tin nào truyền qua, ngược lại nếu thu được giá trị là 1 thì có nghĩa là cho phép mọi thông tin truyền qua. Mô hình LSTM có ba cổng như trên để bảo vệ và điều khiển trạng thái của nơ-ron. Quá trình hoạt động của mô hình LSTM được thông qua bốn

bước cơ bản [2].

**Bước 1:** Bước đầu tiên của mô hình LSTM là quyết định xem thông tin nào cần loại bỏ khỏi trạng thái của nơ-ron. Quyết định này được thực hiện thông qua một lớp với hàm *sigmoid* gọi là cổng chặn (forget gate layer). Đầu vào là  $h_{t-1}$  và  $x_t$  đầu ra là một giá trị nằm trong khoảng  $[0, 1]$  cho trạng thái  $C_{t-1}$ . Giá trị 1 tương đương với giữ lại thông tin, giá trị 0 tương đương với loại bỏ thông tin. Cổng chặn  $f_t$  có thể được tính theo phương trình (1).

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \quad (1)$$

**Bước 2:** Khi cần quyết định thông tin nào cần được lưu lại tại trạng thái nơ-ron. Việc này bao gồm hai thành phần (hay lớp): (i) lớp sử dụng hàm *sigmoid* được gọi là lớp cổng vào (input gate layer) để quyết định các giá trị nào sẽ cập nhật và cổng vào  $i_t$  được tính theo phương trình (2) và (ii) lớp sử dụng hàm *tanh* để tạo ra một vector cho giá trị mới  $\tilde{C}_t$  được thêm vào trong trạng thái.  $\tilde{C}_t$  được tính toán bởi phương trình (3).

$$i_t = \sigma(W_i(h_{t-1}, x_t) + b_i) \quad (2)$$

$$\tilde{C}_t = \tanh \tanh (W_c[h_{t-1}, x_t] + b_c) \quad (3)$$

**Bước 3:** Ở bước này sẽ kết hợp hai thành phần ở bước 2 lại để cập nhật vào trạng thái, từ trạng thái cũ  $C_{t-1}$  thành trạng thái mới  $C_t$ . Mô hình LSTM sẽ nhân trạng thái cũ với hàm  $f_t$  để bỏ đi những thông tin quyết định quên

trước đó. Sau đó, sẽ cộng thêm  $i_t * \tilde{C}_t$ . Đây là giá trị mới mà mô hình muốn cập nhật cho mỗi trạng thái. Trạng thái mới thu được này phụ thuộc vào việc quyết định cập nhật mỗi giá trị trạng thái ra sao. Trạng thái mới  $C_t$  có thể được tính bởi phương trình (4).

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (4)$$

**Bước 4:** Trong mô hình LSTM cần quyết định thông tin đầu ra. Đầu ra này cần dựa trên trạng thái và thường sẽ được lọc bớt thông tin. Đầu tiên, áp dụng lớp hàm *sigmoid* để quyết định xem phần nào của trạng thái dự định sẽ đưa ra. Sau đó, sẽ đẩy trạng thái qua hàm *tanh* (giá trị trong khoảng -1 và 1) và nhân với một cổng đầu ra với hàm *sigmoid* (output sigmoid gate) để giữ lại những phần muốn đưa ra ngoài.

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \quad (5)$$

$$h_t = o_t * \tanh(C_t) \quad (6)$$

Mô hình LSTM sử dụng các công thức nêu trên và lặp lại tại từng bước thời điểm  $t$ . Thông tin của các trạng thái được điều khiển bởi cấu trúc các cổng chặn  $f_t$ , cổng vào  $i_t$  và cổng ra  $o_t$ . Trong đó cổng chặn  $f_t$  chính là tư tưởng chủ đạo của mô hình LSTM khi cho phép điều khiển lượng thông tin đầu vào  $h_{t-1}$  từ các thời điểm trước. Cổng ra  $o_t$  và

trạng thái ẩn  $h_t$  có thể được tính theo phương trình (5) và (6). Các thông số chính của mô hình LSTM có thể kể đến là tỷ lệ học, số lượng lớp ẩn, số epochs. Giá trị thích hợp của các thông số trên sẽ được xác định dựa vào chuỗi số liệu dòng chảy thực đo trên lưu vực nghiên cứu.

### 3.2. Các chỉ tiêu đánh giá sai số

Bốn chỉ tiêu sai số, bao gồm sai số quân phương (*RMSE*), sai số trung bình tuyệt đối (*MAE*), hệ số tương quan ( $r$ ) và hệ số *NSE* (Nash–Sutcliffe efficiency) đã được sử dụng để đánh giá sự phù hợp giữa lưu lượng dòng chảy tính toán và thực đo. Sai số quân phương và sai số tuyệt đối trung bình biểu thị độ lớn trung bình của sai số. Sai số quân phương xét đến bình phương của sai số nên khi sai số lớn thì khi bình phương sẽ càng lớn hơn vì vậy mà sai số sẽ thiên lớn trong khi đó các giá trị thiên nhỏ có thể bị bỏ qua (sai số càng nhỏ bình phương sẽ càng nhỏ) [3, 4]. Chính vì thế, sai số tuyệt đối trung bình được sử dụng để bổ sung cho sai số quân phương. Hệ số tương quan  $r$  và *NSE* càng gần 1 thì kết quả tính toán càng chính xác.

### 3.3. Thiết lập các tính toán

**Bảng 2:** Các trường hợp tính toán khi sử dụng mô hình LSTM

Trạm mưa Kí hiệu	Hiệp Đức	Nông Sơn	Tiên Phước	Trà My
LSTM(HĐ)	✓			
LSTM(NS)		✓		
LSTM(TP)			✓	
LSTM(TM)				✓

Chuỗi số liệu mưa và dòng chảy tại các trạm được sử dụng để thực hiện các tính toán mô phỏng khác nhau khi sử dụng mô hình bộ nhớ dài ngắn LSTM. Cụ thể, chuỗi số liệu trong khoảng thời gian từ ngày 1/1/1977 đến ngày 14/10/2007 (tương ứng với 70% chiều dài của chuỗi số liệu thu thập) đã được sử dụng để đào

tạo mô hình, trong khi chuỗi số liệu trong khoảng thời gian từ ngày 15/10/2007 đến 31/12/2020 đã được sử dụng cho kiểm định mô hình [2, 4-5]. Số liệu tại từng trạm mưa đã được sử dụng làm dữ liệu đầu vào (như thống kê trong Bảng 2). Chi tiết kết quả dòng chảy cho bước đào tạo và kiểm định mô hình LSTM ứng với

từng trạm mưa được thể hiện trong Mục 4.

## 4. KẾT QUẢ VÀ THẢO LUẬN

### 4.1. Kết quả đào tạo mô hình

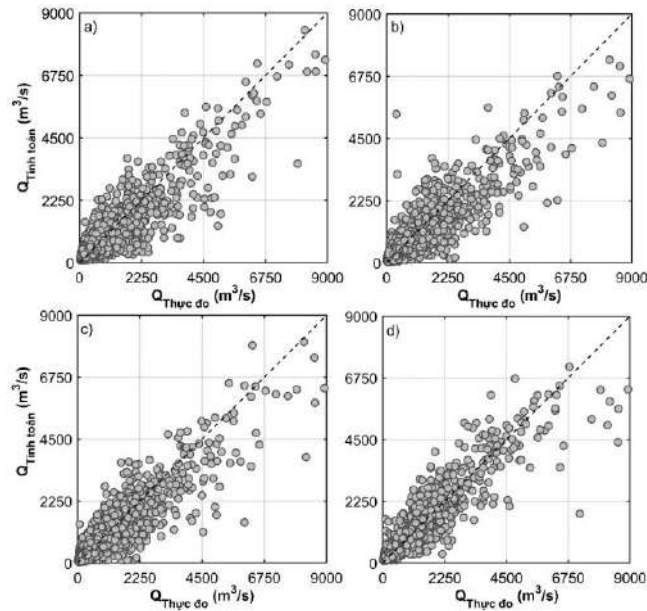
Khi sử dụng chuỗi số liệu mưa ngày tại mỗi trạm làm dữ liệu đầu vào, các thông số của mô hình LSTM được xác định theo phương pháp thử sai. Kết quả tỷ lệ học = 0,001, số lượng lớp ẩn = 128 và số epochs = 100 đã được xác định cho quá trình đào tạo mô hình. Hình 3 thể hiện kết quả so sánh giữa lưu lượng dòng chảy tính toán (ứng với các giá trị tối ưu của thông số mô hình LSTM ở trên) và dòng chảy thực đo, trong khi đó giá trị của các chỉ tiêu đánh giá sai số được thống kê chi tiết trong Bảng 3. Kết quả tính toán thể hiện rằng mô hình LSTM thể hiện rất tốt chuỗi số liệu dòng chảy thực đo trong lưu vực nghiên cứu. Hệ số tương quan  $r$  lớn hơn 0,91, trong khi hệ số  $NSE$  thay đổi từ 0,82 đến 0,88. Sai số quân phương ( $RMSE$ ) của dòng chảy dao động từ 195 đến 236 m<sup>3</sup>/s và sai số trung bình tuyệt đối ( $MAE$ ) biến đổi trong khoảng giữa 91

và 96 m<sup>3</sup>/s (Bảng 3). Giá trị của  $RMSE$  và  $MAE$  nêu trên chỉ bằng từ 1,03 đến 2,6% biên độ của dòng chảy ngày đo đạc tại trạm.

Trong 4 trạm mưa trên lưu vực, mô hình LSTM cho kết quả mô phỏng chuỗi dòng chảy thực đo tốt nhất ứng với trường hợp sử dụng số liệu mưa tại trạm Trà My. Hệ số tương quan  $r$  và hệ số  $NSE$  lần lượt bằng 0,94 và 0,88, trong khi giá trị  $RMSE$  và  $MAE$  của dòng chảy nhỏ hơn 2,2% biên độ của lưu lượng dòng chảy thực đo quan trắc tại Nông Sơn. Các kết quả tính toán trên cũng phù hợp với hệ số tương quan giữa mưa tại Trà My và lưu lượng dòng chảy tại Nông Sơn ( $r = 0.669$ ) thống kê trong Bảng 1. Mặt khác, mô hình LSTM cho kết quả mô phỏng lưu lượng dòng chảy thực đo ít chính xác hơn khi sử dụng số liệu mưa tại Nông Sơn làm dữ liệu đầu vào của mô hình. Điều đó thể hiện rằng mưa tại trạm Nông Sơn không phản ánh chính xác đặc trưng mưa trên lưu vực nghiên cứu.

**Bảng 3: Giá trị của các chỉ tiêu đánh giá sai số cho đào tạo mô hình LSTM**

Các chỉ tiêu sai số Kí hiệu	$r$	$NSE$	$RMSE$		$MAE$	
			(m <sup>3</sup> /s)	(%)	(m <sup>3</sup> /s)	(%)
LSTM(HĐ)	0,92	0,85	217,77	2,44	96,06	1,08
LSTM(NS)	0,91	0,82	235,84	2,64	94,16	1,06
LSTM(TP)	0,92	0,85	219,01	2,46	94,55	1,06
LSTM(TM)	0,94	0,88	195,00	2,19	91,76	1,03



Hình 3: Lưu lượng dòng chảy tính toán và thực đo: a) LSTM(HĐ), b) LSTM(NS), c) LSTM(TP) và d) LSTM(TM) cho đào tạo mô hình

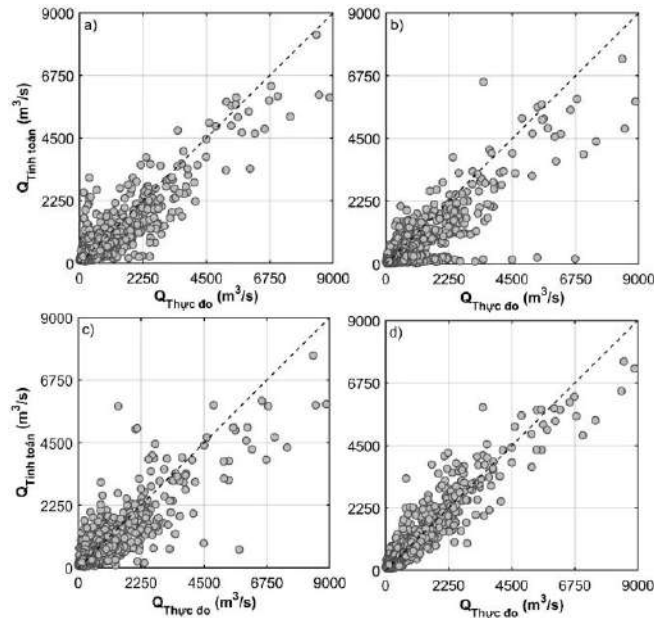
#### 4.2. Kết quả kiểm định mô hình

Để kiểm định mô hình LSTM, chuỗi số liệu mưa và dòng chảy trong khoảng thời gian từ ngày 15/10/2007 đến 31/12/2020 (tương ứng với 30% chiều dài của chuỗi số liệu thu thập) đã được sử dụng. Giá trị tối ưu của các thông số xác định trong bước đào tạo cũng được sử dụng trong mô

phỏng. Kết quả so sánh lưu lượng dòng chảy tính toán và thực đo khi sử dụng số liệu mưa tại các trạm (Hiệp Đức, Nông Sơn, Tân Phước và Trà My) cho bước kiểm định mô hình LSTM được thể hiện như trên Hình 4, trong khi giá trị của các chỉ tiêu đánh giá sai số được thống kê chi tiết trong Bảng 4.

**Bảng 4: Giá trị của các chỉ tiêu đánh giá sai số cho kiểm định mô hình LSTM**

Các chỉ tiêu sai số Kí hiệu	$r$	NSE	RMSE		MAE	
			(m <sup>3</sup> /s)	(%)	(m <sup>3</sup> /s)	(%)
			LSTM(HĐ)	0,91	0,82	253,91
LSTM(NS)	0,87	0,73	312,26	3,52	137,26	1,55
LSTM(TP)	0,89	0,79	280,12	3,15	119,31	1,34
LSTM(TM)	0,95	0,90	195,64	2,20	92,96	1,05



Hình 4: Lưu lượng dòng chảy tính toán và thực đo: a) LSTM(HĐ), b) LSTM(NS), c) LSTM(TP) và d) LSTM(TM) cho kiểm định mô hình

Kết quả kiểm định thể hiện rằng mô hình LSMT tái hiện khá tốt lưu lượng dòng chảy thực đo tại cửa ra của lưu vực nghiên cứu trong khoảng thời gian từ ngày 15/10/2007 đến 31/12/2020 khi sử dụng số liệu mưa tại các trạm khác nhau trên lưu vực. Cụ thể, hệ số tương quan  $r$  thay đổi từ 0,87 đến 0,95, thể hiện rằng mô hình đã thể hiện rất tốt xu hướng thay đổi của chuỗi lưu lượng dòng chảy thực đo theo thời gian. Hệ số  $NSE$  dao động từ 0,73 đến 0,90. Giá trị của sai số quân phương ( $RMSE$ ) thay đổi từ 195 đến 313  $m^3/s$  (tương ứng bằng từ 2,2 đến 3,52% biên độ lưu lượng dòng chảy thực đo). Sai số tuyệt đối trung bình ( $MAE$ ) thay đổi từ 93 đến 138  $m^3/s$  và sai số trên nhỏ hơn 1,55% lưu lượng dòng chảy ngày lớn nhất thực đo tại Nông Sơn. Đồng thời, kết quả đào tạo và kiểm định mô hình đều thể hiện rằng khi sử dụng số liệu mưa tại trạm Trà My làm dữ liệu đầu vào, mô hình LSTM cho kết quả mô phỏng lưu lượng dòng chảy thực đo tốt nhất so với việc sử dụng số liệu mưa của trạm Nông Sơn hoặc Hiệp Đức hoặc Tiên Phước. Điều đó có nghĩa rằng số liệu mưa tại trạm Trà My hoàn toàn có thể được sử dụng để tính toán mô phỏng dòng chảy trên lưu vực nghiên cứu

trong trường hợp số liệu mưa bị hạn chế, khi sử dụng mô hình LSTM.

Như thể hiện trên Hình 3 và Hình 4, sự khác biệt giữa lưu lượng dòng chảy tính toán và thực đo vẫn còn tồn tại, đặc biệt là ứng với các trường hợp lưu lượng dòng chảy lớn. Các kết quả tính toán lưu lượng dòng chảy lớn nhất từ mô hình LSTM vẫn đang thiên nhỏ so với thực đo, bất kể việc sử dụng một trong bốn trạm mưa trên lưu vực. Nguyên nhân dẫn đến hiện tượng trên là do giả thiết mưa tại các trạm rời rạc trên là lượng mưa trên toàn lưu vực. Do đó, để cải thiện kết quả tính toán lưu lượng dòng chảy trên lưu vực nghiên cứu khi sử dụng mô hình LSTM, lượng mưa trung bình lưu vực (theo đa giác Thiessen hay trung bình số học) nên được sử dụng. Vấn đề trên sẽ được xem xét, thực hiện trong nghiên cứu tiếp theo khi sử dụng mô hình LSTM cho mô phỏng và dự báo dòng chảy trên lưu vực nghiên cứu phục vụ cho quy hoạch, điều tiết, khai thác và sử dụng bền vững nguồn nước trên lưu vực nghiên cứu.

## 5. KẾT LUẬN

Mô hình LSTM và chuỗi số liệu mưa tại bốn trạm (Nông Sơn, Hiệp Đức, Tiên Phước và Trà My) đã được sử dụng để tính toán lưu lượng



dòng chảy tại cửa ra của lưu vực sông Thu Bồn tính đến trạm thủy văn Nông Sơn. Dựa trên các kết quả đã trình bày, một số kết luận chính của nghiên cứu như sau:

(i) Mô hình LSTM tái hiện rất tốt chuỗi số liệu lưu lượng dòng chảy ngày thực đo trong thời kỳ nhiều năm từ 1/1/1977 đến 31/12/2020. Hệ số tương quan  $r$  thay đổi từ 0,87 đến 0,95, trong

khi hệ số  $NSE$  dao động từ 0,73 đến 0,90. Giá trị của  $RMSE$  và  $MAE$  nhỏ hơn 3,52% biên độ lưu lượng dòng chảy ngày thực đo.

(ii) Khi sử dụng mô hình LSTM, dữ liệu mưa tại trạm Trà My cho các kết quả mô phỏng chuỗi lưu lượng dòng chảy ngày thực đo tại Nông Sơn tốt nhất, với hệ số tương quan  $r$  lớn hơn 0,94 và hệ số  $NSE$  lớn hơn 0,88.

## TÀI LIỆU THAM KHẢO

- [1] Rajae T. & Jafari H. (2020). Two decades on the artificial intelligence models advancement for modeling river sediment concentration: state-of-the-art. *Journal of Hydrlogy*, 588, 125011.
- [2] Hochreiter S. & Schmidhuber J. (1997). Long short-term memory. *Neural Computation*, 9, 1735–1780.
- [3] Hien, L. X., Hùng, H. V., Lee, G. (2018). Xây dựng mô hình mạng nơ-ron hồi quy dựa trên phần mềm mã nguồn mở để dự báo lưu lượng dòng chảy. *Hội nghị Khoa học thường niên trường Đại học Thủy lợi 2018*, 18/11, 560-562.
- [4] Pham Van C. & Le H. (2022). Estimation of the daily flow in river basins using the data-driven model and traditional approaches: an application in the Hieu river basin, Vietnam. *Water practice and technology*, 18(1), 166-230.
- [5] Nguyen H.D, Pham Van C., Nguyen Q-H., Bui Q-T (2023). Daily streamflow prediction based on the long short-term memory algorithm: a case study in the Vietnamese Mekong Delta. *Journal of Water and Climate Change*, 14(4), 1247-1267.
- [6] Vũ Thị Thu Lan và ccs (2019). Cân bằng nước lưu vực sông Vu Gia - Thu Bồn bằng mô hình Mike hydro basin. *Tạp chí Khí tượng Thủy văn*, số 708, 1–12.