

ỨNG DỤNG MÔ HÌNH HỌC MÁY RANDOM FOREST MÔ PHỎNG DÒNG CHẢY CỦA LƯU VỰC SÔNG VỆ-QUẢNG NGÃI

Trần Tuấn Thạch

Trường Đại học Thủy lợi

Tóm tắt: Tài nguyên nước của lưu vực sông có vai trò thiết yếu đối với sự phát triển kinh tế-xã hội. Việc mô phỏng xác định dòng chảy trên sông do đó rất cần thiết nhằm đưa ra chiến lược quy hoạch phân bổ tài nguyên nước một cách bền vững cũng như có kế hoạch ứng phó với các thiệt hại do lũ gây ra. Trong nghiên cứu này, mô hình học máy Random Forest (RF) được ứng dụng để mô phỏng dòng chảy trên lưu vực sông Vệ tỉnh Quảng Ngãi dựa trên nền nguồn ngữ lập trình Python. Tài liệu khí tượng thủy văn của các trạm đo trong lưu vực sông Vệ sử dụng trong nghiên cứu được quan trắc từ năm 1981÷2022. Số liệu quan trắc để mô phỏng được chia làm 2 phần, 70% số liệu mẫu (từ năm 1981÷2010) được dùng để đào tạo mô hình và 30% số liệu (từ năm 2011÷2022) được dùng làm kiểm định mô hình. Mô hình RF được xây dựng để mô phỏng dòng chảy trên sông Vệ tại trạm An Chi với trường hợp sử dụng lượng mưa như yếu tố đầu vào tại trạm khí tượng Ba Tơ và trạm An Chi. Dựa trên kết quả mô phỏng cho thấy: 1) Mô hình RF cho kết tính toán lưu lượng có độ chính xác cao so với số liệu quan trắc ($R^2 \geq 0,77$; $RMSE \leq 96,6 \text{ m}^3/\text{s}$; $MAPE \leq 3,91\%$); 2) Kết quả mô phỏng lưu lượng sử dụng trạm đo mưa Ba Tơ có mức độ chính xác hơn so với sử dụng mưa của trạm An Chi. Trong đó, giai đoạn đào tạo mô hình RF với trạm đo mưa Ba Tơ ($R^2=0,90$; $RMSE=63,4 \text{ m}^3/\text{s}$; $MAPE=1,84\%$) cao hơn so với mô phỏng của mô hình RF cho trạm đo mưa An Chi ($R^2=0,84$; $RMSE=85,4 \text{ m}^3/\text{s}$; $MAPE=1,97\%$). Tương tự, trong giai đoạn kiểm định mô hình RF cũng cho thấy xu hướng tương tự với độ chính xác mô phỏng của trạm mưa Ba Tơ cao hơn trạm mưa An Chi, các chỉ số đánh giá độ chính xác của mô hình lần lượt là ($R^2=0,85$; $RMSE=78,4 \text{ m}^3/\text{s}$; $MAPE=3,51\%$) và ($R^2=0,77$; $RMSE=96,6 \text{ m}^3/\text{s}$; $MAPE=3,91\%$).

Từ khóa: Dòng chảy trên lưu vực sông, RF, lưu vực sông Vệ.

Summary: Water resources in river basins play an indispensable role in human socio-economic development. Therefore, simulating river flows is necessary to develop a sustainable water resources analysis strategy and plan for flood damage. Based on Python programming, this study adopted the Random Forest (RF) model to simulate the daily streamflow in the Ve River basin, Quang Ngai Province. Hydrometeorological data of stations used in the study were observed from the years 1981÷2022. The observed data for simulation is divided into 2 components, the first one is that 70% of the sample data (from 1981÷2010) was used to train the model, and the second is that 30% of the data (from 2011÷2022) was used to test the model. The RF model was built to simulate the streamflow on the Ve River at An Chi station with rainfall scenarios as an input factor at the Ba To meteorological station and An Chi station. Based on the simulated results, it concludes several highlight points: 1) The RF model gives highly accurate calculated results compared to the observed results at An Chi station ($R^2 \geq 0,77$; $RMSE \leq 96,6 \text{ m}^3/\text{s}$; $MAPE \leq 3,91\%$); 2) Simulation using Ba To rain gauge station has a more accurate simulation speed than using rain at An Chi point, in which the RF model training duration for the Ba To station ($R^2=0,90$; $RMSE=63,4 \text{ m}^3/\text{s}$; $MAPE=1,84\%$) is higher than that for An Chi station ($R^2=0,84$; $RMSE=85,4 \text{ m}^3/\text{s}$; In the RF model for validating period, a similar trend is also represented acceptable the simulated streamflow of Ba To station higher than that of An Chi station with the accuracy model criteria indices for validated period, being ($R^2=0,85$; $RMSE=78,4 \text{ m}^3/\text{s}$; $MAPE=3,51\%$) and ($R^2=0,77$; $RMSE=96,6 \text{ m}^3/\text{s}$; $MAPE=3,91\%$).

Keywords: Daily streamflow, RF, Ve River Basin.

1. GIỚI THIỆU CHUNG

Dòng chảy trên các lưu vực sông đóng góp vai trò rất quan trọng cho các hoạt động kinh tế

của con người, đặc biệt đối với phát triển sản xuất nông nghiệp. Do đó, việc xác định dòng chảy trên sông có ý nghĩa lớn trong việc quản lý phân bổ tài nguyên nước một cách hợp lý. Để xác định dòng chảy hiện nay có nhiều phương thức khác nhau như: Quan trắc, ứng dụng các mô hình thủy lực, thủy văn, ...vv.

Ngày nhận bài: 05/11/2025

Ngày thông qua phản biện: 26/12/2025

Ngày duyệt đăng: 20/01/2026

Tuy nhiên, việc ứng dụng các mô hình thủy văn, thủy lực thương mại như: MIKE NAM, HEC-RAS, ...vv đòi hỏi số liệu đo đạc về (địa hình, địa chất, ...vv) đầy đủ nên thường khó khăn trong việc mô phỏng khi nguồn kinh phí hạn hẹp, đặc biệt đối với lưu vực sông lớn. Bên cạnh các mô hình vật lý này được xây dựng dựa trên quan hệ giữa mưa và dòng chảy với các đặc trưng của riêng từng lưu vực thì các mô hình sử dụng các thuật toán học sâu (Deep learning-DL) và học máy (Machine learning-ML) đã được ứng dụng trong nhiều lĩnh vực, trong đó có thủy văn và thủy lực nhằm mô phỏng, dự đoán và dự báo các yếu tố khí tượng, thủy văn của các lưu vực sông nghiên cứu.

Các mô hình học máy (ML) có thể chia thành 4 nhóm chính bao gồm: Học giám sát (Supervised Learning); học không giám sát (Unsupervised Learning); học nửa giám sát (Semi-supervised Learning); học tăng cường (Reinforcement Learning). Trong lĩnh vực tài nguyên nước hiện nay đã có nhiều ứng dụng sử dụng ML để mô phỏng và dự báo về các yếu tố khí hậu và thủy văn như: mưa, lưu lượng, mực nước, ...vv. Tác giả Patel sử dụng mô hình ANN (Artificial Neural Network) để dự đoán dòng chảy theo thời gian của lưu vực thượng nguồn sông Cauvery [1]. Trong một nghiên cứu khác, Peiman Parisouj đã ứng dụng thật toán ANN dạng cải tiến ANN-BP (Artificial Neural Network with backpropagation) và so sánh với các mô hình học máy khác (SVR-support Vector Regression, ELM-Extreme Learning Machine) để dự đoán dòng chảy cho 4 lưu vực sông lớn của Hoa Kỳ [2]. Kourosh Mohammad lại sử dụng mô hình ARMA (Autoregressive moving average) mô phỏng dòng chảy từ chuỗi số liệu 68 năm của sông Karun [3]. Tương tự, tác giả Maha Shabbir sử dụng các mô hình ML khác nhau, trong đó có ARMA như là công cụ hữu hiệu trong việc dự báo lưu lượng dòng chảy [4]. Bên cạnh các thuật toán ML (ANN hoặc

ARMA) được đưa vào sử dụng từ lâu trong lĩnh vực thủy văn học, các thuật toán ML cải tiến được đề xuất, trong đó có mô hình Random Forest (RF) bởi tính ưu điểm của nó. RF được áp dụng cho việc (phân loại, xây dựng phương trình hồi quy, ...vv) thông qua cây ra quyết định (Decision tree) [5]. RF là một dạng của thập toán học máy có giám sát (supervised learning) với ưu điểm linh hoạt và dễ sử dụng bằng cách tạo ra cây quyết định trên các mẫu dữ liệu được chọn ngẫu nhiên, được dự đoán từ mỗi cây và chọn giải pháp tốt nhất bằng cách bỏ phiếu quyết định. RF đã được ứng dụng nhiều trên thế giới trong lĩnh vực quản lý tài nguyên nước, đặc biệt là mô phỏng dòng chảy theo mùa. Tác giả Bing Li đã so sánh mô hình RF với ANN trong việc dự đoán mực nước của hồ Poyang ở Trung Quốc và cho thấy RF cho kết quả mô phỏng tốt hơn so với ANN [6]. Một nghiên cứu khác của tác giả Mohammad Ali Ghorbani cũng đưa ra quan điểm về tính hiệu quả trong việc dùng mô hình RF trong việc dự báo dòng chảy của lưu vực sông Dulhunty and Herbert của Úc [7]. Dựa trên các nghiên cứu về ứng dụng ML trong mô phỏng dự báo dòng chảy trên lưu vực sông thì có thể thấy mô hình RF có nhiều ưu điểm trong việc xử lý chuỗi số lớn liệu theo thời gian và mức độ mô phỏng có độ chính xác cao [8].

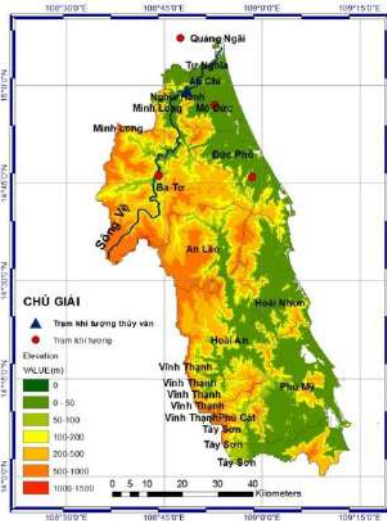
Nghiên cứu này tập trung vào phân tích và mô phỏng dòng chảy của lưu vực sông Vệ, tỉnh Quảng Ngãi sử dụng mô hình RF. Kết quả mô phỏng được so với số liệu quan trắc lưu lượng nhằm đánh giá mức độ phù hợp của mô hình.

2. PHƯƠNG PHÁP NGHIÊN CỨU

2.1 Phạm vi và đối tượng nghiên cứu

Sông Vệ chảy trên địa bàn các huyện Ba Tơ, Đức Phổ, Minh Long, Mộ Đức, Nghĩa Hành, và Tư Nghĩa của tỉnh Quảng Ngãi (**Hình 1**). Sông có diện tích lưu vực khoảng 1.263 km² và tuyến sông chính dài khoảng 90km. Dòng chính được hình thành từ khu vực núi phi Tây thuộc huyện Ba Tơ, sau đó chảy qua huyện Tư

Nghĩa, và cuối cùng đổ ra biển tại cửa Cỏ Lũy và Đức Lợi. Sông Vệ đóng vai trò quan trọng đối với các hoạt động kinh tế của tỉnh Quảng Ngãi. Nguồn nước của sông Vệ phục vụ trực tiếp cho sản xuất chủ yếu cho các hộ dân trên địa bàn các huyện Nghĩa Hành, Minh Long, Tư Nghĩa, và Mộ Đức. Số liệu về lưu lượng của sông Vệ được quan trắc thường xuyên qua các năm tại trạm thủy văn An Chỉ (diện tích lưu vực được tính từ trạm đến đầu nguồn là 814 km²).



Hình 1: Ranh giới của lưu vực Sông Vệ

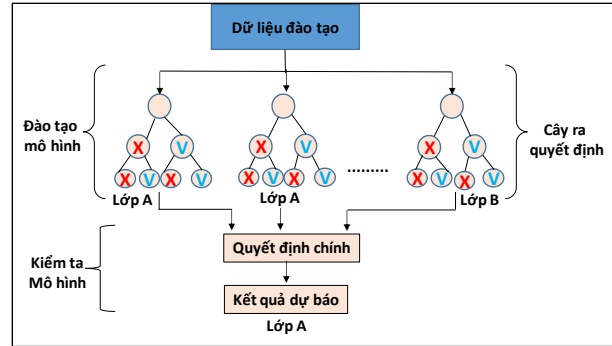
2.2 Phương pháp nghiên cứu

Mô hình Radom forest

Thập toán được xây dựng bằng cách tạo ra cây quyết định (Decision trees) trong quá trình đào tạo. Mỗi cây được tạo ra bằng việc sử dụng một tập con ngẫu nhiên của tập dữ liệu để đo lường. Tổng hợp kết quả của các cây thông qua việc quyết định đối với nhiệm vụ phân loại hoặc bằng cách tính trung bình đối với hàm hồi quy [9], sơ đồ RF như sau:

Bảng 1: Các trường hợp tính toán của mô hình Random Forest

Trạm	Số liệu đầu vào (mm)	Số liệu đầu ra (m ³ /s)	Ghi chú
An Chỉ	Mưa (X)	Lưu lượng (Q)	Đào tạo (1981÷2010)
An Chỉ	Mưa (X)	Lưu lượng (Q)	Kiểm định (2011÷2022)



Hình 2: Sơ đồ thuật toán Random Forest

Các bước thực hiện mô hình RF bao gồm: 1) Chọn các mẫu ngẫu nhiên từ tập dữ liệu đã cho; 2) Thiết lập cây quyết định cho từng mẫu và nhận kết quả dự đoán từ mỗi quyết định cây; 3) Bỏ phiếu cho mỗi kết quả dự đoán; 4) Chọn kết quả được bỏ phiếu nhiều nhất là dự đoán cuối cùng. Thông số tối ưu của mô hình RF được ứng dụng để mô phỏng dòng chảy lưu vực Sông Vệ trong đào tạo và kiểm định được lựa chọn bằng phương pháp Trial and Error. Trong đó, n-stimators (số lượng các cây quyết định được sử dụng trong mô hình RF) được lựa chọn là 100.

Dữ liệu mô hình Radom forest

Số liệu lượng mưa theo ngày được thu thập tại trạm khí tượng Ba Tư và trạm An Chỉ từ năm 1981÷2022. Lưu lượng của sông Vệ được đo tại vị trí trạm khí tượng thủy văn An Chỉ trong thời gian tương ứng với quan trắc mưa. Số liệu thu thập được chia làm 2 phần, trong đó: 70% số liệu mẫu (năm 1981÷2010) dùng cho đào tạo mô hình (Training) và 30% số liệu mẫu (2011÷2022) dùng cho kiểm định mô hình RF (Validating) như trong (Bảng 1).

Trạm	Số liệu đầu vào (mm)	Số liệu đầu ra (m^3/s)	Ghi chú
Ba Tơ	Mưa (X)	Lưu lượng (Q)	Đào tạo (1981÷2010)
Ba Tơ	Mưa (X)	Lưu lượng (Q)	Kiểm định (2011÷2022)

Chỉ tiêu đánh giá mô hình

Độ chính xác của mô hình RF được đánh giá thông qua so sánh giữa số liệu lưu lượng thực đo và mô phỏng bằng các chỉ số: Hệ số tương quan (R^2); Độ lệch chuẩn của sai số bình

phương (RMSE-Root Mean Square Error); Sai số tuyệt đối trung bình (MAPE-Mean Absolute Percentage Error). Kết quả đánh giá mô hình được tính cho 2 giai đoạn khác nhau: Đào tạo và Kiểm định.



Hình 3: Lưu lượng (Discharge) và lượng mưa (Rainfall) quan trắc tại các trạm trên sông Vệ

3. KẾT QUẢ VÀ THẢO LUẬN

Số liệu quan trắc số liệu lượng mưa và lưu lượng trên lưu vực sông Vệ, tỉnh Quảng Ngãi (Hình 3, 4, 5) trong nhiều năm cho thấy: Lưu lượng bình quân trên sông tại trạm An Chi là

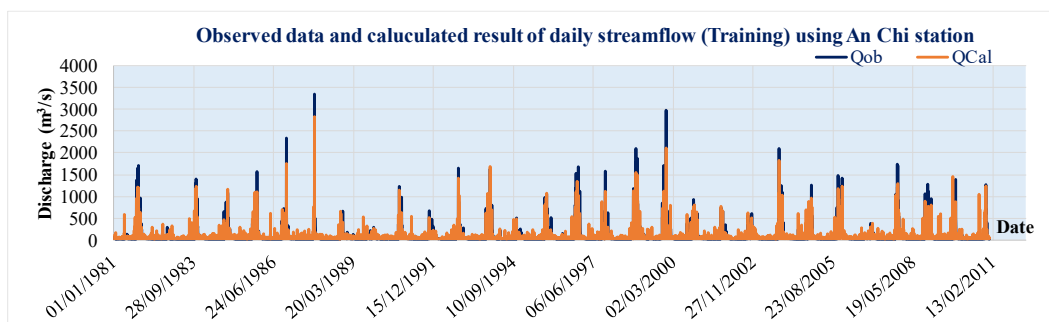
56 m^3/s ; Lưu lượng lớn nhất từng ghi nhận lên đến 3.350 m^3/s vào ngày 19/11/1987 trùng với thời điểm xuất hiện trận mưa kéo dài 2 ngày lên đến 771mm (Ba Tơ) và 549mm (An Chi). Hệ số tương quan giữa mưa và dòng chảy

quan trắc tại trạm Ba Tơ (0,76) so với trạm An Chỉ (0,67).

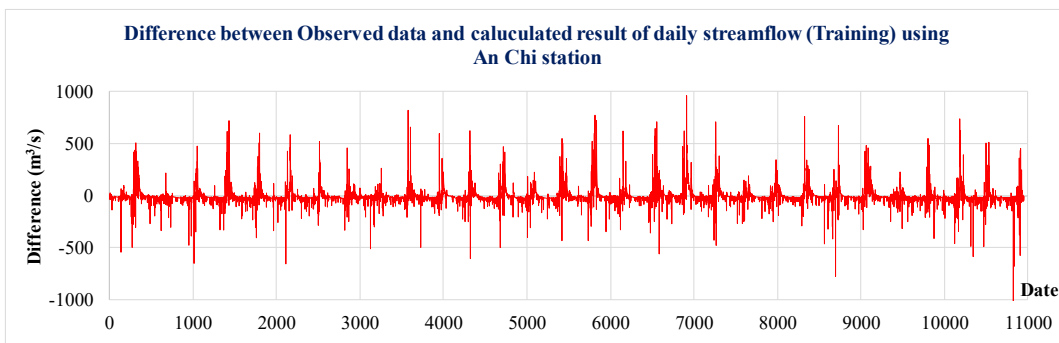
Đào tạo mô hình (Training)

Kết quả đào tạo mô hình RF ứng dụng cho lưu vực sông Vệ được thể hiện trong (Hình 4, 5, 6, 7, 12). Dòng chảy tại trạm An Chỉ được mô phỏng dựa trên số liệu mưa của 2 trạm An Chỉ và Ba Tơ đều cho kết quả có độ chính xác cao với ($R^2 > 0,8$; $RMSE < 90 \text{ m}^3/\text{s}$; $MAPE < 2\%$). Nếu so sánh mô hình RF trường hợp sử dụng số liệu của 2 trạm mưa khác nhau cho thấy:

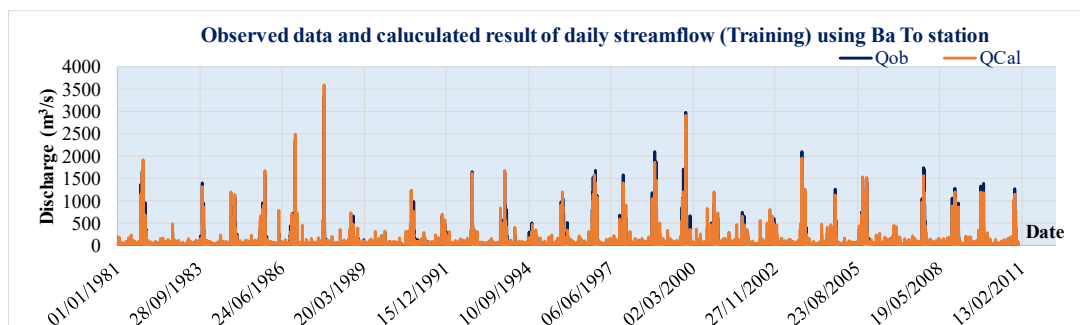
Kết quả mô phỏng mô hình sử dụng trạm đo mưa Ba Tơ làm số liệu đầu vào cho mức độ mô phỏng chính xác hơn trạm An Chỉ. Cụ thể, chỉ số mô hình RF đối với trạm Ba Tơ ($R^2=0,90$; $RMSE=63,4 \text{ m}^3/\text{s}$; $MAPE=1,84\%$) cao hơn so với mô hình RF sử dụng số liệu trạm An Chỉ ($R^2=0,84$; $RMSE=85,4 \text{ m}^3/\text{s}$; $MAPE=1,97\%$). Kết quả này phản ánh mức độ tương quan giữa mưa đo tại trạm Ba Tơ (thượng nguồn sông) so với lưu lượng đo tại trạm An Chỉ (hạ nguồn sông).



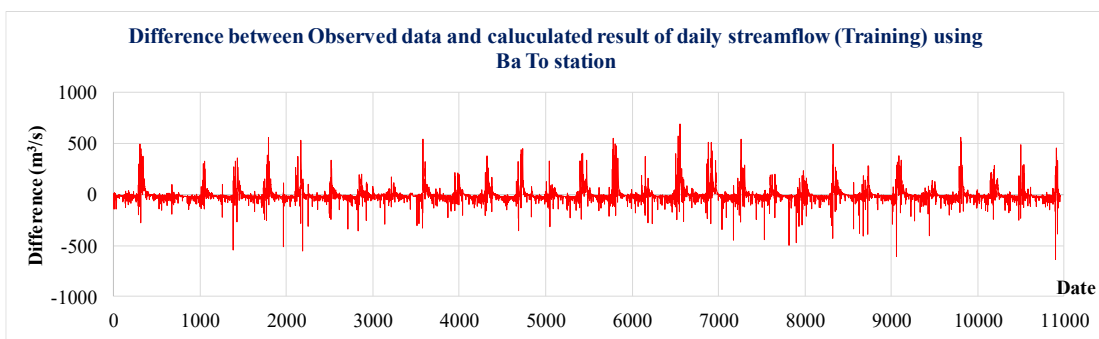
Hình 4: Lưu lượng thực đo (Q_{ob}) và mô phỏng (Q_{cal}) trạm An Chỉ cho đào tạo mô hình sử dụng lượng mưa trạm An Chỉ



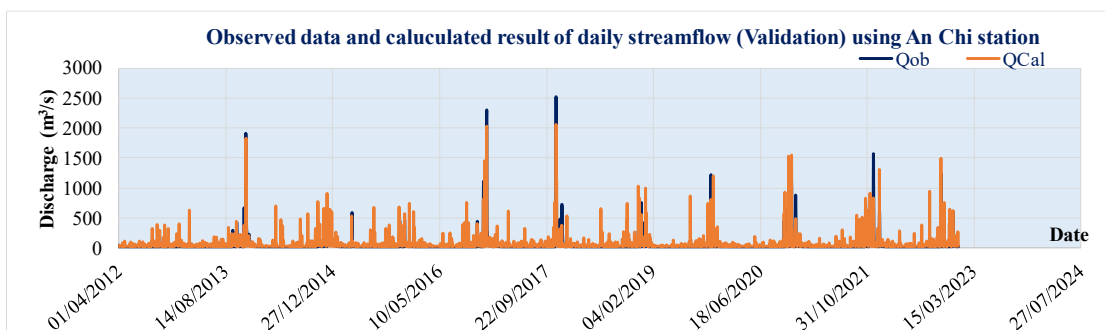
Hình 5: Sự khác biệt giữa (Q_{ob}) và mô phỏng (Q_{cal}) trạm An Chỉ cho đào tạo mô hình sử dụng lượng mưa trạm An Chỉ



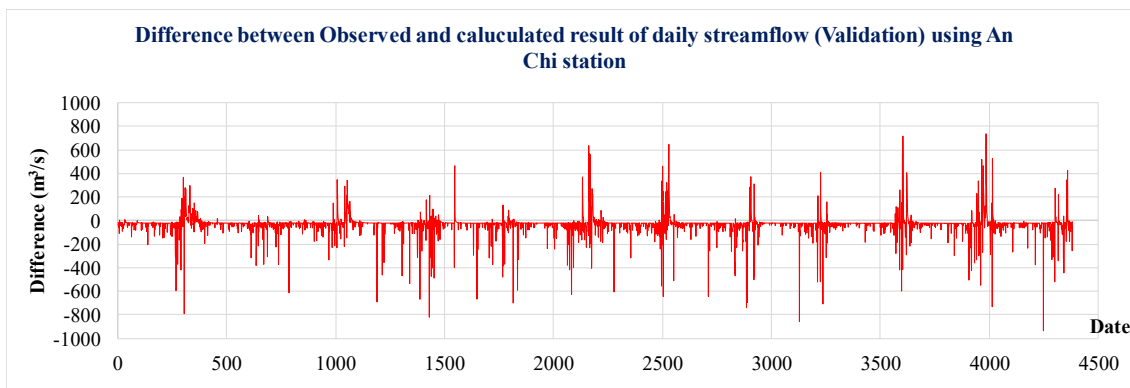
Hình 6: Lưu lượng thực đo (Q_{ob}) và mô phỏng (Q_{cal}) trạm An Chỉ cho đào tạo mô hình sử dụng lượng mưa trạm Ba Tơ



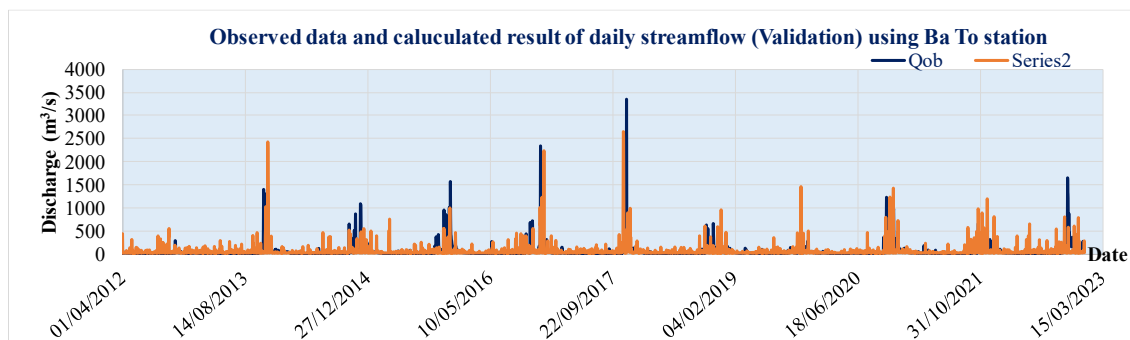
Hình 7: Sự khác biệt giữa (Q_{ob}) và mô phỏng (Q_{cal}) trạm An Chi cho đào tạo mô hình sử dụng lượng mưa trạm Ba To



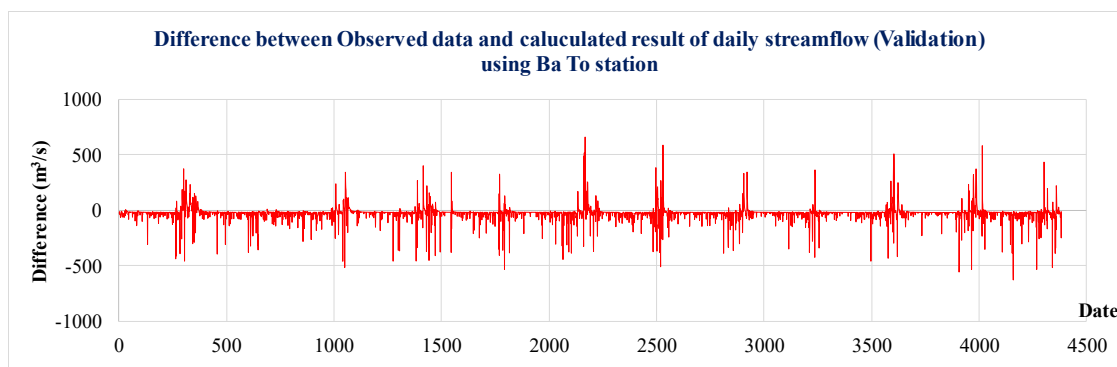
Hình 8: Lưu lượng thực đo (Q_{ob}) và mô phỏng (Q_{cal}) trạm An Chi cho kiểm định mô hình sử dụng lượng mưa trạm An Chi



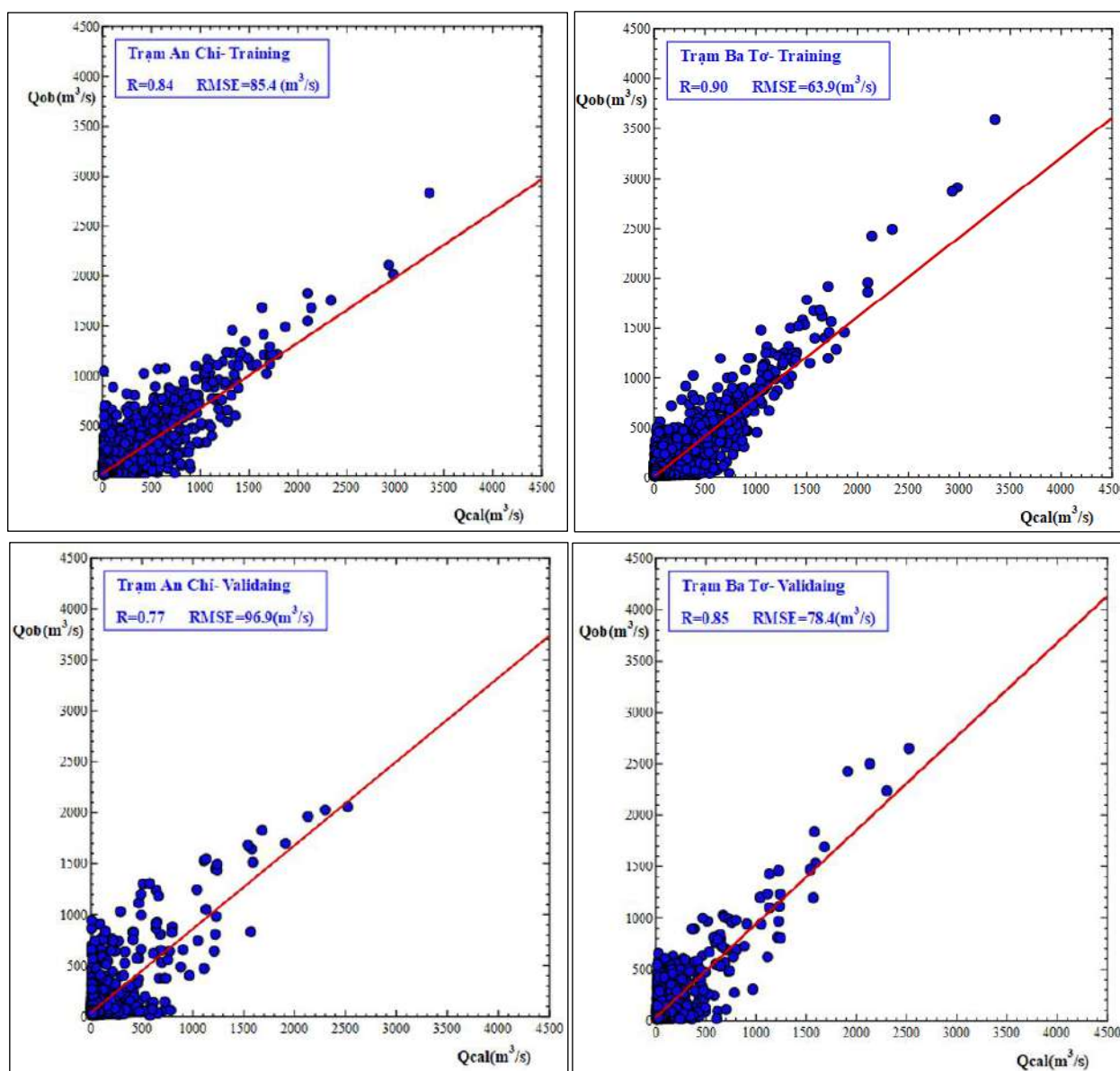
Hình 9: Sự khác biệt giữa (Q_{ob}) và mô phỏng (Q_{cal}) trạm An Chi cho kiểm định mô hình sử dụng lượng mưa trạm An Chi



Hình 10: Lưu lượng thực đo (Q_{ob}) và mô phỏng (Q_{cal}) trạm An Chi cho kiểm định mô hình sử dụng lượng mưa trạm Ba To



Hình 11: Sự khác biệt giữa (Q_{ob}) và mô phỏng (Q_{cal}) trạm An Chỉ cho kiểm định mô hình sử dụng lượng mưa trạm Ba Tơ



Hình 12: Biểu đồ phân tán (scatter plot) giữa thực đo (Q_{ob}) và mô phỏng (Q_{cal}) cho đào tạo mô hình (Training) và kiểm định mô hình (Validaing) sử dụng số liệu mưa trạm An Chỉ và Ba Tơ

Bảng 2: Chỉ tiêu đánh giá sai số cho đào tạo và kiểm định mô hình RF

TT	Trạm	Correlation (R)	Root Mean Square Error RMSE (m ³ /s)	Mean absolute percentage error) MAPE(%)
1	An Chỉ (Đào tạo)	0,84	85,4	1,97
2	Ba Tơ (Đào tạo)	0,90	63,9	1,84
3	An Chỉ (Kiểm định)	0,77	96,9	3,91
4	Ba Tơ (Kiểm định)	0,85	78,4	3,51

Kiểm định mô hình (Validation)

Để kiểm định mô hình, chuỗi số liệu dòng chảy từ năm 2011÷2022 được sử dụng trong nghiên cứu tương đương với 30% mẫu số liệu quan trắc thực tế. Tương tự so với giai đoạn đào tạo mô hình, kết quả tính toán mô phỏng (Hình 8, 9, 10, 11, 12) so với số liệu thực đo cho thấy: 1) Mô hình RF mô phỏng lưu lượng của lưu vực sông Vệ có độ chính xác tương đối cao về mặt chỉ số R^2 , RMSE, và MAPE; 2) Mô hình RF sử dụng số liệu mưa của trạm Ba Tơ cũng cho kết quả mô phỏng chính xác hơn so với việc sử dụng số liệu mưa của trạm An Chỉ. Trong đó, mô hình RF sử dụng số liệu trạm Ba Tơ mô phỏng lưu lượng có độ chính xác cao với các chỉ số ($R^2=0,85$; $RMSE=78,4\text{m}^3/\text{s}$; $MAPE=3,51\%$). Trong khi đó, mô hình RF sử dụng số liệu mưa trạm Ba Chỉ cho mô phỏng lưu lượng với chỉ số đánh giá mô hình thấp hơn ($R^2=0,77$; $RMSE=96,6\text{m}^3/\text{s}$; $MAPE=3,91\%$). Mặc dù vậy, đối với bài toán mô phỏng các yếu tố thủy văn và thủy lực thì đây là kết quả có thể có độ chính xác tương đối cao. Dựa trên kết quả mô phỏng sử dụng số liệu mưa của 2 trạm cho thấy mức độ chính xác của mô hình RF phụ thuộc nhiều vào trạm khí tượng đo mưa trên lưu vực. Trong nghiên cứu này mới chỉ giới hạn đào tạo và kiểm định mô hình RF cho từng

trạm đo riêng lẻ nhưng mức độ chính xác khá cao do lưu vực sông Vệ tương đối nhỏ. Đối với các lưu vực lớn thì việc mô phỏng này cần xem xét yếu tố đầu vào là sự kết hợp các trạm đo mưa trong lưu vực với các trọng số khác nhau.

3. KẾT LUẬN

Nghiên cứu ứng dụng học máy với mô hình Random Forest (RF) để mô phỏng dòng chảy trên lưu vực sông Vệ. Quá trình mô phỏng được chia làm 2 phần: 1) số liệu quan trắc mưa và lưu lượng giai đoạn năm 1981÷2010 được dùng cho đào tạo mô hình; 2) số liệu quan trắc 2 yếu tố mưa và dòng chảy giai đoạn năm 2011÷2022 được sử dụng trong kiểm định mô hình. Để mô phỏng dòng chảy trên lưu vực sông Vệ (quan trắc tại trạm thủy văn An Chỉ) số liệu mưa đầu vào được sử dụng tại trạm Ba Tơ (thượng nguồn sông) và trạm An Chỉ (hạ nguồn và là vị trí đo lưu lượng của sông). Kết quả mô phỏng sử dụng mô hình RF xây dựng trên nền ngôn ngữ lập trình Python cho thấy:

Mô hình RF cho kết quả mô phỏng có độ chính xác cao với giá trị ($R^2 \geq 0,77$; $RMSE \leq 96,6\text{ m}^3/\text{s}$; $MAPE \leq 3,91\%$;

Mô hình RF sử dụng số liệu đầu vào là lượng mưa của trạm Ba Tơ cho kết quả mô phỏng (đào tạo và kiểm định) cao hơn so với số liệu

mưa sử dụng của trạm An Chi.

TÀI LIỆU THAM KHẢO

- [1] Patel, S.S., Ramachandran, P. A Comparison of Machine Learning Techniques for Modeling River Flow Time Series: The Case of Upper Cauvery River Basin. *Water Resour Manage* (2015). <https://doi.org/10.1007/s11269-014-0705-0>
- [2] Parisouj, P., Mohebzadeh, H. & Lee, T. Employing Machine Learning Algorithms for Streamflow Prediction: A Case Study of Four River Basins with Different Climatic Zones in the United States. *Water Resour Manage* (2020). <https://doi.org/10.1007/s11269-020-02659-5>
- [3] Kourosh Mohammadi, 2006, Parameter estimation of an ARMA model for river flow forecasting using goal programming, *Journal of Hydrology*, Volume 331, Issues 1–2, 30 November 2006, Pages 293-299, <https://doi.org/10.1016/j.jhydrol.2006.05.017>
- [4] Shabbir, M., Chand, S. & Iqbal, F. A Novel Hybrid Method for River Discharge Prediction. *Water Resour Manage* 36, 253–272 (2022). <https://doi.org/10.1007/s11269-021>
- [5] Pham Minh Hải, 2019, Khái niệm về phương pháp Random Forest trong cuộc cách mạng Machine Learning và định hướng ứng dụng trong lĩnh vực Viễn Thám, *Tạp chí khoa học và Đo đạc bản đồ* 39-3/2019, DOI: 10.54491/jgac.2019.39.344
- [6] Bing Li, 2016, Comparison of random forests and other statistical methods for the prediction of lake water level: a case study of the Poyang Lake in Chin, *Hydrology Research* (2016) 47 (S1): <https://doi.org/10.2166/nh.2016.264>
- [7] Ghorbani, M.A., Deo, R.C., Kim, S. et al. Development and evaluation of the cascade correlation neural network and the random forest models for river stage and river flow prediction in Australia. *Soft Comput* 24, 12079–12090 (2020). <https://doi.org/10.1007/s00500-019-04648-2>
- [8] Thach, T.T. Multiple data-driven approaches for estimating daily streamflow in the Kone River basin, Vietnam. *Earth Sci Inform* (2024). <https://doi.org/10.1007/s12145-024-01390-8>
- [9] Random Forest Algorithm in Machine Learning, <https://www.geeksforgeeks.org/random-forest-algorithm-in-machine-learning/>