TẠP CHÍ
KHÍ TƯỢNG THỦY VĂN

Check for updates

Bài báo khoa học

Ước tính lượng mưa sử dụng dữ liệu vệ tinh Himawari-8 dựa trên mô hình học máy Random Forest

Nguyễn Vinh Thư[1], Bùi Thị Khánh Hòa[1]*, Nguyễn Minh Cường[1], Hoàng Thị Thanh Thuật[1], Nguyễn Thị Hoàng Anh[1]

[1] Trung tâm Mạng lưới khí tượng thủy văn quốc gia, Tổng cục Khí tượng Thủy văn, Hà Nội, Việt Nam; vinhthu73@gmail.com; nguyenminhcuong30596@gmail.com; hoangthithanhthuat99@gmail.com; khanhhoa303@gmail.com; hoanganhck@gmail.com

***Tác giả liên hệ:** khanhhoa303@gmail.com; Tel.: +84–916591270

Ban Biên tập nhận bài: 5/9/2023; Ngày phản biện xong: 2/10/2023; Ngày đăng bài: 25/11/2023

Tóm tắt: Bài báo sử dụng phương pháp học máy Random Forest (RF) để nghiên cứu ước tính lượng mưa từ dữ liệu vệ tinh Himawari-8. Dữ liệu được sử dụng là 09 kênh ảnh riêng lẻ và 36 kênh ảnh tổ hợp của vệ tinh Himawari-8 và dữ liệu đo mưa bề mặt để ước tính lượng mưa cho khu vực Việt Nam. Phương pháp RF được áp dụng thử nghiệm cho 04 đợt mưa lớn xảy ra trong các năm 2019, 2020 và 2023. Kết quả nghiên cứu cho thấy rằng phương pháp RF đã mô phỏng được trọng tâm của các đợt mưa lớn. Hệ số tương quan giữa lượng mưa ước tính từ vệ tinh và lượng mưa quan trắc (R) đều đạt từ 0,8 trở lên, giá trị sai số tuyệt đối trung bình (MAE) dưới 1,1 mm và sai số bình phương trung bình gốc (RMSE) dưới 2,17 mm. Tuy nhiên phương pháp RF có xu hướng cho kết quả ước tính lượng mưa thiên thấp hơn so với lượng mưa quan trắc thực tế.

Từ khóa: Ước tính lượng mưa; Vệ tinh Himawari-8; Random Forest.

1. Giới thiệu

Độ chính xác của số liệu quan trắc lượng mưa đóng vai trò rất quan trọng trong việc khai thác, quản lý tài nguyên nước và dự báo khí tượng thủy văn. Các quan trắc bằng trạm đo mưa bề mặt thường được sử dụng để đo trực tiếp lượng mưa trên bề mặt Trái Đất. Tuy nhiên, các phép đo bằng trạm đo mưa này cũng có một số nhược điểm, chẳng hạn như chúng không có khả năng mô phỏng chính xác sự phân bố của trường lượng mưa theo không gian, vì thiết bị đo mưa bề mặt chỉ cung cấp phạm vi quan trắc theo từng vị trí điểm trong khu vực. Việc sử dụng radar để quan trắc thời tiết hiện nay cho phép ước tính lượng mưa trên các khu vực rộng hơn, nhưng phạm vi quan trắc vẫn bị giới hạn ở bề mặt đất liền và các vùng ven biển [1]. Với tốc độ phát triển nhanh chóng của các công cụ đo đạc lượng mưa gián tiếp, các quan trắc từ vệ tinh viễn thám đã trở thành một công cụ thích hợp để cải thiện những nhược điểm trên bằng cách cung cấp phạm vi quan trắc không gian đồng nhất hơn và có độ phân giải gần thời gian thực. Ngày nay, vệ tinh địa tĩnh Himawari-8 được sử dụng khá rộng rãi và phổ biến. Vệ tinh Himawari-8 được phóng vào ngày 07/10/2014 và được vận hành phát triển bởi Cơ quan Khí tượng Nhật Bản (JMA), nó bắt đầu hoạt động và phân phối dữ liệu từ ngày 07/07/2015 [2]. Vệ tinh Himawari-8 cung cấp dữ liệu từ 16 kênh ảnh bao gồm 03 kênh ảnh thị phổ (VIS), 03 kênh cận hồng ngoại (NIR) và 10 kênh ảnh hồng ngoại (IR). Quá trình hình thành và phát triển của các đám mây đối lưu gây ra mưa thường xảy ra trong thời gian ngắn và dữ liệu từ vệ tinh Himawari-8 ngoài việc có thể quan trắc bao phủ cả một khu vực rộng lớn thì nó cũng cung cấp thông tin hữu ích để ước tính định lượng mưa.

Trong những năm gần đây, đã có khá nhiều nghiên cứu ứng dụng dữ liệu vệ tinh vào việc ước tính lượng mưa. Các nghiên cứu này chủ yếu sử dụng dữ liệu của kênh ảnh IR trong dữ liệu vệ tinh theo thời gian thực để truy xuất các sản phẩm mưa dựa trên mối quan hệ lượng mưa với nhiệt độ thu được từ các kênh hồng ngoại. Từ những năm 1997, tác giả [3] đã nghiên cứu sử dụng chênh lệch nhiệt độ đỉnh mây của các kênh IR trong dữ liệu vệ tinh địa tĩnh của Nhật Bản Geostationary Meteorological Satellite-5 (GMS-5) để ước tính lượng mưa. Kết quả nghiên cứu cho thấy sai số và mức độ phân tán lượng mưa tương đối lớn nhưng lượng mưa cực đại và lượng mưa tích lũy ước tính cho kết quả tương đối tốt. Trong nghiên cứu [4–5] cũng đã sử dụng các kênh ảnh IR của vệ tinh địa tĩnh Phong Vân 2 và Himawari-8 để đưa ra bộ dữ liệu lượng mưa trên khu vực Trung Quốc và Hàn Quốc. Các đặc điểm phát xạ của mây và các tham số vật lý như độ dày quang học, độ cao đỉnh mây và pha của đám mây, nhiệt độ đỉnh mây có thể được khai thác một cách hiệu quả để xác định sự hình thành và phát triển của đám mây. Năm 2020, tác giả [1] đã trình bày về thuật toán học máy dùng để ước tính các tham số đám mây và năng lượng bức xạ bề mặt từ dữ liệu vệ tinh Himawari-8. Có thể thấy các thuật toán học máy đã trở thành một công cụ mạnh mẽ liên kết dữ liệu vệ tinh với việc ước tính lượng mưa trong những năm gần đây. Hiện nay, nhiều nghiên cứu về các kỹ thuật học máy khác nhau đã được cung cấp rộng rãi, do đó nó mở ra một hướng mới cho việc mô phỏng các mối quan hệ giữa dữ liệu vệ tinh với việc khôi phục trường lượng mưa, ví dụ như kỹ thuật sử dụng mạng thần kinh nhân tạo (ANN) [1, 6–10]; thuật toán học máy có giám sát dạng vector hỗ trợ (SVM) [11–13], cây quyết định (DT) [14–15], học sâu (DL) [16–17]. Bên cạnh đó, thuật toán rừng ngẫu nhiên (RF) [18], cũng được sử dụng rộng rãi và rất thích hợp sử dụng trong việc ước tính lượng mưa viễn thám [19–24]. RF là một thuật toán học máy đồng bộ, RF có thể xử lý hiệu quả trên các tập dữ liệu lớn và cũng có thể dễ dàng tính toán song song. Hơn nữa nó không yêu cầu đặc tả của một mô hình dữ liệu cơ bản, nó cung cấp khả năng nắm bắt các mô hình liên kết phi tuyến tính và có thể xử lý các biến dự báo có độ tương quan cao [19]. Năm 2014, tác giả [21] đã sử dụng dữ liệu vệ tinh MSG-SEVIRI và thuật toán RF để xác định cường độ mưa của các sự kiện mưa một cách liên tục (trong thời gian ban ngày, khi hoàng hôn và đêm), kết quả nghiên cứu cho thấy mô hình cho kết quả rất tốt trong việc phát hiện khu vực mưa. Trong nghiên cứu sử dụng thuật toán RF để phát hiện các sự kiện mưa lớn [22] thì RF với thời gian dự báo cố định là 30 phút và 60 phút cho kết quả hoạt động khá tốt (xác suất phát hiện tất cả các loại điều kiện thời tiết lên tới gần 90%). Tác giả [23] đã tích hợp dữ liệu Himawari-8 và sản phẩm thời tiết số (NWP) để đưa ra phân loại các điểm ảnh có mưa và không có mưa bằng phương pháp phân loại RF, sau đó mô phỏng thành công ước tính cường độ mưa.

Dữ liệu vệ tinh Himawari-8 đã được sử dụng rộng rãi trong những nghiên cứu tại Việt Nam từ năm 2015 đến nay. Điển hình là công trình nghiên cứu [25] đã ứng dụng số liệu ảnh mây vệ tinh Himawari-8 trong dự báo, cảnh báo mưa dông cho khu vực Đồng bằng sông Cửu Long. Tác giả [26] đã nghiên cứu tính toán được lượng mưa bằng việc sử dụng dữ liệu vệ tinh, radar và đo mưa tại trạm; kết quả của nghiên cứu này cũng được cho là tương đối khả quan hơn so với việc đơn thuần chỉ sử dụng một loại thông tin. Tác giả [27] đã tính toán định lượng mưa sử dụng dữ liệu vệ tinh địa tĩnh MTSAT áp dụng cho khu vực Việt Nam; mô hình tính toán mô phỏng lượng mưa là mô hình mạng ANN. Nhìn chung, kết quả chỉ ra rằng phương pháp ANN đã nâng cao được độ chính xác của việc ước lượng mưa từ vệ tinh MTSAT. Nghiên cứu [28] đã đánh giá và so sánh các dữ liệu mưa vệ tinh cho lưu vực sông Cả, kết quả cho thấy khả năng nhận diện ngày mưa và ngày không mưa là khá tốt. Tuy nhiên, tương quan về lượng mưa ngày giữa lượng mưa quan trắc và lượng mưa ước tính từ vệ tinh còn tương đối thấp. Nói chung, các kết quả nghiên cứu tại Việt Nam cũng đã khả quan và cải thiện hơn giai đoạn trước, tuy nhiên thông tin về lượng mưa đưa ra chưa đạt độ tin cậy cao, các phương pháp ước tính chủ yếu áp dụng với lượng mưa ngày mà chưa đề cập tới vấn đề chạy nghiệp vụ trong thời gian thực vì bản chất mô hình mạng ANN còn phức tạp. Vì vậy,

trong nghiên cứu này đề xuất phương pháp RF đơn giản và dễ dàng tính toán hơn trong việc ước tính lượng mưa từ vệ tinh tiệm cận thời gian thực.

Như vậy rõ ràng là các phương pháp học máy khác nhau đã được sử dụng rộng rãi trong các ứng dụng viễn thám, khí tượng và thủy văn để mô hình hóa hiệu quả các mối quan hệ phi tuyến tính giữa các yếu tố dự báo và biến khí tượng. Hơn nữa, các thuật toán học máy cũng thể hiện là rất hiệu quả khi xử lý dữ liệu đa chiều và phức tạp. Ở Việt Nam hiện nay, các nghiên cứu về phương pháp RF sử dụng trong ước tính lượng mưa từ vệ tinh vẫn còn khá mới mẻ và chưa có nhiều tác giả nghiên cứu, đề cập đến. Vì vậy, trong nghiên cứu này kỹ thuật học máy RF được sử dụng để tính toán định lượng mưa từ vệ tinh Himawari cho khu vực Việt Nam. Việc ứng dụng thuật toán RF này trong ước tính lượng mưa từ vệ tinh hy vọng sẽ thúc đẩy hiệu quả hơn nữa ứng dụng khôi phục lượng mưa trong dự báo thời tiết.

2. Số liệu và phương pháp

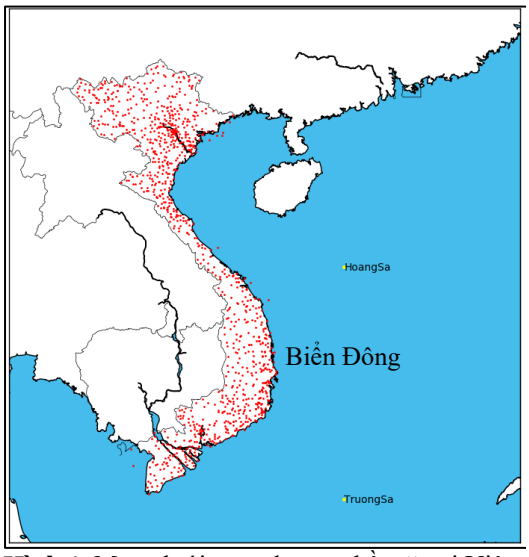
2.1. Số liệu

Số liệu vệ tinh: dữ liệu vệ tinh địa tĩnh Himawari-8, có độ phân giải không gian và thời gian cao với tần suất được ghi lại sau mỗi 10 phút, các dữ liệu quan trắc giai đoạn 2019-2020 đã được sử dụng để tạo bộ số liệu nghiên cứu trong bài báo này. Bài báo sử dụng các bước sóng trung tâm của ba kênh ảnh hơi nước WV (6,2; 6,9; và 7,3 μm), có thể thấy ba dải kênh ảnh WV này nhạy cảm với tầng đối lưu từ tầng trung đến tầng cao và các dải này cho phép ước tính sự phân bố lượng ẩm theo chiều thẳng đứng từ sự khác biệt về nhiệt độ chói giữa các dải phổ 6,2; 6,9; và 7,3 μm. Bên cạnh đó, sáu kênh ảnh hồng ngoại IR (8,6; 9,6; 10,4; 11,2; 12,4 và 13,3 μm) được sử dụng để theo dõi bức xạ IR từ các đỉnh đám mây. Sự chênh lệch nhiệt độ ΔT đại diện cho sự chênh lệch nhiệt độ chói giữa hai kênh ảnh. Để tính toán lượng mưa từ vệ tinh Himawari-8 trong nghiên cứu này sử dụng tổ hợp chênh lệch nhiệt độ chói của 2 kênh ảnh khác nhau, như đã nêu ra ở trên ta đã lựa chọn 09 kênh ảnh vậy ta sẽ có C_9^2 = 36 kênh ảnh tổ hợp và 9 kênh ảnh riêng lẻ được sử dụng và được thể hiện trong bảng 1.

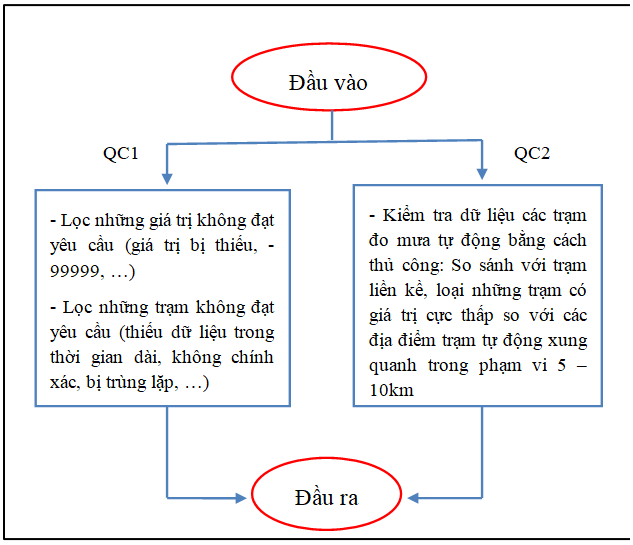
Bảng 1. Tổ hợp các kênh ảnh quan trắc bằng vệ tinh Himawari-8 được sử dụng để ước tính cường độ mưa.

μm	**6,2**	**6,9**	**7,3**	**8,6**	**9,6**	**10,4**	**11,2**	**12,4**	**13,3**
6,2	6,2	6,2-6,9	6,2-7,3	6,2-8,6	6,2-9,6	6,2-10,4	6,2-11,2	6,2-12,4	6,2-13,3
6,9		6,9	6,9-7,3	6,9-8,6	6,9-9,6	6,9-10,4	6,9-11,2	6,9-12,4	6,9-13,3
7,3			7,3	7,3-8,6	7,3-9,6	7,3-10,4	7,3-11,2	7,3-12,4	7,3-13,3
8,6				8,6	8,6-9,6	8,6-10,4	8,6-11,2	8,6-12,4	8,6-13,3
9,6					9,6	9,6-10,4	9,6-11,2	9,6-12,4	9,6-13,3
10,4						10,4	10,4-11,2	10,4-12,4	10,4-13,3
11,2							11,2	11,2-12,4	11,2-13,3
12,4								12,4	12,4-13,3
13,3									13,3

Số liệu đo mưa bề mặt: là số liệu của khoảng 2000 trạm khí tượng tự động/ đo mưa tự động trong giai đoạn 2019-2020. Trong số liệu đo mưa tự động này luôn tiềm ẩn các sai số, vì vậy cần phải thực hiện việc kiểm soát chất lượng dữ liệu này (QC) để phát hiện và loại bỏ các giá trị sai số trong dữ liệu. Sau khi lại bỏ toàn bộ những trạm không đạt yêu cầu, kết quả thu được 866 trạm đo mưa bề mặt với tần suất 10 phút trong giai đoạn 2019-2020 đạt yêu cầu để đưa vào mô hình RF (Hình 1). Hình 2 đưa ra sơ đồ các bước thực hiện của phương pháp QC để loại bỏ các trạm khí tượng tự động/ đo mưa tự động có chất lượng dữ liệu chưa tốt.



Hình 1. Mạng lưới trạm đo mưa bề mặt tại Việt Nam.



Hình 2. Sơ đồ phương pháp kiểm soát chất lượng dữ liệu trạm đo mưa bề mặt.

2.2. Phương pháp nghiên cứu

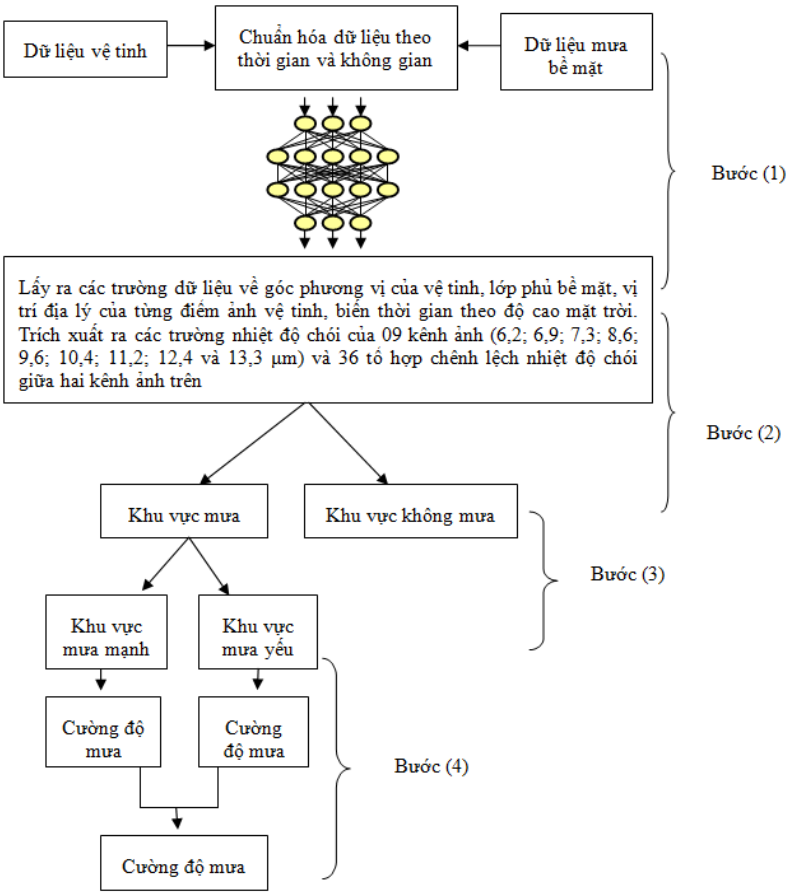
Nghiên cứu này sử dụng thuật toán RF để ước tính định lượng mưa từ các kênh ảnh của vệ tinh Himawari-8. Nguyên tắc cơ bản của phương pháp RF trước tiên là tạo ra nhiều cây phân loại và hồi quy dựa trên dữ liệu huấn luyệnđể thu được kết quả dự đoán cuối cùng theo số phân loại nhiều nhất (phân loại: mưa hay không mưa; mưa mạnh hay mưa yếu) hoặc lấy trung bình (hồi quy, cường độ mưa) của từng cây.

Đầu vào của mô hình RF bao gồm 09 kênh ảnh riêng lẻ (6,2; 6,9; 7,3; 8,6; 9,6; 10,4; 11,2; 12,4 và 13,3) và 36 kênh ảnh tổ hợp từ 09 kênh ảnh này được lấy từ các kênh ảnh của vệ tinh Himawari-8; ngoài ra các dữ liệu về góc phương vị của vệ tinh, lớp phủ bề mặt, vị trí địa lý của từng điểm ảnh vệ tinh, biến thời gian theo độ cao mặt trời cũng được thu thập để ước tính định lượng mưa. Hình 3 trình bày tổng quan các bước của thuật toán RF để ước tính lượng mưa từ vệ tinh Himawari-8.

Thuật toán ước tính lượng mưa từ vệ tinh Himawari-8 được thực hiện theo các bước sau:

Bước 1 (xử lý dữ liệu đầu vào): Dữ liệu đầu vào gồm 09 kênh ảnh riêng lẻ của vệ tinh Himawari trong giai đoạn 2019-2020 được đưa vào chuẩn hóa và đồng bộ về thời gian và không gian với bộ dữ liệu đo mưa bề mặt của 866 trạm cùng giai đoạn. Ở bước này từ số liệu gốc của các kênh ảnh sẽ được đồng bộ về thời gian là 10 phút và không gian lưới tính là từ -5^{o}N - 40^{o}N, 95^{o}E - 140^{o}E bao gồm kích thước 1125 × 1125 điểm lưới, độ phân giải 4km. Số liệu lượng mưa được căng lên lưới này và nội suy ra giá trị điểm mưa theo lưới vệ tinh để đồng nhất, được sử dụng làm cờ hiệu để xác nhận là có mưa hay không mưa, mưa lớn hay mưa nhỏ. Bộ dữ liệu ban đầu sẽ được tỉa thưa bớt các trường hợp mà giá trị lượng mưa bề mặt bằng 0 để bộ dữ liệu đưa vào training sẽ được giảm kích thước, quy mô bộ số liệu tính toán và tối ưu máy tính chạy (lọc khoảng 10%). Sau đó, kích thước của bộ dữ liệu training được tính bằng tỉ lệ giữa tổng số trường hợp có mưa nhân với 1.5 trên tổng số trường hợp không mưa (lượng mưa = 0), còn lại là bộ dữ liệu test. Các trọng số của mỗi thành phần trong mô hình training được xác định thông qua quá trình luyện mạng từ các bộ dữ liệu training được thành lập.

Bước 2 (phân loại khu vực có mưa và không có mưa): Từ dữ liệu gốc ở bước 1 thì sẽ trích xuất ra được trường dữ liệu nhiệt độ chói của 14 kênh ảnh này và độ chênh lệch nhiệt độ chói của 10 kênh ảnh tổ hợp. Sau đó được đưa vào mô hình training để phân loại ra vùng có mưa/ không mưa, vùng có mưa được gán bằng 1 và ngược lại không mưa bằng 0. Bộ dữ liệu sau khi đã xác định được vùng có mưa sẽ được chia theo tỉ lệ 80:20 cho bộ dữ liệu training và bộ dữ liệu test để xác định vùng mưa mạnh và mưa yếu.



Hình 3. Sơ đồ thuật toán ước tính lượng mưa từ vệ tinh Himawari-8.

Bước 3 (phân loại khu vực mưa mạnh và mưa yếu): sẽ tiến hành phân loại các ô pixel mây gây mưa thành ô pixel mây gây mưa mạnh và mưa yếu. Nghiên cứu sử dụng ngưỡng mưa 0-1,2 mm/ 10 phút để phân loại khu vực mưa yếu và trên 1,2 mm/ 10 phút là khu vực mưa mạnh. Sau đó xây dựng được 04 mô hình đó là mô hình mưa, không mưa, mưa mạnh, mưa yếu dựa trên tập số liệu vệ tinh Himawari-8 và đo mưa bề mặt giai đoạn 2019-2020.

Bước 4 (ước tính cường độ mưa): Cường độ mưa của các ô pixel điểm ảnh mưa mạnh và mưa yếu được ước tính từ 04 mô hình xây dựng được ở bước 3. Ở bước 4 sẽ xử lý làm trơn để trường lượng mưa được mịn hơn khi sử dụng các phép nội suy ở các bước trên.

2.3. Phương pháp đánh giá

Để đánh giá khả năng ước tính lượng mưa từ vệ tinh Himawari-8 của phương pháp RF, nghiên cứu đã sử dụng các chỉ số thống kê để thực hiện việc so sánh kết quả ước tính định lượng mưa của phương pháp RF với số liệu đo mưa bề mặt như: ME (cho biết xu hướng lệch trung bình của dự báo so với quan trắc, không thể hiện độ lớn của sai số, ME = 0 là hoàn hảo), MAE (thể hiện biên độ trung bình của sai số mô hình nhưng không nói lên xu hướng lệch của dự báo và quan trắc, MAE = 0 là hoàn hảo), RMSE (thể hiện độ lớn trung bình của sai số, khi giá trị RMSE càng gần MAE thì sai số dự báo càng ổn định), MAPE (thể hiện sai số phần trăm tuyệt đối trung bình giữa quan trắc và dự báo, nó dùng để đánh giá hiệu suất dự báo), R (dùng để đánh giá mối quan hệ tuyến tính giữa tập giá trị dự báo và tập giá trị quan trắc, R = 1 là hoàn hảo), R^2 (thể hiện tỉ lệ phương sai (mức độ biến thiên) của một biến phụ thuộc được giải thích bởi một biến độc lập hay xác định mức độ mà một hoặc nhiều biến độc lập ảnh hưởng đến một biến phụ thuộc, tuy nhiên nó không có biết mô hình dự báo là tốt hay xấu cũng như không biết được dự báo và quan trắc có sai lệch hay không, ví dụ R^2 = 0,5 thì

nghĩa là một nửa mức độ thay đổi giá trị dự báo được có thể được giải thích bằng các giá trị quan trắc).

Chỉ số ME: $$ME = \frac{1}{n}\sum_{i=1}^{n}(S_i - G_i) \quad (1)$$

Chỉ số MAE: $$MAE = \frac{1}{n}\sum_{i=1}^{n}|S_i - G_i| \quad (2)$$

Chỉ số RMSE: $$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(S_i - G_i)^2} \quad (3)$$

Chỉ số MAPE: $$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|S_i - G_i|}{G_i} \quad (4)$$

Hệ số tương quan R: $$R = \frac{\sum_{i=1}^{n}(S_i - \bar{S})(G_i - \bar{G})}{\sqrt{\sum_{i=1}^{n}(S_i - \bar{S})^2 \sum_{i=1}^{n}(G_i - \bar{G})^2}} \quad (5)$$

Hệ số xác định R^2: $$R^2 = 1 - \frac{\sum_{i=1}^{n}(S_i - G_i)^2}{[\sum_{i=1}^{n}(S_i - \bar{G}_i)^2 + \sum_{i=1}^{n}(S_i - G_i)^2]} \quad (6)$$

Trong các công thức trên, n là số điểm có dữ liệu, S_i là lượng mưa ước lượng từ vệ tinh theo phương pháp RF được nghiên cứu, G_i là mưa quan trắc bề mặt, $\bar{S}$ là lượng mưa ước lượng trung bình từ vệ tinh và $\bar{G}$ là lượng mưa trung bình quan trắc từ bề mặt.

3. Kết quả và thảo luận

Phương pháp RF được sử dụng để ước tính lượng mưa từ vệ tinh và đo mưa bề mặt để tạo ra trường lượng mưa có chất lượng cao với độ phân giải không gian cao bằng cách kết hợp ưu điểm của các quan trắc đo mưa bề mặt và ước tính định lượng mưa từ vệ tinh. Kết quả đánh giá của phương pháp RF với ước tính định lượng mưa từ vệ tinh và đo mưa bề mặt, thời gian ước tính 10 phút cho 04 trường hợp mưa: trường hợp thứ nhất từ ngày 01/08/2019 đến ngày 04/08/2019; trường hợp thứ hai từ ngày 02/09/2019 đến ngày 05/09/2019; trường hợp thứ ba từ ngày 02/03/2020 đến ngày 04/03/2020 và trường hợp cuối cùng xảy ra từ ngày 24/06/2023 đến ngày 27/06/2023.

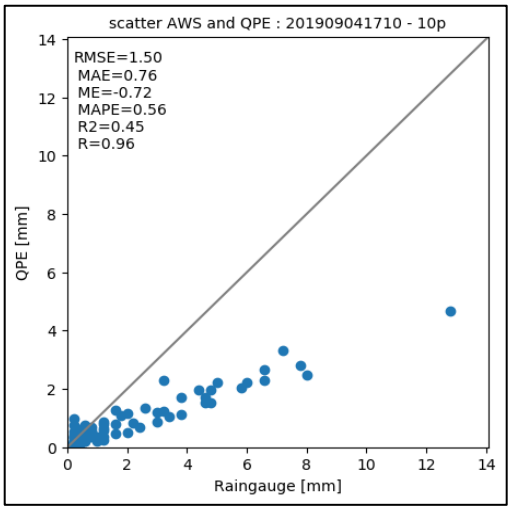
Áp dụng phương pháp RF, một vài kết quả ước tính lượng mưa từ vệ tinh Himawari-8 tại một vài ốp dự báo điển hình của 04 trường hợp mưa cụ thể được đánh giá thông qua các chỉ số thống kê (Bảng 2).

Bảng 2. Kết quả đánh giá ước tính lượng mưa 10 phút của 04 trường hợp mưa.

Thời gian	RMSE	MAE	ME	MAPE	R^2	R
Trường hợp 1: từ 10h00 đến 10h50 UTC ngày 03/08/2019						
10h00	1,05	0.62	-0,11	1,07	0,50	0,83
10h10	1,26	0,68	-0,14	1,03	0,48	0,80
10h20	1,13	0,68	-0,16	1,18	0,49	0,81
10h30	1,09	0,61	-0,13	1,11	0,49	0,84
10h40	1,33	0,69	-0,16	1,26	0,44	0,82
10h50	1,06	0,59	-0,14	1,01	0,48	0,84
Trường hợp 2: từ 17h00 đến 17h50 UTC ngày 04/09/2019						
17h00	1,94	0,95	-0,86	0,61	0,36	0,91
17h10	1,50	0,76	-0,72	0,56	0,45	0,96
17h20	1,35	0,77	-0,74	0,53	0,41	0,94
17h30	1,29	0,71	-0,66	0,60	0,37	0,91

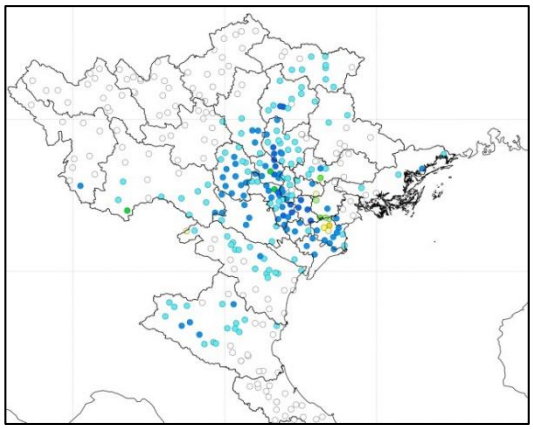
Thời gian	RMSE	MAE	ME	MAPE	R^2	R
17h40	1,48	0,78	-0,66	1,09	0,27	0,83
17h50	1,35	0,69	-0,62	0,57	0,30	0,89
Trường hợp 3: từ 04h00 đến 04h40 UTC ngày 04/03/2020						
04h00	1,55	0,87	-0,84	0,68	0,21	0,93
04h10	1,71	0,86	-0,85	0,62	0,25	0,91
04h20	2,14	1,05	-0,98	0,67	0,26	0,89
04h30	2,17	1,10	-1,02	0,62	0,29	0,91
04h40	1,79	0,89	-0,85	0,63	0,27	0,93
Trường hợp 4: từ 19h00 đến 19h50 UTC ngày 26/06/2023						
19h00	1,74	0,79	-0,69	0,64	0,43	0,85
19h10	1,74	0,76	-0,51	0,89	0,41	0,81
19h20	1,81	0,79	-0,63	0,80	0,43	0,91
19h30	1,42	0,61	-0,48	0,72	0,40	0,88
19h40	1,31	0,64	-0,53	0,73	0,32	0,87
19h50	1,08	0,57	-0,46	0,82	0,33	0,85

Kết quả đánh giá 04 trường hợp cụ thể của đợt mưa nghiên cứu cho thấy rằng, đối với tất cả các thời điểm ở các trường hợp khác nhau thì chỉ số ME đều nhỏ hơn 0, điều này cho thấy mưa ước lượng từ vệ tinh luôn được đánh giá thấp hơn so với mưa bề mặt. Các giá trị RMSE biểu thị độ lớn trung bình của sai số, trong các trường hợp nghiên cứu, độ lớn trung bình của sai số dao động từ 1,05 (10h00 ngày 03/08/2019) đến 2,17 (04h30 ngày 04/03/2020), MAE dao động từ 0,57 (19h50 ngày 26/6/2023) đến 1,1 (4h30 ngày 04/03/2020). Chỉ số MAPE dao động trong khoảng từ 0,5-1,3, MAPE có giá trị thấp nhất là vào 17h20 ngày 04/09/2019 với MAPE bằng 0,53. Giá trị hệ số xác định R^2 cho biết mức độ phụ thuộc của biến dự báo được giải thích bởi biến độc lập, trong các trường hợp trên R^2 dao động trong khoảng từ 0,2 đến 0,5, cho thấy có khoảng từ 20 đến 50% mức độ biến thiên dự báo được có thể được giải thích bằng các giá trị quan trắc. Mối quan hệ tương quan giữa lượng mưa ước tính được từ vệ tinh và lượng mưa quan trắc đạt giá trị rất tốt với R đều đạt trên 0,8. Tại 17h10 ngày 04/09/2019 giá trị hệ số tương quan đạt giá trị cao nhất với R lên tới 0,96. Hình 4 đưa ra biểu đồ scatter mưa vệ tinh và mưa bề mặt tại thời điểm 17h10 ngày 04/09/2019, ta thấy với lượng mưa lớn hơn 2mm/10 phút, phương pháp vẫn chưa thể cho kết quả ước lượng mưa tốt, với các điểm không nằm tập trung xung quanh đường chéo chính mà có sự phân tán lệch về một phía.

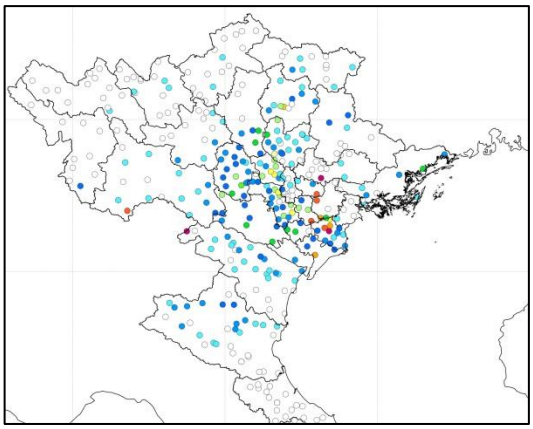


Hình 4. Biểu đồ scatter mưa vệ tinh và mưa bề mặt tại thời điểm 17h10 UTC ngày 04/09/2019.

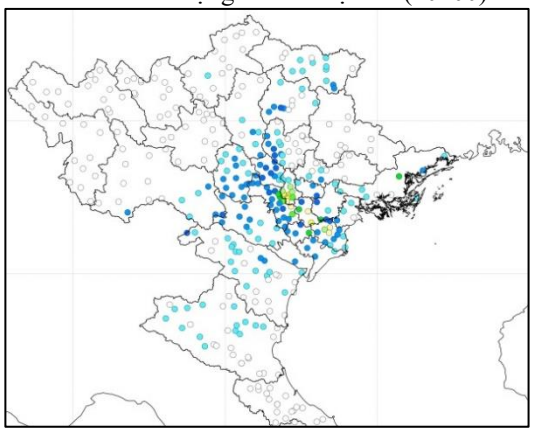
Mô phỏng trường lượng mưa cho 04 trường hợp mưa được ước tính. Kết quả chỉ ra rằng, phương pháp đã nắm bắt được trọng tâm của khu vực xảy ra mưa lớn. Ước tính lượng mưa từ vệ tinh cho dự báo khống tại một số điểm trạm, bên cạnh đó, tại một số trạm đo mưa bề mặt trong thực tế xảy ra mưa nhưng phương pháp vẫn chưa thể nắm bắt được. Giá trị ước tính lượng mưa từ vệ tinh tại một số điểm vẫn có sự sai lệch nhất định so với thực tế. Tại điểm mưa cực trị lượng mưa ước lượng đều cho thấy xu hướng thiên thấp hơn so với mưa tại trạm quan trắc. Dưới đây, trình bày một trường hợp cụ thể về trường lượng mưa được mô phỏng theo không gian giữa ước tính lượng mưa từ vệ tinh Himawari-8 sử dụng phương pháp RF so với trạm đo mưa bề mặt tại thời điểm từ 10h00 đến 10h50 UTC ngày 03/08/2019 với thời gian ước tính 10 phút (Hình 5). Kết quả ước tính lượng mưa thu được từ vệ tinh được quy về các điểm trạm để so sánh với lượng mưa thực tế tại điểm trạm.



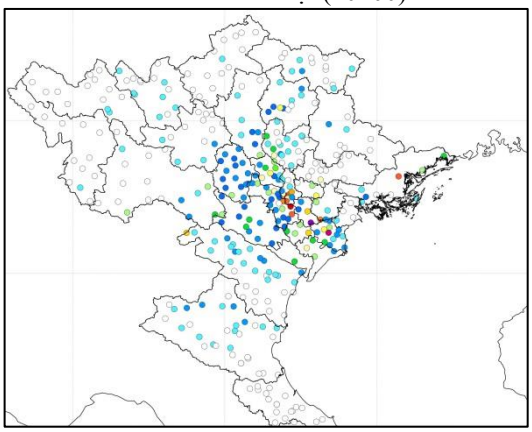
a. Ước tính lượng mưa từ vệ tinh (10h00)



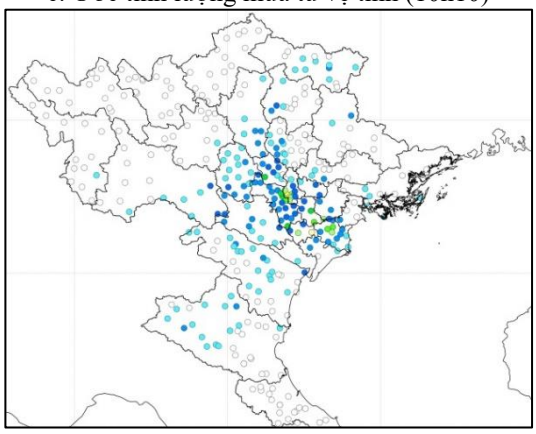
b. Đo mưa bề mặt (10h00)



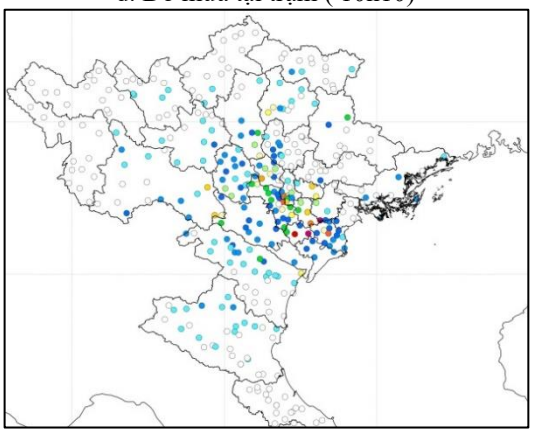
c. Ước tính lượng mưa từ vệ tinh (10h10)



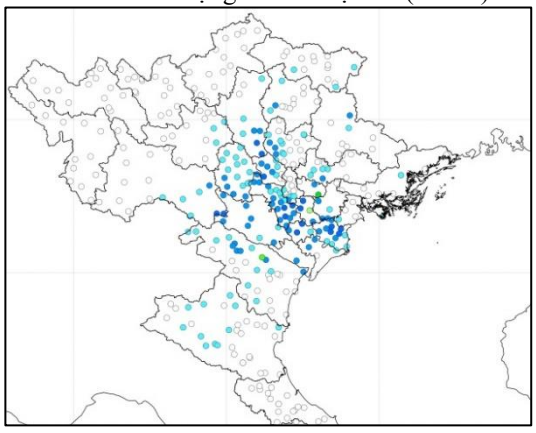
d. Đo mưa tại trạm (10h10)



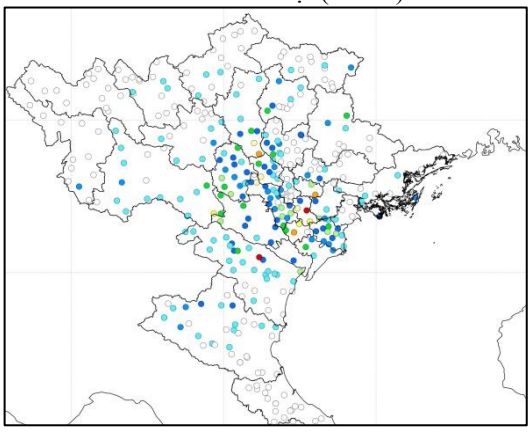
e. Ước tính lượng mưa từ vệ tinh (10h20)



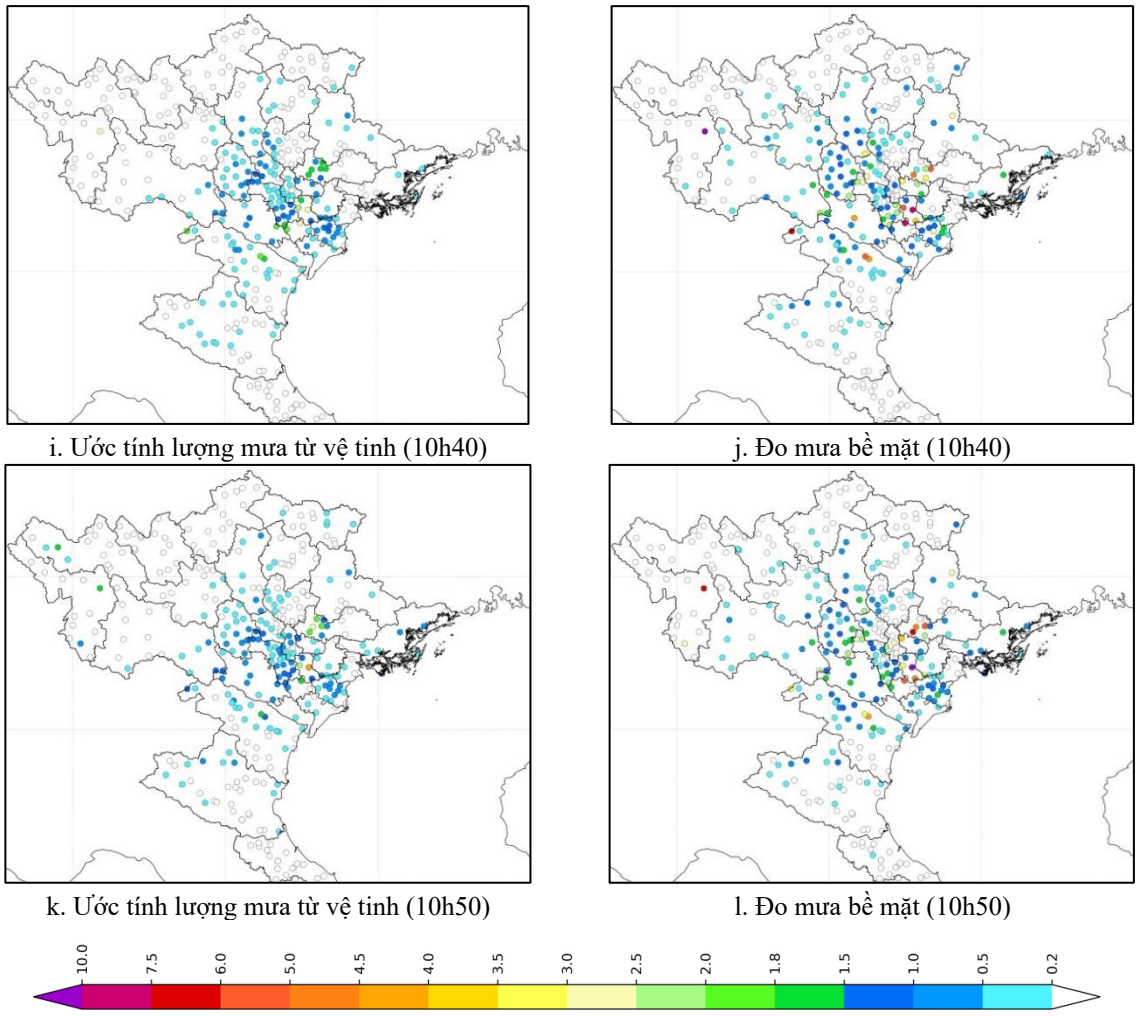
f. Đo mưa bề mặt (10h20)



g. Ước tính lượng mưa từ vệ tinh (10h30)



h. Đo mưa bề mặt (10h30)



i. Ước tính lượng mưa từ vệ tinh (10h40) j. Đo mưa bề mặt (10h40)

k. Ước tính lượng mưa từ vệ tinh (10h50) l. Đo mưa bề mặt (10h50)

Hình 5. Kết quả mô phỏng trường lượng mưa ước tính từ vệ tinh và trạm đo mưa bề mặt với thời gian ước tính 10 phút từ 10h00 đến 10h50 UTC ngày 03/08/2019.

Tại thời điểm ước tính định lượng mưa 10 phút từ 10h00 đến 10h50 ngày 03/08/2019, số liệu mưa thực tế cho thấy các điểm xảy ra mưa vừa đến mưa to, tập trung tại các tỉnh thuộc khu vực Bắc Bộ, tỉnh Thanh Hóa và tỉnh Nghệ An. Kết quả mô phỏng của trường lượng mưa từ vệ tinh cho thấy phương pháp RF đã nắm bắt được trọng tâm của đợt mưa lớn ở quanh khu vực đồng bằng Bắc Bộ. Tuy nhiên, so sánh từ hình 5a đến 5l thì có thể thấy rằng, giá trị của các điểm mưa cực trị được ước tính từ vệ tinh nhỏ hơn so với điểm mưa xảy ra ở thực tế. Ngoài ra, dự báo không xảy ra ở một số điểm trạm trên các khu vực các tỉnh Lạng Sơn, Cao Bằng. Tại một số điểm trạm thuộc khu vực các tỉnh Lào Cai, Lai Châu, Điện Biên, Sơn La phương pháp ước lượng mưa từ vệ tinh chưa nắm bắt được một số điểm trạm xảy ra mưa.

4. Kết luận

Trong nghiên cứu này, một phương pháp ước tính định lượng mưa dựa trên mô hình thuật toán RF đã được đề xuất thử nghiệm. Nghiên cứu đã trình bày về thuật toán RF sử dụng trong ước tính lượng mưa từ dữ liệu vệ tinh Himawari-8 với dữ liệu của 09 kênh ảnh (6,2; 6,9; 7,3; 8,6; 9,6; 10,4; 11,2; 12,4; và 13,3 μm), 36 tổ hợp từ các kênh ảnh đó và số liệu của 866 trạm đo mưa bề mặt. Phương pháp RF được đề xuất trong nghiên cứu trên đã đạt được những tiến bộ trong việc truy xuất ước tính lượng mưa từ vệ tinh. Với 04 trường hợp mưa được nghiên cứu, kết quả chỉ ra rằng: nghiên cứu thành công trong việc xác định trọng tâm của đợt mưa lớn. Tuy nhiên vẫn có những điểm trạm mưa ước lượng từ vệ tinh cho dự báo không và có những điểm mưa ước lượng vẫn chưa thể nắm bắt được chính xác điểm mưa.

Tại các điểm mưa cực trị, giá trị của lượng mưa ước tính từ vệ tinh cho kết quả thiên thấp hơn so với thực tế. Mối quan hệ tuyến tính giữa lượng mưa quan trắc và lượng mưa mô phỏng từ vệ tinh đạt ở mức đánh giá rất tốt với giá trị hệ số tương quan R đạt từ 0,8 trở lên; với giá trị chỉ số MAE dưới 1,1mm và RMSE dưới 2,17 mm.

Đóng góp của tác giả: Xây dựng ý tưởng nghiên cứu: N.V.T., B.T.K.H., N.M.C., N.T.H.A.; Lựa chọn phương pháp nghiên cứu: N.V.T., B.T.K.H., N.T.H.A.; Xử lý số liệu: N.M.C., H.T.T.T.; Phân tích mẫu: H.T.T.T., N.M.C., B.T.K.H.; Lấy mẫu: B.T.K.H., H.T.T.T., N.T.H.A.; Viết bản thảo bài báo: N.V.T., B.T.K.H., H.T.T.T., N.T.H.A. Chỉnh sửa bài báo: B.T.K.H., H.T.T.T., N.T.H.A.

Lời cảm ơn: Tập thể tác giả trân trọng cảm ơn sự hỗ trợ của đề tài nghiên cứu khoa học cấp quốc gia "Nghiên cứu ứng dụng dòng quang để ước lượng và dự báo định lượng mưa cho toàn lãnh thổ Việt Nam trên cơ sở số liệu vệ tinh Himawari và Radar thời tiết", mã số ĐTĐL.CN-58/21.

Lời cam đoan: Tập thể tác giả cam đoan bài báo này là công trình nghiên cứu của tập thể tác giả, chưa được công bố ở đâu, không được sao chép từ những nghiên cứu trước đây; không có sự tranh chấp lợi ích trong nhóm tác giả.

Tài liệu tham khảo

1. Hsu, K.; Gao, X.; Sorooshian, S.; Gupta, H.V. Precipitation Estimation from Remotely Sensed Information Using Artificial Neural Networks. *J. Appl. Meteor*. **1997**, *36(9)*, 1176–1190. Doi:10.1175/1520-0450(1997)036<1176:PEFRSI>2.0.CO;2.
2. Bessho, K.; Date, K.; Hayashi, M.; Ikeda, A.; Imai, T.; Inoue, H.; Kumagai, Y.; Miyakawa, T.; Murata, H.; Ohno, T.; Okuyama, A.; Oyama, R.; Sasaki, Y.; Shimazu, Y.; Shimoji, K.; Sumida, Y.; Suzuki, M.; Taniguchi, H.; Tsuchiyama, H.; Uesawa, D.; Yokota, H.; Yoshida, R. An Introduction to Himawari-8/9 – Japan's New-Generation Geostationary Meteorological Satellites. *J. Meteorolog. Soc. Jpn*. **2016**, *94(2)*, 151–183. Doi:10.2151/jmsj.2016-009.
3. Kurino, T. A satellite infrared technique for estimating "deep/shallow" precipitation. *Adv. Space. Res.* **1997**, *19(3)*, 511–514. doi:10.1016/s0273-1177(97)00063-x.
4. Wang, C.; Tang, G.; Xiong, W.; Ma, Z.; Zhu, S. Infrared Precipitation Estimation using Convolutional neural network for FengYun satellites. *J. Hydrol.* **2021**, *603(22)*, 127113. Doi: 10.1016/j.jhydrol.2021.127113.
5. Yuko, K.; Keiji, I.; Hidenori, S.; Kakuji, O. Study on Extracting Precipitation Information Using Infrared Bands of Himawari-8. The 40th Asian Conference on Remote Sensing (ACRS 2019), October 14-18, 2019, Daejeon Convention Center (DCC), Daejeon, Korea, 2019.
6. Bellerby, T.; Todd, M.; Kniveton, D.; Kidd, C. Rainfall Estimation from a Combination of TRMM Precipitation Radar and GOES Multispectral Satellite Imagery through the Use of an Artificial Neural Network. *J. Climate Appl. Meteor*. **2000**, *39(12)*, 2115–2128. Doi:10.1175/1520-0450(2001)040<2115:REFACO>2.0.CO;2.
7. Grimes, D.I.F.; Coppola, E.; Verdecchia, M.; Visconti, G. A Neural Network Approach to Real-Time Rainfall Estimation for Africa Using Satellite Data. *J. Hydrol.* **2003**, *4(6)*, 1119–1133. doi:10.1175/1525-7541(2003)004<1119:annatr>2.0.co;2.
8. Hong, Y.; Hsu, K.; Sorooshian, S.; Gao, X. Precipitation Estimation from Remotely Sensed Imagery Using an Artificial Neural Network Cloud Classification System. *J. Climate Appl. Meteor*. **2004**, *43(12)*, 1834–1853. doi:10.1175/jam2173.1.

9. Rivolta, G.; Marzano, F.S.; Coppola, E.; Verdecchia, M. Artificial neural-network technique for precipitation nowcasting from satellite imagery, Adv. *Geosci.* **2006**, *7*, 97–103. Doi: 10.5194/adgeo-7-97-2006.
10. Abhishek, K.; Kumar, A.; Ranjan, R.; Kumar, S. A rainfall prediction model using artificial neural network. 2012 IEEE Control and System Graduate Research Colloquium (ICSGRC), Shah Alam, Selangor, Malaysia (2012.07.16-2012.07.17), 82–87. Doi:10.1109/ICSGRC.2012.6287140.
11. Mountrakis, G.; Im, J.; Ogole, C. Support vector machines in remote sensing: A review. *ISPRS J. Photogramm. Remote Sens.* **2011**, *66(3)*, 247–259. doi:10.1016/j.isprsjprs.2010.11.001.
12. Hamidi, O.; Poorolajal, J.; Sadeghifar, M.; Abbasi, H.; Maryanaji, Z.; Faridi, H.R.; Tapak, L. A comparative study of support vector machines and artificial neural networks for predicting precipitation in Iran. *Theor. Appl. Climatol.* **2015**, *119*, 723–731.
13. Yin, G.; Yoshikane, T.; Yamamoto, K.; Kubota, T.; Yoshimura, K. A support vector machine-based method for improving real-time hourly precipitation forecast in Japan. *J. Hydrol.* **2022**, 612, 128125. Doi: 10.1016/j.jhydrol.2022.128125.
14. Chaudhary, S.; Dhanya, C.T. Decision tree-based reduction of bias in monthly IMERG satellite precipitation dataset over India. *H2Open J.* **2020**, *3(1)*, 236–255. Doi:10.2166/h2oj.2020.124.
15. Ma, L.; Zhang, G.; Lu, E. Using the gradient boosting decision tree to improve the delineation of hourly rain areas during the summer from advanced Himawari imager data. *J. Hydrometeorol.* **2018**, *19*, 761–776.
16. Tao, Y.; Gao, X.; Ihler, A.; Sorooshian, S.; Hsu, K. Precipitation Identification with Bispectral Satellite Information Using Deep Learning Approaches. *J. Hydrometeorol.* **2017**, *18(5)*, 1271–1283. Doi:10.1175/jhm-d-16-0176.1.
17. Moraux, A.; Dewitte, S.; Cornelis, B.; Munteanu, A. Deep learning for precipitation estimation from satellite and rain gauges measurements. *Remote Sens.* **2019**, *11(21)*, 2463. doi:10.3390/rs11212463.
18. Breiman, L. Random Forests. Mach. *Learn.* **2001**, *45*, 5–32. https://doi.org/10.1023/A:1010933404324.
19. Kühnlein, M.; Appelhans, T.; Thies, B.; Nauss, T. Improving the accuracy of rainfall rates from optical satellite sensors with machine learning — A random forests-based approach applied to MSG SEVIRI. *Remote Sens. Environ*. **2014**, *141*, 129–143. Doi:10.1016/j.rse.2013.10.026.
20. Hirose, H.; Shige, S.; Yamamoto, M.K.; Higuchi, A. High Temporal Rainfall Estimations from Himawari-8 Multiband Observations Using the Random-Forest Machine-Learning Method. *J. Meteorolog. Soc. Jpn*. **2019**, *97(3)*, 689–710. doi:10.2151/jmsj.2019-040.
21. Kühnlein, M.; Appelhans, T.; Thies, B.; Nauß, T. Precipitation Estimates from MSG SEVIRI Daytime, Nighttime, and Twilight Data with Random Forests. J. Appl. *Meteorol. Climatol.* **2014**, *53(11)*, 2457–2480. Doi:10.1175/jamc-d-14-0082.1.
22. Das, S. A random forest algorithm for nowcasting of intense precipitation events. *Adv. Space Res*. **2017**, *60(6)*, 1271–1282. Doi:10.1016/j.asr.2017.03.026.
23. Min, M.; Bai, C.; Guo, J.; Sun, F.; Liu, C.; Wang, F.; Xu, H.; Tang, S.; Li, B.; Di, D.; Dong, L.; Li, J. Estimating Summertime Precipitation from Himawari-8 and Global Forecast System Based on Machine Learning. IEEE Trans. *Geosci. Remote Sens*. **2019**, *57(5)*, 2557–2570. doi:10.1109/TGRS.2018.2874950.
24. Nazli, T.; Boris, T.; Natalia, H.; Jörg, B. Random forest-based rainfall retrieval for Ecuador using GOES-16 and IMERG-V06 data. *Eur. J. Remote Sens*. **2021**, *54(1)*, 117–139. Doi:10.1080/22797254.2021.1884002.

25. Công, T.T.; Quyền L.N.; Gián, N.M.; Quyết L.Đ. Ứng dụng số liệu ảnh vệ tinh Himawari trong dự báo và cảnh báo mưa dông cho khu vực Đồng bằng sông Cửu Long. *Tạp chí Khí tượng Thủy văn* **2020**, *719*, 1–13.
26. Thắng, N.V.; Cường, H.Đ.; Mậu, N.Đ.; Thư, N.V.; Quốc, P.K. Xác định lượng mưa kết hợp từ thông tin vệ tinh radar và đo mưa tại trạm. *Tạp chí Khí tượng Thủy văn* **2013**, *631*, 28–34.
27. Thư, N.V. Nghiên cứu phương pháp xác định lượng mưa trên cơ sở ảnh mây vệ tinh địa tĩnh MTSAT cho khu vực Việt Nam. Đề tài Nghiên cứu khoa học cấp Bộ, 2010.
28. Hải, B.T.; Tuấn, N.V. Nghiên cứu đánh giá và so sánh các dữ liệu mưa vệ tinh độ phân giải cao lưu vực sông Cả. *Tạp chí Khí tượng Thủy văn* **2018**, *695*, 17–28.

Quantitative Estimating Precipitation using Himawari-8 satellite data based on the Random Forest machine learning

Nguyen Vinh Thu[1]*, Bui Thi Khanh Hoa[1], Nguyen Minh Cuong[1], Hoang Thi Thanh Thuat[1], Nguyen Thi Hoang Anh[1]

[1] National Center for Meteorological and Hydrological Network, Vietnam Meteorological and Hydrological Administration; vinhthu73@gmail.com; nguyenminhcuong30596@gmail.com; hoangthithanhthuat99@gmail.com; khanhhoa303@gmail.com; hoanganhck@gmail.com

Abstract: The paper uses the Machine Learning Random Forest (RF) method for quantitative precipitation estimation based on Himawari -8 satellite data. The data used are 09 individual image channels and 36 combined image channels of Himawari-8 satellite data and rain gauge data to estimate rainfall for the Vietnam area. The RF method is experimentally applied to 04 heavy rainfall events occurring in 2019, 2020, and 2023. The results show that the RF method has simulated the center of heavy rain areas. The correlation coefficient between the estimated rainfall from the satellite and the observed rainfall (R) reaches 0.8 or more, with the Mean Absolute Error value (MAE) is less than 1.1 mm and the Root Mean Squared Error (RMSE) is less than 2.17mm. However, the RF method shows the estimated rainfall which tends to be lower than the observed rainfall.

Keywords: Rainfall estimation; Himawari-8 satellite; Random Forest.