

Bài báo khoa học

Nghiên cứu xây dựng lượng mưa trên lưới kết hợp giữa dữ liệu mưa vệ tinh GSMaP và mưa tự động tại Việt Nam

Thái Thị Bé Vân^{1,2*}, Lê Thị Thanh Nga³

¹ Trung tâm Dự báo Khí tượng Thủy văn Quốc gia; thaivan.meteo@gmail.com

² Học viên cao học Trường Đại học Khoa học Tự nhiên, Đại học Quốc gia Hà Nội; thaivan.meteo@gmail.com

³ Trung tâm Ứng dụng công nghệ khí tượng thủy văn; thanhnga1324@gmail.com

*Tác giả liên hệ: thaivan.meteo@gmail.com; Tel.: +84–941365898

Ban Biên tập nhận bài: 3/4/2023; Ngày phản biện xong: 22/5/2023; Ngày đăng bài: 25/5/2023

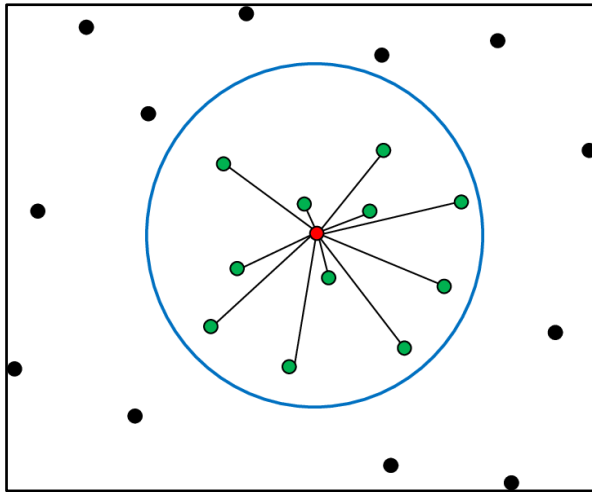
Tóm tắt: Bài báo trình bày kết quả tính toán lượng mưa trên lưới bằng việc kết hợp giữa số liệu mưa vệ tinh GSMaP và các trạm đo mưa tự động tại Việt Nam trong 04 tháng, tháng 8 và 9 của năm 2020 và 2021 thông qua phương pháp phân tích khách quan. 80 trạm quan trắc mưa tự động (tương đương 40% số liệu) đã được tách riêng để làm số liệu độc lập, 60% số trạm đưa vào chương trình phân tích khách quan để kết hợp 2 loại số liệu với nhau. Kết quả cho thấy, bộ số liệu kết hợp cho chất lượng ước lượng mưa tốt hơn khoảng 2–5% so với số liệu GSMaP gốc; số liệu kết hợp với bán kính ảnh hưởng 30km (L30) có sai số RMSE nhỏ nhất, cải thiện khoảng 5% so với số liệu GSMaP và có thể ứng dụng làm đầu vào cho các bài toán dự báo cực ngắn lượng mưa cũng như dự báo lũ, lũ quét và sạt lở đất.

Từ khóa: Mưa trên lưới; GSMaP; Phân tích khách quan; Mưa tự động; Số liệu mưa kết hợp.

1. Giới thiệu

Mục đích chính của bước phân tích khách quan của thuật toán đồng hóa số liệu là sự kết hợp thông tin từ các trường nền (trường phân tích, dự báo của mô hình, trường ước lượng mưa vệ tinh, ước lượng mưa từ radar, trường tái phân tích lượng mưa...), số liệu quan trắc địa phương (quan trắc trạm thủ công, tự động...) nhằm tạo ra một trường ban đầu tốt nhất có thể thông qua cực tiểu hàm giá (*cost function*), nhờ đó giúp nâng cao chất lượng trường phân tích, tạo ra ước lượng tốt nhất cho thời điểm ban đầu. Các nghiên cứu trước đây đã chỉ ra, trong các yếu tố tạo nên chất lượng của một hệ thống dự báo tốt thì dữ liệu đầu vào, trường ban đầu là một phần rất quan trọng, đặc biệt là khi các nguồn dữ liệu ngày càng nhiều, thông tin đo đạc đa dạng, bằng cả phương pháp gián tiếp và trực tiếp [1–3]). Các mô hình dự báo (số trị, thống kê tuyến tính, thống kê phi tuyến...) phụ thuộc rất lớn vào trường ban đầu: kết quả dự báo sẽ hoàn toàn sai lệch khi điều kiện ban đầu của các mô hình dự báo, đặc biệt là dự báo định lượng mưa không chính xác. Mục đích của phân tích khách quan là xác định chính xác nhất sử dụng tất cả thông tin có sẵn kết hợp với các số liệu bổ sung và trường nền để thu được trường điều kiện ban đầu tối ưu, làm đầu vào cho các mô hình dự báo khác nhau. Theo Daley, vấn đề xác định điều kiện ban đầu cho công cụ dự báo là rất cần thiết, tuy nhiên, đây là một quá trình vô cùng phức tạp do không chỉ cần phân tích khách quan mà còn phải kiểm tra tính nhất quán theo không gian và các mối quan hệ động lực [4]. [5] đã sử dụng phương pháp phân tích khách quan thực nghiệm và đưa ra một chương trình nội suy cho độ

cao địa thế vị bằng việc vẽ các vòng tròn quanh từng điểm lưới, chỉ những điểm quan trắc nằm trong vòng tròn (của từng điểm lưới) mới có ảnh hưởng đến giá trị của điểm lưới đó (Hình 1). Phương pháp này có nhược điểm là nếu dữ liệu quá thừa và phân bố dữ liệu theo thời gian không đồng đều sẽ không đủ để khởi tạo trường điều kiện ban đầu.



Hình 1. Điểm lưới (chấm đỏ), các điểm quan trắc (màu xanh lá cây) nằm trong vùng bán kính ảnh hưởng (màu xanh dương) mới có ảnh hưởng tới giá trị phân tích tại điểm lưới (màu đỏ).

[6–7] đã giới thiệu phương pháp hiệu chỉnh liên tiếp để khắc phục lượng dữ liệu quan trắc nghèo nàn bằng việc bổ sung thêm các thông tin từ trường nền, trường dự báo hạn ngắn hoặc từ các phân tích trước đó cho trường ban đầu. Nhìn chung có ba loại chính của kỹ thuật đồng hóa dữ liệu: phương pháp thực nghiệm của Gilchrist và Cressman, phương pháp biến phân dựa trên

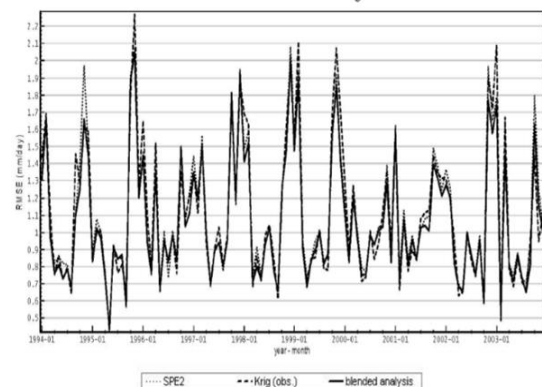
lý thuyết điều khiển tối ưu [8] và phương pháp thống kê dựa trên lý thuyết về ước lượng thống kê tối ưu [9] áp dụng trong lĩnh vực khí tượng.

Bước ngoặt của đồng hóa số liệu là các phương pháp biến phân (3D-VAR, 4D-VAR) lần đầu tiên được giới thiệu trong khí tượng bởi [10–11]. Bằng cách tối ưu hóa một hàm giá (*cost function*), bài toán biến phân tập trung vào việc tìm kiếm trạng thái khí quyển có khả năng xảy ra cao nhất ứng với một tập quan trắc và một trạng thái nền cho trước. Ưu điểm của phương pháp này là hàm giá được cực tiểu hoá trên toàn miền và các quan trắc được dàn đều, tránh được tình huống “mắt trâu” từ một điểm có giá trị quá lớn. Nhìn chung cùng với sự phát triển mạnh mẽ của máy tính, cho phép thực hiện được các thuật toán phân tích khách quan một cách nhanh chóng và có hiệu quả, nên đồng hóa số liệu trong khí tượng chỉ thực sự phát triển mạnh trong khoảng 20 năm gần đây mặc dù đã được quan tâm từ đầu những năm 50 của thế kỷ 20.

[12] nghiên cứu đề xuất thuật toán kết hợp giữa ước lượng mưa vệ tinh SPEs và số liệu đo mưa tại trạm cho Canada. Kết quả cho thấy, khi so sánh với ước lượng mưa vệ tinh và nội suy dữ liệu từ trạm đo bằng phương pháp kriging thì ước lượng mưa kết hợp (blended analysis) giữa hai loại dữ liệu có RMSE nhỏ nhất (Hình 2), ít sai lệch nhất và có thể được sử dụng cho tất cả các mùa, ngoài ra, khu vực có mật độ trạm đo càng thấp thì số liệu đồng hóa càng tốt, thể hiện ưu điểm khi so với phương pháp nội suy kriging.

[13] đã kết hợp giữa số liệu mưa vệ tinh GSMaP và số liệu mưa tại trạm trên khắp Australia. Kết quả đánh giá trong 20 năm (2001 đến 2020) đã cho thấy tính khả thi của phương pháp đề xuất: bộ dữ liệu kết hợp (GSMaP-bl, Hình 3) cho kết quả tốt hơn so với các bộ dữ liệu không dựa trên số liệu quan trắc khác cũng như có phân bố mưa mưa thực tế hơn so với bộ số liệu khí hậu hiện có của Australia (AGCD), đặc biệt ở những khu vực không có điểm đo mưa.

[14] đã kết hợp số liệu mưa ước lượng vệ tinh CMORPH và số liệu đo mưa tại trạm trong khoảng thời gian 5 tháng (từ tháng 4–9/2007). Bộ số liệu mưa vệ tinh sau khi được hiệu chỉnh sẽ kết hợp với bộ số liệu tại trạm thông qua phương pháp nội suy tối ưu OI (*OI - Optimal*



Hình 2. So sánh sai số RMSE cho 100 trạm giữa số liệu SPE, Kriging và ước lượng mưa kết hợp [12].

Interpolation). Bộ số liệu sau khi kết hợp sẽ được kiểm chứng chéo với bộ số liệu độc lập. Kết quả cho thấy bộ số liệu sau khi kết hợp đã tối ưu hơn, hạn chế được sai số hơn so với các bộ số liệu ban đầu.

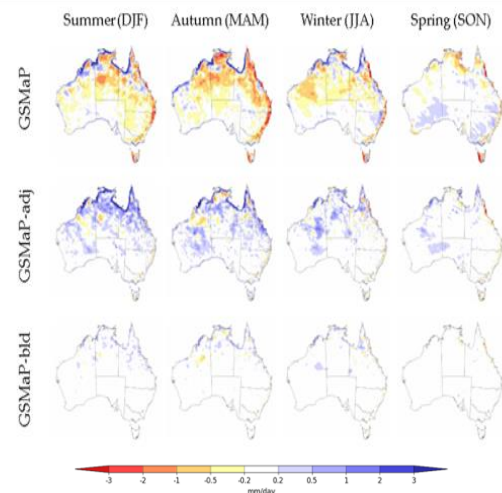
Ngoài ra, một số công trình nghiên cứu và đánh giá về số liệu ước lượng mưa vệ tinh GSMaP đã được một số tác giả thực hiện. Các kết quả cho thấy chất lượng của GSMaP trong mùa hè tốt hơn so với mùa đông [15], khu vực đồng bằng tốt hơn so với khu vực trung du và miền núi trong đó sai số RMSE dao động khoảng 4,8 mm/ngày ở khu vực Himalaya [16] và khoảng từ 4,93 đến 5,16 cho khu vực Việt Nam [17]. Các giá trị POD dao động khoảng 91% và FAR dao động trong khoảng 12–16%, điều đó có nghĩa là GSMaP khá tốt trong việc ước tính các sự kiện mưa và tỷ lệ các sự kiện cảnh báo sai chỉ khoảng 12–16% [18].

Tại Việt Nam, [19–20] đã đánh giá chất lượng số liệu mưa vệ tinh GSMaP trong các tình huống mưa lớn và ứng dụng vào tính toán lũ cho lưu vực sông Mã (Hình 4). Trong nghiên cứu này tác giả đã so sánh sản phẩm mưa GSMaP_MKV, GSMaP_NRT, GSMaP_NOW với dữ liệu thực đo cho 11 trận mưa vừa và lớn cho các trạm mưa trên và lân cận của lưu vực sông Mã thời đoạn 6 giờ trong 2 năm 2017 (6 trận mưa) và 2018 (5 trận mưa). Kết quả cho thấy, các sản phẩm mưa GSMaP có khả năng phát hiện được phân bố theo không gian, diễn biến mưa theo thời gian của các trận mưa. Ngoài ra, ba sản phẩm GSMaP đã được đánh giá và nhóm tác giả nhận thấy số liệu GSMaP_MKV cho kết quả tốt nhất, số liệu GSMaP_NOW cho kết quả kém nhất [19–20].

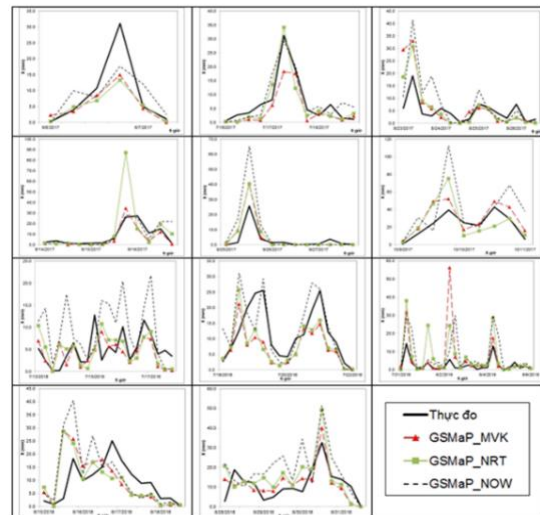
[21] đã nghiên cứu xây dựng bộ số liệu lượng mưa ngày VNGP_1deg trên lưới 1×1 độ kinh vĩ cho Việt Nam. [21] đã được kết hợp với nhau sử dụng phương pháp nội suy Cressman để đã tạo ra một bộ số liệu mưa ngày trên lưới 1×1 kinh vĩ (VnGP_1deg), cho giai đoạn từ 10/1996–12/2007. Kết quả đánh giá bộ dữ liệu VnGP_1deg với bộ số liệu mưa APHORODITE của Nhật Bản cho một số trường hợp mưa cụ thể cho thấy ưu điểm nổi trội của VnGP_1deg cả về phương diện phân bố thời gian và không gian.

Dữ liệu quan trắc hàng ngày từ 481 trạm đo mưa được sử dụng để xây dựng bộ dữ liệu lượng mưa mới cho Việt Nam dựa trên kỹ thuật nội suy Spheremap [22]. Tập dữ liệu mới được gọi là tập dữ liệu lượng mưa trên lưới cho Việt Nam (VnGP) giai đoạn 1980–2010. Kết quả đánh giá, so sánh cho thấy VnGP có chất lượng khá tốt so với các bộ dữ liệu khác vì đã sử dụng các kỹ thuật nội suy khác hoặc các phương pháp sử dụng số lượng đầu vào ít hơn.

[23] đã thử nghiệm phương pháp đồng hóa 3DVAR cho mưa vệ tinh vệ tinh trên lưu vực sông Hồng với việc sử dụng 2 nguồn số liệu mưa vệ tinh GSMaP và số liệu đo mưa toàn cầu GPM thông qua công cụ Delft-FEWS. Kết quả kiểm định cho thấy, mưa phân tích của



Hình 3. So sánh sai số trung bình theo 4 mùa DJF, MAM, JJA và SON giữa các bộ số liệu GSMaP, GSMaP-adj và GSMaP-bld [13].



Hình 4. Đường quá trình mưa thời đoạn 6 giờ trung bình lưu vực theo thực đo và các sản phẩm GSMaP_MKV, GSMaP_NRT và GSMaP_NOW cho 11 trận mưa trong hai năm 2017 và 2018 [19].

GSMaP cho hệ số tương quan $R = 0,53$, trong khi đó GPM chỉ là 0,21. Các nghiên cứu xây dựng các bộ số liệu trước đây chủ yếu tập trung vào lượng mưa ngày và sử dụng các phương pháp nội suy như Cressman, Krigging mà theo đánh giá, so sánh của các công trình gần đây thì có sai số lớn hơn so với phương pháp phân tích khách quan, kết hợp số liệu mưa trên lưới với số liệu quan trắc.

Nghiên cứu sẽ ứng dụng kỹ thuật phân tích khách quan để tạo ra trường ước lượng lượng mưa trên lưới thời gian thực kết hợp trường quan trắc lượng mưa tự động với trường nền là ước lượng mưa từ vệ tinh GSMaP. Kết quả thử nghiệm cho tháng 8–9 năm 2020–2021 để có thể ứng dụng trong các bài toán dự báo lũ cho lưu vực sông Mã. Bộ số liệu tạo ra sẽ được đánh giá và so sánh với các lượng mưa giờ tại các trạm độc lập để chỉ ra được hiệu quả của kỹ thuật phân tích khách quan và trường lượng mưa trên lưới tối ưu có thể ứng dụng trong các bài toán dự báo định lượng mưa hạn cực ngắn, dự báo lũ, cảnh báo lũ quét và sạt lở đất.

2. Số liệu và phương pháp nghiên cứu

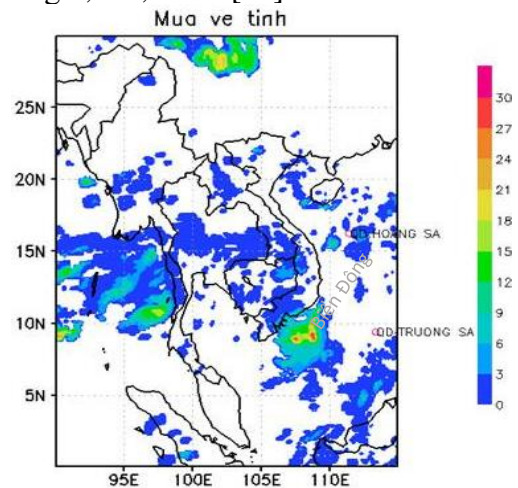
2.1. Số liệu

Số liệu mưa tự động: bộ số liệu quan trắc mưa tự động 1 giờ/1 lần bao gồm khoảng 200 trạm trong tháng 8 và tháng 9 năm 2020–2021 của các trạm trên toàn lãnh thổ Việt Nam được thu thập từ Trung tâm Dự báo khí tượng thủy văn quốc gia, Tổng cục Khí tượng Thủy văn. Sai số quan trắc lượng mưa của các thiết bị đo là khoảng 0,2–0,5 mm [22].

Số liệu Radar: Số liệu radar sử dụng chủ yếu là độ phản hồi vô tuyến của sản phẩm quét tròn (PPI) được thu thập cho những giờ nghỉ ngơi có quan trắc sai của bộ số liệu mưa giờ của các trạm đo mưa tự động trong 2 tháng 8 và 9 của 2 năm 2020–2021. Số liệu được cung cấp bởi Đài Khí tượng Cao không, Tổng cục Khí tượng Thủy văn.

Số liệu mưa ước lượng vệ tinh GSMaP: sử dụng sản phẩm NOW cho số liệu mưa giờ của bộ số liệu ước lượng lượng mưa vệ tinh toàn cầu (GSMaP) của JAXA (Nhật Bản) và là sản phẩm của Nhiệm vụ Đo mưa toàn cầu (PMM – *Precipitation Measuring Mission*) cung cấp trên website: <https://sharaku.eorc.jaxa.jp/GSMaP/index.htm>.

Miền tính được lựa chọn bao gồm khu vực Việt Nam và vùng lân cận, tọa độ 0–30°N; 90°E–115°E (Hình 5).



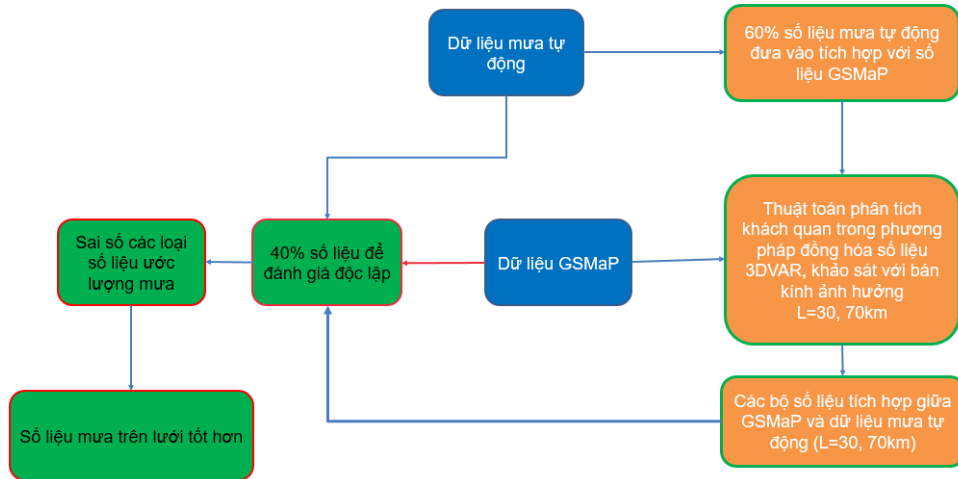
Hình 5. Khu vực thu thập số liệu ước lượng mưa từ vệ tinh GSMaP và số liệu minh họa cho lượng mưa giờ lúc 20 giờ ngày 3/9/2021.

2.2. Phương pháp nghiên cứu

Nghiên cứu này sẽ sử dụng kỹ thuật phân tích khách quan để kết hợp hai loại số liệu là số liệu ước lượng mưa giờ GSMaP của JAXA và số liệu đo mưa tự động từng giờ của Việt Nam sử dụng các giả thiết về sai số trường nền Σ_b của số liệu GSMaP, sai số quan trắc của số liệu mưa tự động Σ_o của các trạm tự động trong mạng lưới đo đạc của Tổng cục KTTV. Tiến hành thử nghiệm cho dữ liệu mưa các tháng 8 và 9 của năm 2020 và 2021, khảo sát với các bán kính ảnh hưởng khác nhau là 30 km và 70 km.

Trong số khoảng 200 trạm đo mưa tự động, sau khi rà soát, chuẩn hoá và loại bỏ sai số, sẽ sử dụng 80 trạm (tương đương khoảng 40% số liệu) để làm số liệu độc lập, đánh giá chất lượng số liệu ước lượng mưa GSMaP và số liệu mưa sau khi phân tích khách quan; còn lại khoảng 60% để đưa vào đồng hóa và khảo sát ảnh hưởng bán kính ảnh hưởng L đối với kết quả tích hợp.

Kết quả khảo sát sẽ được đánh giá dựa trên so sánh sai số của số liệu GSMaP ban đầu (trường nền) và số liệu trên lưới thời gian thực sau khi kết hợp hai loại số liệu kể trên với hai trường hợp bán kính ảnh hưởng $L = 30 \text{ km}$ và $L = 70 \text{ km}$, nhằm mục tiêu xác định được bộ tham số tối ưu cho bài toán đồng hóa số liệu mưa giờ của GSMaP và số liệu mưa tự động của Việt Nam.

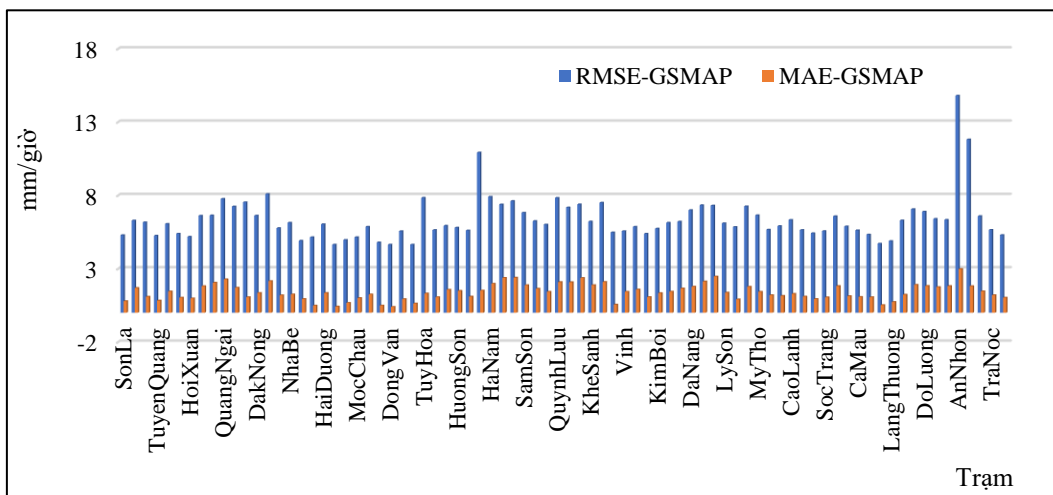


Hình 6. Thiết kế nghiên cứu với mục tiêu tạo ra bộ số liệu mưa trên lưới khi kết hợp số liệu GSMaP và đo mưa tự động sử dụng kỹ thuật phân tích khách quan.

3. Kết quả và thảo luận

3.1. Đánh giá chất lượng số liệu GSMaP

Số liệu GSMaP sẽ được đánh giá với khoảng 80 trạm mưa tự động (chọn ngẫu nhiên, tương đương 40% tổng số trạm) đã được rà soát và đánh dấu các số liệu sai (-99 – missing_value). Đây sẽ là 80 trạm độc lập, không đưa vào phân tích khách quan để đánh giá, so sánh chất lượng số liệu GSMaP gốc và số liệu phân tích khách quan kết hợp hai loại số liệu GSMaP và đo mưa tự động.



Hình 7. Sai số tuyệt đối trung bình MAE (màu cam đậm) và sai số bình phương trung bình RMSE (màu xanh) của số liệu ước lượng mưa GSMaP.

Chất lượng số liệu ước lượng mưa từ vệ tinh GSMaP được đánh giá qua các sai số MAE, RMSE và được thể hiện qua Hình 7, các đặc trưng sai số được cho trong Bảng 1.

Bảng 1. Thông kê sai số MAE và RMSE của bộ số liệu mưa vệ tinh GSMaP.

Đặc trưng	MAE	RMSE
Trung bình	1,41	6,39
Cực đại	2,98	14,81
Cực tiểu	0,40	4,64
Phân vị 25	1,06	5,56
Phân vị 75	1,81	6,92
Độ lệch chuẩn	0,54	1,54

Đối với sai số bình phương trung bình RMSE cho biết biên độ trung bình của sai số ước lượng mưa của từng trạm trong bộ số liệu độc lập trên toàn quốc, sai số RMSE dao động trong khoảng 4,64–14,81mm/giờ, RMSE trung bình khoảng 6,39 mm/giờ và độ lệch chuẩn của sai số là 1,54, tương đương 24% giá trị RMSE trung bình; khoảng sai số RMSE phổ biến (khoảng giá trị từ phân vị 25 đến phân vị 75) là 5,56 tới 6,92 mm/giờ. Như vậy có thể thấy là sai số của bộ số liệu GSMaP vẫn còn khá lớn, 6,40 mm/giờ cho trên khu vực Việt Nam.

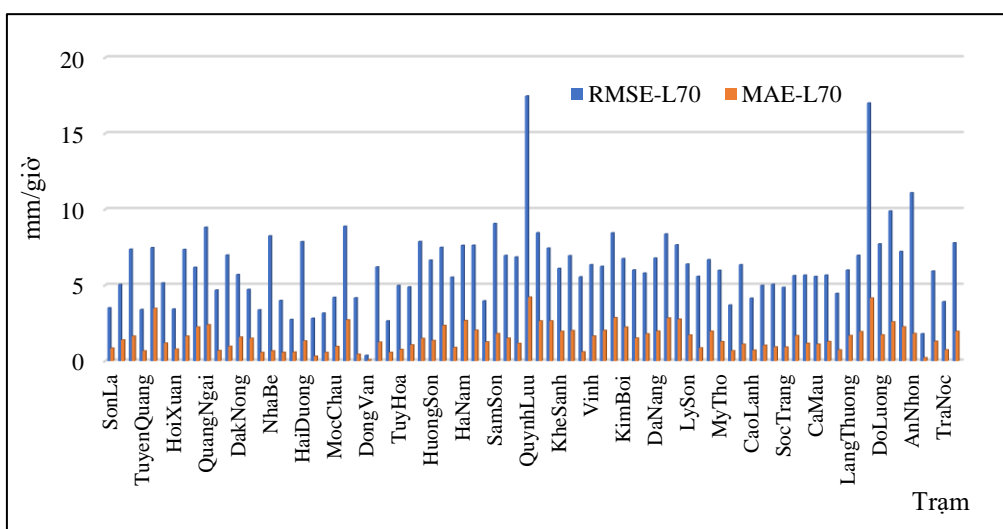
Sai số tuyệt đối trung bình MAE, dao động trong khoảng 0,40–2,98 mm/giờ, MAE trung bình khoảng 1,41 mm/giờ và độ lệch chuẩn của sai số là 0,24, tương đương 38,5% giá trị MAE trung bình; khoảng sai số tuyệt đối trung bình phổ biến là 1,06 tới 1,81 mm/giờ.

Khi so sánh hai loại sai số MAE và RMSE có thể thấy giá trị RMSE lớn hơn rất nhiều so với giá trị MAE, lên tới 4,5 lần (6,39 so với 1,41). Điều này cho thấy trong chuỗi số liệu so sánh giữa GSMaP và quan trắc tự động, có xuất một số giá trị sai số rất lớn, khiến cho giá trị sai số bình phương trung bình bị khuếch đại lên so với sai số tuyệt đối trung bình.

3.2. Đánh giá chất lượng số liệu mưa phân tích khách quan L70

Trong phần này, 60% chuỗi số liệu sau khi được lọc bỏ các quan trắc sai sẽ được sử dụng làm đầu vào cho chương trình phân tích khách quan để kết hợp với trường nền là số liệu GSMaP trên lưới của miền tính từ 90°E–115°E và 5°N–25°N, sau đó được đánh giá lại với bộ số liệu độc lập như đã đề cập ở trên.

Chương trình phân tích khách quan cho phép khảo sát sự ảnh hưởng của kích thước bán kính ảnh hưởng đối với chất lượng số liệu sau đồng hóa. Để thực hiện điều đó, nghiên cứu lấy xấp xỉ sai số trường nền của số liệu GSMaP Sigma_b = 6,5 mm/giờ dựa trên kết quả đánh giá sai số của số liệu GSMaP, sai số quan trắc của số liệu mưa tự động Sigma_o = 0,1 mm/giờ. Hai giá trị khác nhau của bán kính ảnh hưởng L sẽ được khảo sát để so sánh là L = 30 km và L = 70 km. Kết quả đánh giá cho trường hợp L = 70 km được cho trong Hình 8.



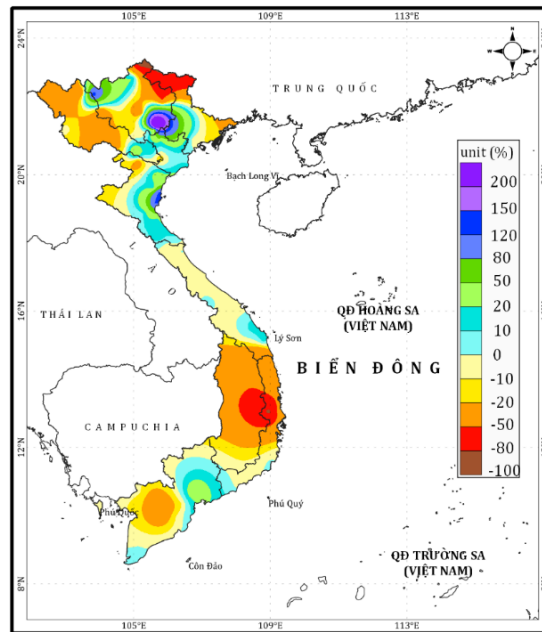
Hình 8. Sai số tuyệt đối trung bình MAE (màu cam đậm) và sai số bình phương trung bình RMSE (màu xanh) của số liệu phân tích khách quan với bán kính ảnh hưởng L = 70 km.

Bảng 2. Thống kê sai số MAE và RMSE của bộ số liệu phân tích khách quan với bán kính ảnh hưởng $L = 70$ km.

Đặc trưng	MAE	RMSE
Trung bình	1,51	6,23
Cực đại	4,21	17,47
Cực tiểu	0,07	0,37
Phân vị 25	0,84	4,28
Phân vị 75	1,96	7,45
Độ lệch chuẩn	0,84	2,61

Sai số RMSE dao động trong khoảng từ cực tiểu là 0,37 mm/giờ tại trạm Đồng Văn cho tới giá trị cực đại 17,47 mm/giờ tại trạm Tĩnh Gia, giá trị RMSE trung bình là 6,23 mm/giờ, độ lệch chuẩn 2,61 mm/giờ; so với giá trị RMSE trung bình của GSMaP là 6,39 thì RMSE của phiên bản L70 đã cải thiện chất lượng khoảng 2,5% sai số RMSE. Khi so sánh phân bố sai số RMSE và MAE của bộ số liệu L70 với số liệu GSMaP cho thấy: Sai số RMSE và MAE đều giảm.

Phân tích phân bố của tỷ lệ thay đổi sai số RMSE trên toàn quốc (Hình 9) khi so sánh số liệu L70 với GSMaP cho thấy, khu vực được cải thiện đáng kể nhất là Tây Nguyên, miền Tây Nam Bộ, khu vực biên giới tỉnh Cao Bằng; trong khi đó khu vực giữa Việt Bắc và đồng bằng Bắc Bộ, trong đó có thủ đô Hà Nội lại ít được cải thiện nhất. Hình 9 cũng cho thấy, số liệu mưa trên các khu vực trung du, vùng núi thường được cải thiện nhiều hơn so với vùng đồng bằng và ven biển. Sai số MAE của bộ số liệu L70 có tăng nhẹ, 1,46 mm/giờ so với sai số MAE của số liệu GSMaP là 1,41 mm/giờ.



RMSE70vsGSMAP

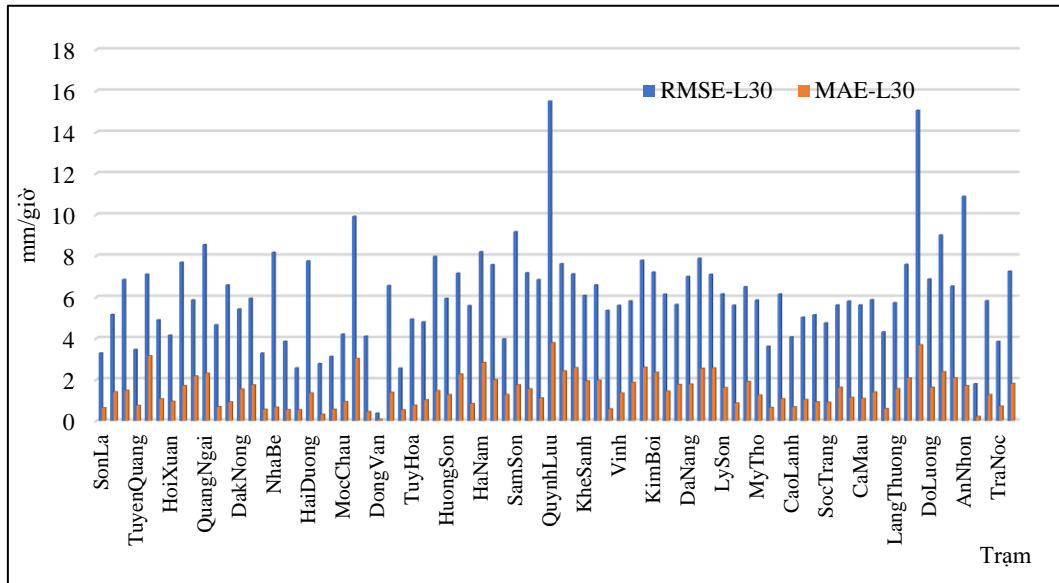
Hình 9. Tỷ lệ thay đổi sai số bình phương trung bình RMSE của số liệu L70 so với số liệu GSMaP (đơn vị: %). Tỷ lệ âm (màu nóng) tương ứng với sai số RMSE giảm, dương (màu lạnh) tương ứng với sai số RMSE tăng.

3.3 Đánh giá chất lượng số liệu mưa phân tích khách quan L30

Khảo sát thứ hai được thực hiện sau khi trong khảo sát với bán kính ảnh hưởng 70 km thấy trên số liệu xuất hiện những vùng “mắt trâu” với tâm mưa có lượng lớn và giảm đều về các phía (không chỉ ra ở đây). Để thử nghiệm, tác giả tiến hành giảm bán kính ảnh hưởng từ 70 km xuống 30 km và đánh giá lại với bộ số liệu độc lập để xem khi giảm bán kính ảnh hưởng sẽ thay đổi các sai số như thế nào. Kết quả đánh giá sai số MAE và RMSE của tất cả

các trạm độc lập được cho trong Hình 10 và thống kê đặc trưng cực đại, cực tiểu, trung bình... của các sai số RMSE, MAE cho các trạm độc lập được cho trong Bảng 3.

Sai số RMSE với trường hợp L30 của các trạm có sai số lớn như Tĩnh Gia đã giảm từ 17,47 xuống 15,50 mm/giờ (Hình 10); giá trị trung bình RMSE cho tất cả 80 trạm độc lập cũng giảm từ 6,23 trong trường hợp L70 xuống còn 6,09 mm/giờ trong trường hợp L30 (Bảng 3), cải thiện khoảng 4,63% so với chất lượng số liệu GSMaP và cải thiện khoảng 2,19% so với trường hợp L70.



Hình 10. Sai số tuyệt đối trung bình MAE (màu cam đậm) và sai số bình phương trung bình RMSE (màu xanh) của số liệu đồng hóa với bán kính ảnh hưởng L = 30 km.

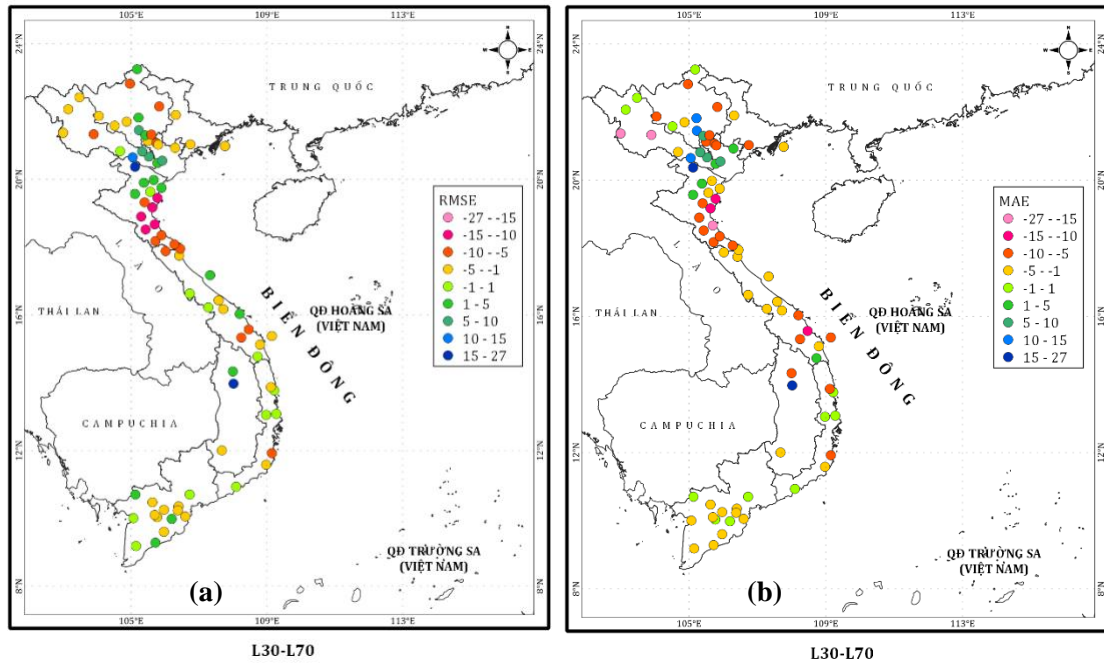
Sai số MAE trung bình của bộ số liệu L30 (1,46 mm/giờ) cũng giảm so với bộ số liệu L70 (1,51mm/giờ), khoảng giá trị phổ biến của MAE của bộ số liệu L30 trong khoảng 0,86 đến 1,91 mm/giờ, sai số MAE cực đại là 3,79 và cực tiểu là 0,37 mm. Đây đều là những trị số sai số khá tốt, đồng nghĩa với việc bộ số liệu có khả năng khai thác trong các bài toán dự báo tiếp theo.

Bảng 3. Thống kê sai số MAE và RMSE của bộ số liệu đồng hóa với bán kính ảnh hưởng L = 30km.

Đặc trưng	MAE	RMSE
Trung bình	1,46	6,09
Cực đại	3,79	15,50
Cực tiểu	0,37	0,37
Phân vị 25	0,86	4,78
Phân vị 75	1,91	7,17
Độ lệch chuẩn	0,79	2,37

Để thấy rõ hơn sự thay đổi và cải thiện của các sai số RMSE, MAE giữa bộ số liệu L30 so với bộ số liệu L70, nghiên cứu tính toán và biểu diễn bản đồ phân bố tỷ lệ giảm sai số RMSE và MAE trong Hình 11. Trong hình này, các trạm có trị số âm (màu nóng) có nghĩa là sai số của bộ số liệu L30 nhỏ hơn so với bộ số liệu L70, ngược lại các trạm có trị số dương (màu lạnh) là có sai số của bộ số liệu L30 lớn hơn bộ số liệu L70.

Như vậy, có thể thấy với cả hai loại sai số RMSE và MAE thì hầu hết các trạm đều có tỷ lệ cải thiện là khoảng từ 1–5% (màu cam), một số nơi cải thiện nhiều hơn trong phiên bản L30 như đối với hầu hết các trạm ở khu vực từ phía nam Thanh Hóa trở vào đến Hà Tĩnh, trong khi khu vực vùng núi Thanh Hóa, khu vực nam đồng bằng Bắc Bộ và phía bắc Thanh Hóa chất lượng lại giảm. Khu vực phía Tây Bắc Bộ cũng có các sai số được giảm đáng kể, đặc biệt là sai số MAE với các trạm ở Điện Biên, Lai Châu có sai số MAE giảm tới trên 15%.



Hình 11. Tỷ lệ giảm sai số RMSE (trái) và MAE (phải) của bộ số liệu L30 so với bộ số liệu L70. Tỷ lệ âm (màu nóng) tương ứng với sai số giảm, dương (màu lạnh) tương ứng với sai số tăng.

4. Kết luận

Tổng hợp các kết quả đánh giá độc lập đối với 03 bộ số liệu mưa: GSMaP, L30 và L70 cho thấy:

- Số liệu ước lượng mưa GSMaP vẫn có sai số tương đối lớn, khoảng 6,5 mm/giờ đối với sai số bình phương trung bình RMSE.
- Khi tích hợp với số liệu quan trắc mưa tự động tại Việt Nam sử dụng kỹ thuật phân tích khách quan đã giúp giảm sai số. Tuy nhiên các đợt mưa lớn vẫn có sai số lớn nên dẫn đến sai số RMSE vẫn còn cao hơn nhiều so với sai số MAE.
- Bán kính ảnh hưởng càng lớn thì sai số được cải thiện càng ít. Kết quả cho thấy bộ số liệu L70 có sai số lớn hơn bộ số liệu L30, nguyên nhân có thể là do mưa là yếu tố cục bộ, việc hạn chế bán kính ảnh hưởng của số liệu quan trắc mưa là cần thiết, tránh giả thiết các vùng lân cận cũng có cùng đặc điểm mưa.
- Bộ số liệu L30 đã giúp cải thiện đáng kể sai số ở khu vực vùng núi, đặc biệt là vùng núi Tây Bắc, một số nơi có sai số MAE được cải thiện tới trên 15%.

Tổng hợp lại, bộ số liệu L30 cho kết quả phân tích khách quan, kết hợp số liệu GSMaP và số liệu đo mưa tự động có sai số RMSE trung bình trên toàn quốc tốt nhất, khoảng 6mm/giờ, cải thiện khoảng 5% so với số liệu GSMaP, trong đó sai số cực tiểu 0,37 mm/giờ, sai số cực đại là 15,5 mm/giờ. Như vậy, số liệu tạo ra có thể sử dụng được làm đầu vào cho các bài toán dự báo cực ngắn lượng mưa, dự báo lũ, lũ quét và sạt lở đất. Tuy nhiên cần lưu ý lượng mưa tính toán được vẫn có sai số lớn trong các trường hợp mưa lớn.

Đóng góp của tác giả: Xây dựng ý tưởng nghiên cứu: L.T.T.N.; Xử lý số liệu: L.T.T.N., T.T.B.V.; Viết bản thảo bài báo: L.T.T.N., T.T.B.V.; Chỉnh sửa bài báo: L.T.T.N.

Lời cảm ơn: Bài báo là một phần kết quả của luận văn thạc sĩ của học viên Thái Thị Bé Vân dưới sự hướng dẫn của TS. Hoàng Phúc Lâm, sử dụng số liệu và phương pháp nghiên cứu và là kết quả của các nhiệm vụ TNMT.2022.06.06 và TNMT.2022.02.12.

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. Lorenz E.N. The predictability of hydrodynamic flow. *Trans. N. Y. Acad. Sci.* **1963**, 25(4), 409–432.
2. Krishnamurthy, V. Predictability of weather and climate. *Earth Space Sci.* **2019**, 6(7), 1043–1056.
3. Ferrari, F.; Cassola, F.; Tuju, P.E.; Stocchino, A.; Brotto, P.; Mazzino, A. Impact of Model Resolution and Initial/Boundary Conditions in Forecasting Flood-Causing Precipitations. *Atmosphere* **2020**, 11, 592.
4. Daley, R. Atmospheric Data Analysis. Cambridge Atmospheric and Space Science Series, 1991, pp. 457.
5. Gilchrist, B.; Cressman, G. An experiment in objective analysis. *Tellus* **1954**, 6, 309.
6. Bergthorsson, P.; Doos, B. Numerical weather map analysis. *Tellus* **1955**, 7(3), 329–340.
7. Cressman, G.P. An operational objective analysis system. *Mon. Weather Rev.* **1959**, 87, 367.
8. Lions, J.L. Contrôle optimal de syst`emes gouvern`es par des`equations aux d`eriv`ees partielles”. Dunod, 1968.
9. Bennett, A.F. Inverse Modeling of the Ocean and Atmosphere. Cambridge University Press, Cambridge. 2022.
10. Lewis, J.M.; Derber, J.C. The use of adjoint equations to solve a variational adjustment problem with convective constraints. *Tellus* **1985**, 37A, 309–322.
11. Talagrand, O.; Courtier, P. Variational assimilation of meteorological observations with the adjoint vorticity equation. Part I: Theory. *Q. J. R. Meteorolog. Soc.* **1987**, 113, 1311–1328.
12. Lin, A.; Wang, X.L. An algorithm for blending multiple satellite precipitation estimates with in situ precipitation measurements in Canada. *J. Geophys. Res. Atmos.* **2011**, 116, D21111. <https://doi.org/10.1029/2011JD016359>.
13. Chua, Z.W.; Kuleshov, Y.; Watkins, A.B.; Choy, S.; Sun, C.A. Two-Step Approach to Blending GSMaP Satellite Rainfall Estimates with Gauge Observations over Australia. *Remote Sens.* **2022**, 14, 1903. <https://doi.org/10.3390/rs14081903>.
14. Xie, P.; Xiong, A.Y. A conceptual model for constructing high-resolution gauge-satellite merged precipitation analyses. *J. Geophys. Res. Atmos.* **2011**, 116, D21106. <https://doi.org/10.1029/2011JD016118>.
15. Tian, Y.; Peters-Lidard, C.D.; Adler, R.F.; Kubota, T.; Ushio, T. Evaluation of GSMaP Precipitation Estimates over the Contiguous United States. *J. Hydrometeorol.* **2010**, 566–574. <https://doi.org/10.1175/2009JHM1190.1>
16. Mandira Shrestha và cs. Verification of GSMaP Rainfall Estimates over the Central Himalayas. *J. Jpn. Soc. Civil Eng. Ser B1* **2012**, 67(4), I_37–I_42. Doi:10.2208/jscejhe.67.I_37.
17. Hung, M.K.; et al. Application of GSMaP Satellite data in precipitation estimation and nowcasting: evaluations for October 2019 to January 2020 period for Vietnam, *VNJ. Hydrometeorol.* **2020**, 8, 79–93. doi: 10.36335/VNJHM.2020(5).80-94.
18. Nurlatifah, A.; Susanti, I.; Sipayyung, S.B.; Latifah, H. Application of GSMaP on estimating rainfall condition in Jakarta during 16 December. AIP Conference Proceedings Published by American Institute of Physics. **2021**, 2366, 060027. Doi:10.1063/5.0060373,
19. Kiên, N.T.; An, N.L.; Thành, L.Đ. Đánh giá chất lượng mưa vệ tinh GSMaP mô phỏng mưa lớn - Ứng dụng cho lưu vực sông Mã. *Tạp chí khoa học kỹ thuật Thủy lợi và Môi trường* **2019**, 64, 76–83.

20. Saito, K.; Hung, M.K.; Hung, N.V.; Vinh, N.Q.; Tien, D.D. Heavy rainfall in central Viet Nam in December 2018 and modification of precipitation analysis at VNMHA. *VNJ. Hydrometeorol.* **2020**, 5, 65–78. Doi: 10.36335/VNJHM.2020(5).65-79.
21. Đức, T.A.; Tân, P.V.; Thành, N.Đ. Xây dựng bộ số liệu lượng mưa ngày VNGP_1deg trên lưới 1'×1' kinh vĩ cho Việt Nam. *Tạp chí Khí tượng Thủy văn* **2010**, 590, 42–48.
22. Thanh, N.X.; Thanh, N.D.; Kamomera, H.; Long, T.T.; Matsumoto, J.; Inoue, T.; Tan, P.V. The Vietnam Gridded Precipitation (VnGP) Dataset: Construction and Validation. *Sci. Online Lett. Atmos.* **2016**, 12, 291–296.
23. Huyền, T.T.; Chanh, K.Q.; DĐuwcs. T.Q. Đồng hóa mưa vệ tinh bằng phương pháp 3Dvar, áp dụng thử nghiệm cho lưu vực sông Hồng. *Tạp chí khoa học biến đổi khí hậu* **2019**, 10, 15–20.
24. Thông tư số 30/2018/TT-BTNMT: Quy định kỹ thuật về quan trắc và cung cấp thông tin, dữ liệu khí tượng thủy văn đối với trạm khí tượng thủy văn chuyên dùng.

Constructing gridded rainfall by combining GSMaP satellite rainfall estimation and automatic rain gauge observation in Vietnam

Thai Thi Be Van^{1,2*}, Le Thi Thanh Nga³

¹ National Center for Hydro-Meteorological Forecasting; thaivan.meteo@gmail.com

² Master student, Hanoi University of Sciences, Hanoi National University; thaivan.meteo@gmail.com

³ Center for Hydro-Meteorological Technology Application; thanhnga1324@gmail.com

Abstract: This article presents the results of the construction of gridded rainfall by combining GSMaP satellite rainfall estimation data and automatic rain gauge observation in Vietnam in four months of August and September 2020 and 2021 using objective analysis method. 80 automatic rain gauge stations (equivalent to 40% of the data) have been separated to be used as independent data, 60% of the stations are input for the objective analysis program to combine the two data types. The results show that combined data improve the quality of rainfall estimation by 2-5% compared to original GSMaP data; the combined data with influence radius of 30km (L30) have the smallest RMSE which improves about 5% compared to GSMaP data and can be used as input for rainfall nowcasting, flood forecast and flash flood and landslide warning programs.

Keywords: Gridded rainfall; GSMaP; Objective analysis; Automatic rain gauge; Combined rainfall data.