TẠP CHÍ
KHÍ TƯỢNG THỦY VĂN

Bài báo khoa học

Nghiên cứu hiệu chỉnh sản phẩm dự báo mưa hạn nội mùa cho khu vực Bắc Trung Bộ bằng phương pháp phân vị với xấp xỉ Gamma (QM–G)

Hoàng Thị Thu Hương[1]*, Nguyễn Văn Lượng[1], Phan Văn Vinh[1], Phạm Thanh Hà[2]

[1] Đài Khí tượng Thủy văn khu vực Bắc Trung Bộ; hoanghuong.btb@gmail.com; Luongnvkttv@gmail.com; Vinhpv@gmail.com@gmail.com
[2] Trường Đại học Khoa học Tự nhiên – Đại học Quốc gia Hà Nội; phamthanhha5693@gmail.com

*Tác giả liên hệ: hoanghuong.btb@gmail.com; Tel: +84–945698793

Ban Biên tập nhận bài: 4/3/2023; Ngày phản biện xong: 15/4/2023; Ngày đăng bài: 25/5/2023

Tóm tắt: Khoảng thời gian từ 10 ngày đến 2 tháng (hạn nội mùa) có ý nghĩa quan trọng trong quản lý nguồn nước, điều tiết hồ chứa, cả thuỷ lợi và thuỷ điện, cũng như lập nông lịch ở khu vực Bắc Trung Bộ. Tuy nhiên, cũng không thể phủ nhận được thực tế rằng, ở Việt Nam nói chung và Bắc Trung Bộ nói riêng bài toán dự báo mưa hạn nội mùa chỉ mới được đặt những viên gạch đầu tiên, kĩ năng dự báo vẫn còn thấp. Để có thể ứng dụng vào thực tiễn đòi hỏi phải có nhiều hơn các thử nghiệm, các công trình nghiên cứu hoặc tìm ra các công cụ dự báo mạnh mẽ hơn. Nhóm tác giả đã thực hiện nghiên cứu hiệu chỉnh dự báo mưa hạn nội mùa của mô ECMWF dựa trên phương pháp phân vị với xấp xỉ gamma (QM–G) cho độ phân giải thời gian 5 ngày với từng hạn dự báo (3–6 tuần). Nghiên cứu sử dụng hai bộ số liệu chính: bộ số liệu dự báo mưa hạn nội mùa của mô hình ECMWF cho 20 năm trong giai đoạn 2000–2019 và bộ số liệu mưa vệ tinh TRMM. Kết quả cho thấy: Sau khi hiệu chỉnh, các giá trị sai số như ME, MAE, RMSE giảm đi rõ rệt, đặc biệt là từ tháng 12 đến tháng 3 năm sau. Riêng tháng 7–9, tuy sai số đã giảm so với trước hiệu chỉnh nhưng vẫn còn nhiều vùng (chủ yếu là phía Nam khu vực) vẫn còn tồn tại sai số lớn. Tuy nhiên, giá trị tương quan CORR sau hiệu chỉnh không cải thiện, thậm chí có tháng còn giảm so với trước khi hiệu chỉnh.

Từ khóa: Hạn nội mùa; Hiệu chỉnh sai số; ECMWF; QM–G; Bắc Trung Bộ.

1. Mở đầu

Hiện nay, Đài KTTV khu vực Bắc Trung Bộ đã thực hiện cung cấp các bản tin dự báo mưa cho các đơn vị liên quan chi tiết đến hàng ngày, 3–5–10 ngày. Ngoài ra, Đài còn cung cấp bản tin dự báo tháng, dự báo hạn mưa hạn mùa từ 3 đến 6 tháng. Tuy nhiên, công tác dự báo mưa hạn dự báo từ 10 ngày cho đến 2 tháng của Đài còn rất nhiều hạn chế. Bên cạnh đó, khoảng thời gian từ 10 ngày đến 2 tháng có ý nghĩa quan trọng quản lý nguồn nước, điều tiết hồ chứa, cả thuỷ lợi và thuỷ điện, cũng như lập nông lịch nhưng Đài chưa tập trung khai thác cũng như cung cấp được các thông tin dự báo được chi tiết đến từng độ phân giải thời gian và các hạn dự báo khác nhau. Hay nói cách khác, vấn đề dự báo mưa hạn nội mùa ở Đài khu vực Bắc Trung Bộ là mới và chưa được thực hiện trên phạm vi của Đài.

Trong những năm gần đây, những thông tin từ dự báo nội mùa đã được cung cấp từ nhiều trung tâm trên thế giới, như như hệ thống dự báo mùa CFS của NCEP (Hoa Kỳ), dự báo mùa và nội mùa của Trung tâm dự báo hạn vừa châu Âu (ECMWF)… [1] đã nghiên cứu đánh giá kết quả dự báo mưa hạn nội mùa ở châu Phi bằng cách sử dụng các dự báo từ ba mô hình (ECMWF, UKMO và NCEP); [2] đã nghiên cứu dự báo mưa hạn nội mùa trên một số đại dương và đất liền ở vùng Nhiệt đới Châu Á từ mô hình CFSv2; [3] đã đánh giá kỹ năng dự báo mưa trong tháng từ hệ thống dự báo toàn cầu (EPS)… Tuy nhiên, sản phẩm của các mô hình này luôn tiềm ẩn sai số. Các sai số này ngoài sự không hoàn hảo của chính các mô hình còn có sự đóng góp của sai số từ điều kiện ban đầu và điều kiện biên. Vì tồn tại các sai số trong cả quá trình nên sản phẩm của các mô hình đều cần được hiệu chỉnh để nhận sản phẩm dự báo phù hợp với số liệu quan trắc. Nếu phương pháp hiệu chỉnh có hiệu quả có thể sử dụng sản phẩm dự báo cho các ứng dụng thực tiễn.

Về hiệu chỉnh mưa, phương pháp hiệu chỉnh phân vị (*QM – Quantile Mapping*) đã được áp dụng nhiều trong việc hiệu chỉnh sản phẩm mưa. Có thể kể đến một số nghiên cứu, [4] đã nghiên cứu khả năng hiệu chỉnh sai lệch của các phương pháp phân vị đối với các biến nhiệt độ và lượng mưa; [5–6] đã trình bày chi tiết về phương pháp QM–G; [7] đã sử dụng hai phương pháp là phân vị với xấp xỉ gamma (QM–G) và Bayesian xác suất kết hợp (BJP) để hiệu chỉnh sản phẩm mưa dự báo hạn mùa từ mô hình phổ khu vực (RSM); [8] đã nghiên cứu dự tính số ngày nắng nóng cho Việt Nam dựa trên số liệu nhiệt độ đã được hiệu chỉnh theo phương pháp phân vị (*Quantile mapping*) tổ hợp đa mô hình; [9] đã nghiên cứu hiệu chỉnh cường độ và tần suất mưa ngày từ các mô hình khí tượng toàn cầu cho trạm Láng bằng phương pháp Quantile mapping….

Xuất phát từ tầm quan trọng về vấn đề hiệu chỉnh mưa hạn nội mùa, nghiên cứu này đã thực hiện hiệu chỉnh sản phẩm mưa hạn nội mùa của mô hình ECMWF bằng phương pháp phân vị với xấp xỉ gamma **(**QM–G) cho độ phân giải thời gian 5 ngày (tổng lượng mưa 5 ngày) với từng hạn dự báo (3–6 tuần). Phần tiếp theo của bài báo sẽ trình bày về số liệu và phương pháp nghiên cứu (mục 2); kết quả và phân tích (mục 3); và cuối bài báo là kết luận (mục 4).

2. Số liệu và phương pháp nghiên cứu

2.1. Số liệu

Nghiên cứu sử dụng hai bộ số liệu: bộ số liệu dự báo mưa hạn nội mùa của mô hình ECMWF cho 20 năm trong giai đoạn 2000–2019 và bộ số liệu mưa vệ tinh TRMM làm số liệu quan trắc.

a) Bộ số liệu dự báo mưa hạn nội mùa từ mô hình toàn cầu ECMWF

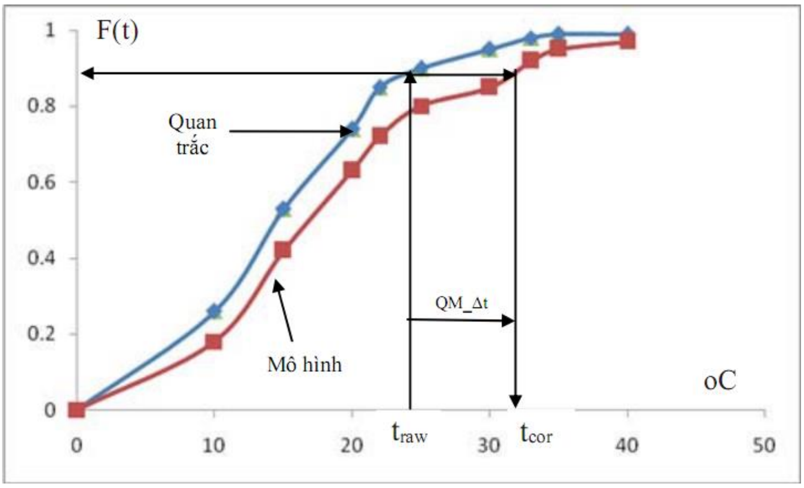
Số liệu dự báo về lượng mưa ngày từ cơ sở dữ liệu dự báo lại hạn nội mùa (*Reforecast*) từ mô hình ECMWF trong 20 năm 2000–2019. Các dự báo lại của ECMWF cung cấp bao gồm 11 thành phần (member dự báo) và mỗi sản phẩm dự báo có hạn dự báo 46 ngày với độ phân giải không gian là 1.5°×1.5°. Để chọn tốt hơn các điểm lưới nằm trong các tiểu vùng khác nhau, đề tài sử dụng bộ dữ liệu dự báo đã nội suy về độ phân giải 0.25°×0.25°. Các bản tin dự báo của ECMWF đưa ra liên tục với tần suất 2 lần/tuần vào thứ hai và thứ năm, trong đó bao gồm thông tin về dự báo lại trong 20 năm tính từ thời điểm có sản phẩm dự báo nghiệp vụ mới nhất. Độ phân giải thời gian của các dự báo là hàng ngày và dữ liệu được cung cấp miễn phí trực tuyến từ website: https://apps.ecmwf.int/datasets/data/s2s/.

b) Bộ số liệu mưa quan trắc từ vệ tinh TRMM

Số liệu mưa vệ tinh (Dữ liệu mưa TMPA 3B42V7): Là sản phẩm số liệu mưa 03 giờ, độ phân giải 0.25°×0.25° của TMPA 3B42V7 được tải về từ NASA (https://pmm.nasa.gov/data-access/downloads/trmm) trong 20 năm 2000–2019.

2.2. Phương pháp hiệu chỉnh phân vị với xấp xỉ gamma (QM–G)

Phương pháp hiệu chỉnh phân vị với xấp xỉ gamma trong nghiên cứu này dựa trên giả định ban đầu rằng cả phân bố cường độ được quan trắc và dự báo đều gần đúng với phân bố gamma (Hình 1). Việc hiệu chỉnh được thực hiện dựa vào phân bố của quan trắc và được thực hiện cho từng phân vị. [10–11] đã chỉ ra rằng sử dụng hàm gamma cho biến mưa hiệu quả hơn so với các biến khác.



Hình 1. Sơ đồ phương pháp hiệu chỉnh phân vị (*Quantile mapping*) [8].

2.3. Các chỉ số đánh giá định lượng

Để ứng dụng được một phương pháp hiệu chỉnh vào trong nghiệp vụ dự báo đòi hỏi trước hết là đánh giá được kết quả sau hiệu chỉnh. Có rất nhiều phương pháp đánh giá, trong khóa luận này sẽ sử dụng phương pháp đánh giá bằng các chỉ số thống kê.

Với F là giá trị dự báo từ mô hình, O là giá trị quan trắc. Các chỉ số thống kê được tính như sau:

a) Sai số trung bình hay sai số hệ thống ME (*Mean Error*)

$$ME=\frac{1}{N}\sum_{i=1}^{n}(F_i - O_i) \quad (1)$$

Chỉ số ME biểu thị sai số trung bình của mô hình so với quan trắc, cho biết xu hướng lệch trung bình của giá trị dự báo so với giá trị quan trắc nhưng không phản ánh độ lớn của sai lệch, nó. Giá trị ME nằm trong khoảng $-\infty$ đến $+\infty$. Với giá trị ME = 0 mô hình được coi là "hoàn hảo". Nếu ME dương đồng nghĩa với giá trị mô hình cao hơn giá trị quan trắc. ME âm thì giá trị mô hình thấp hơn giá trị quan trắc.

b) Sai số tuyệt đối trung bình (*MAE– Mean Absolute Error*)

$$MAE= \frac{1}{N}\sum_{i=1}^{n}|F_i - O_i| \quad (2)$$

Chỉ số MAE biểu thị biên độ trung bình của sai số mô hình nhưng không nói lên xu hướng lệch của giá trị dự báo và quan trắc. Nó sử dụng để đo độ chính xác. Giá trị MAE nằm trong khoảng (0, +∞), giá trị của nó càng nhỏ thì độ chính xác càng lớn. Khi MAE = 0, giá trị của mô hình hoàn toàn trùng khớp với giá trị quan trắc, mô hình được xem là "lý tưởng".

c) Sai số trung bình quân phương (*RMSE– Root Mean Square Error*)

$$RMSE=\sqrt{\frac{1}{N}\sum_{i=1}^{n}(F_i - O_i)^2} \quad (3)$$

Chỉ số RMSE biểu thị độ lớn trung bình của sai số. Đặc biệt RMSE rất nhạy với những giá trị sai số lớn. Phép toán bình phương trong căn bậc hai có tác dụng khuếch đại sai số. Do đó nếu các giá trị F và O có độ lệch không lớn thì RMSE càng gần MAE, sai số mô hình càng ổn định. Giá trị RMSE nằm trong khoảng (0, +∞).

d) Hệ số tương quan (*Correlation Coefficient*)

$$CORR = \frac{\frac{1}{n}\sum_{i=1}^{n}(F_i-\bar{F})(O_i-\bar{O})}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(F_i-\bar{F})^2\frac{1}{n}\sum_{i=1}^{n}(O_i-\bar{O})^2}} \quad (4)$$

Chỉ số CORR dùng để đo mối quan hệ tuyến tính giữa dự báo và quan trắc, chỉ phản ánh quan hệ đồng biến hay nghịch biến. Tự nó không cung cấp thông tin về độ nghiêng của đường hồi quy. Chỉ số nhạy cảm với giá trị lớn, tản mạn, không nhạy cảm với bias. Giá trị CORR nằm trong khoảng (–1, 1) hệ số tương quan càng gần 1 thì càng tốt. Giá trị lý tưởng trong trường hợp này là CORR = 1.

2.4. Các chỉ số đánh giá pha

a) Chỉ số brier score

$$BS = \frac{1}{N}\sum_{i=1}^{N}(F_i - O_i)^2 \quad (5)$$

Brier score là chỉ số đánh giá mức độ chính xác của các dự báo xác suất. Trong đó Fi là xác suất của lần dự báo thứ i và Oi là quan trắc tương ứng với phiên dự báo đó, Oi nhận giá trị là 1 nếu sự kiện xảy ra và 0 nếu sự kiện không xảy ra, N là số lượng các sự kiện được dự báo. BS có giá trị nằm trong khoảng (0–1) với kết quả lý tưởng là 0 – tương ứng với độ chính xác hoàn hảo và giá trị tệ nhất là 1 tương ứng với dự báo sai hoàn toàn. Giá trị BS cho chúng ta biết sai số bình phương trung bình của giá trị dự báo xác suất.

b) Đường cong ROC

Đường cong ROC (*receiver operating characteristic curve*) là đường cong thể hiện mối quan hệ giữa xác suất phát hiện POD (*probability of detection*) và tỷ suất cảnh báo sai FAR (*False alarm rate*) tại một ngưỡng xác suất nào đó. Nếu đường cong nằm dọc theo đường chéo no skill, nó thể hiện dự báo không có kỹ năng. Nếu đường cong càng hướng xa về phía góc trên bên trái của biểu đồ, nó hàm ý dự báo càng có kỹ năng cao (khi POD có xác suất cao hơn FAR). Nếu đường cong nằm dưới đường chéo no skill có nghĩa là kỹ năng dự báo âm (FAR cao hơn POD).

Bảng 1. Tình huống có thể xảy ra trong đánh giá dự báo sự kiện mưa có/không xảy ra

Dự báo	**Quan Trắc**		
	Có	**Không**	**Tổng dự báo**
Có	H	F	H+F
Không	M	CN	M+CN
Tổng quan trắc	H+M	F+CN	H+F+M+CN

Trong đó H = dự báo có, quan trắc có; M = dự báo không, quan trắc có; F = dự báo có, quan trắc không; CN = dự báo không, quan trắc không.

+ Xác suất phát hiện

$$POD = \frac{H}{H+M} \quad (6)$$

POD cho biết khả năng thành công của mô hình, có giá trị trong khoảng (0, 1), POD = 1 là giá trị lý tưởng mô hình được xem là hoàn hảo.

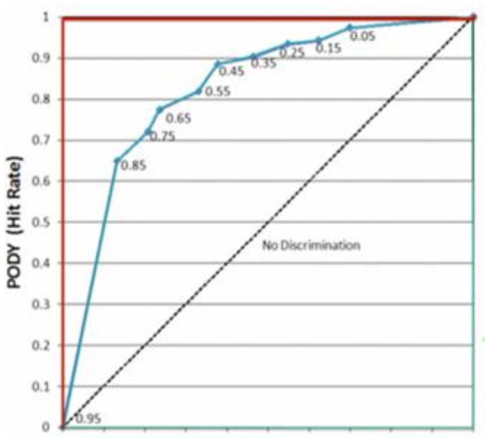
+ Tỷ suất cảnh báo sai

$$FAR = \frac{F}{H+F} \quad (7)$$

FAR cho biết tỷ lệ mô phỏng/dự báo khống của mô hình (mô hình cho kết quả có nhưng thực tế hiện tượng không xảy ra). Giá trị FAR biến đổi từ (0, 1), tối ưu FAR = 0.

c) AUC (*Area under the ROC Curve*)

AUC là khoảng diện tích được tính từ đường cong ROC đến điểm dưới cùng bên phải của đồ thị, giá trị AUC cung cấp thước đo hiệu suất tổng hợp trên tất cả các ngưỡng phân loại có thể. Một cách diễn giải, AUC là xác suất mà mô hình xếp hạng một ví dụ tích cực ngẫu nhiên cao hơn một ví dụ tiêu cực ngẫu nhiên. AUC nằm trong khoảng giá trị từ 0 đến 1. Một mô hình có dự đoán sai 100% có



Hình 2. Đồ thị đường cong ROC

AUC là 0.0; mô hình có dự đoán chính xác 100% có AUC là 1.0. Mô hình không có kỹ năng có giá trị AUC là 0,5.

3. Kết quả và phân tích

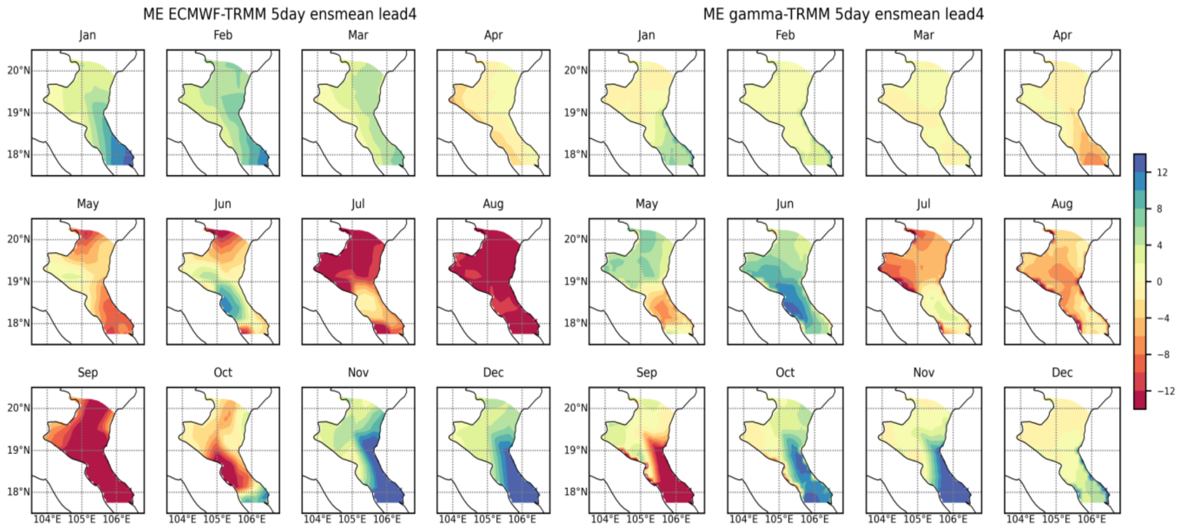
3.1. Kết quả sai số trung bình ME

Như đã trình bày ở chương II, mưa dự báo hạn nội mùa từ mô hình EC có hạn dự báo là 45 ngày. Đối với yếu tố dự báo là tổng lượng mưa 5 ngày, sẽ có 45/5 = 9 hạn dự báo (9 leadtime). Kết quả hiệu chỉnh của từng leadtime sẽ được trình bày dưới đây.

Bảng 2. Kết quả đánh giá ME tổng lượng mưa 5 ngày giữa TRMM và S2S sau hiệu chỉnh cho khu vực Bắc Trung Bộ trong 20 năm từ 2000–2019 đối với các leadtime khác nhau.

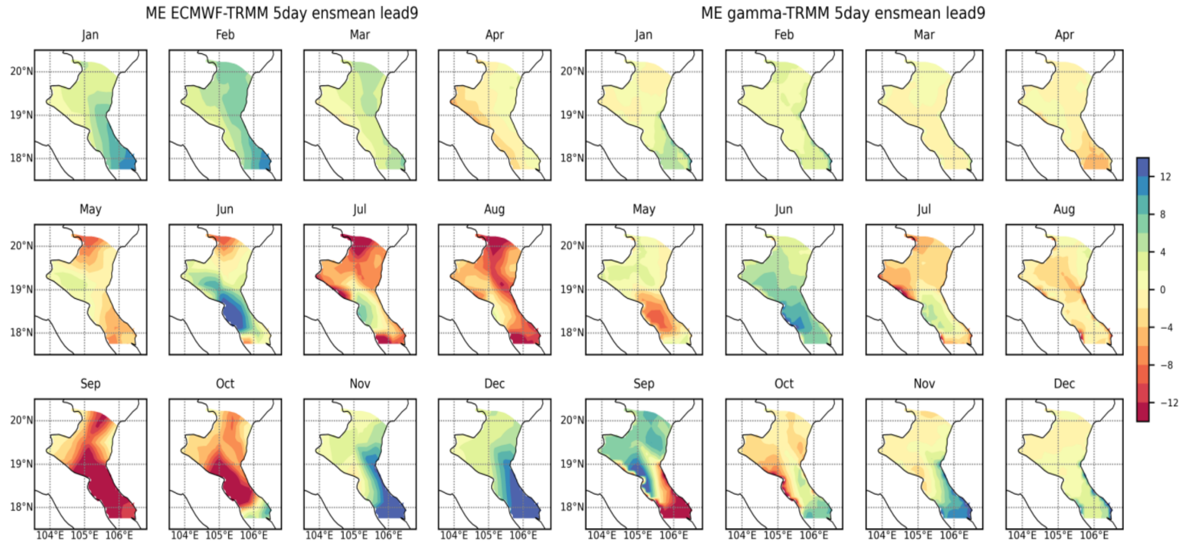
Leadtime	Tháng 1	Tháng 2	Tháng 3	Tháng 4	Tháng 5	Tháng 6	Tháng 7	Tháng 8	Tháng 9	Tháng 10	Tháng 11	Tháng 12
1	3,812	0,242	–1,778	–0,861	0,566	0,452	–2,282	2,166	–3,976	3,434	4,59	1,975
2	2,804	–0,122	–0,102	–1,45	7,573	4,017	–7,959	–1,929	–8,025	2,712	7,366	2,36
3	2,254	0,684	–1,072	–0,738	5,401	8,326	–8,173	–4,805	–9,345	5,389	8,749	0,845
4	1,204	1,429	–0,041	–1,224	2,51	7,264	–4,047	–3,842	–5,462	5,292	5,096	0,765
5	1,492	1,098	–0,813	–0,599	–0,318	4,851	–1,957	–1,414	–1,451	1,787	3,82	0,297
6	1,036	2,017	0,051	–1,622	–2,998	4,356	2,335	–2,03	2,536	–0,33	3,276	–0,286
7	1,186	1,64	–0,923	–0,507	–2,031	4,942	–1,114	–0,799	0,247	1,071	2,492	0,557
8	1,296	1,527	0,009	–1,414	–2,38	5,82	1,011	–3,117	3,493	–0,203	1,248	0,382
9	1,013	1,759	–0,223	–0,921	–0,823	5,794	–1,899	–1,272	3,219	–0,685	1,65	0,585

Từ bảng 2 có thể thấy, sai số ME sau hiệu chỉnh từ tháng 12–3 bé hơn so với tháng 5–10. Tháng 10–2 và tháng 6, ME xu hướng thiên dương (chứng tỏ giá trị dự báo có xu hướng cao hơn giá trị quan trắc). Các tháng còn lại hầu như có xu hướng thiên âm (chứng tỏ giá trị dự báo có xu hướng thấp hơn giá trị quan trắc). Trong giai đoạn từ tuần thứ 3 đến tuần thứ 6 (tương ứng với leadtime 4–9) có thể thấy sai số ME tương đối bé, bé hơn so với các leadtime 1–3.



Hình 3. Kết quả đánh giá ME tổng lượng mưa 5 ngày giữa TRMM và S2S trước và sau hiệu chỉnh cho khu vực Bắc Trung Bộ trong 20 năm từ 2000–2019 đối với leadtime = 4.

Từ hình 3 ta có thể thấy, đối với mưa dự báo trước hiệu chỉnh, từ tháng 11 đến tháng 3, ME có xu hướng thiên dương (chứng tỏ giá trị dự báo có xu hướng cao hơn giá trị quan trắc) 4–8 mm; riêng phía nam khu vực từ tháng 11–12 lớn hơn 12 mm. Từ tháng 5 đến tháng 10, ME có xu hướng thiên âm (chứng tỏ giá trị dự báo có xu hướng thấp hơn giá trị quan trắc); riêng tháng 7–9 nhiều nơi bé hơn –12 mm. Sau khi hiệu chỉnh có thể thấy, hầu hết các tháng có xu hướng thiên dương (trong khi các tháng 12–4 có sai số khá bé 0–2 thì các tháng 6, 9, 10,11 vẫn còn nhiều vùng sai số lớn, tuy nhiên đã giảm đáng kể so với trước khi hiệu chỉnh.



Hình 4. Kết quả đánh giá ME tổng lượng mưa 5 ngày giữa TRMM và S2S trước và sau hiệu chỉnh cho khu vực Bắc Trung Bộ trong 20 năm từ 2000–2019 đối với leadtime = 9.

Từ hình 4 ta có thể thấy, đối với mưa dự báo trước hiệu chỉnh, tương tự như leadtime = 4, từ tháng 11 đến tháng 3, ME có xu hướng thiên dương (chứng tỏ giá trị dự báo có xu hướng cao hơn giá trị quan trắc) 2–8 mm; riêng phía nam khu vực từ tháng 11–12 lớn hơn 12 mm. Từ tháng 4 đến tháng 10, ME có xu hướng thiên âm (chứng tỏ giá trị dự báo có xu hướng thấp hơn giá trị quan trắc) –4 đến –8 mm; riêng từ tháng 8–10 nhiều nơi bé hơn –12 mm. Sau khi hiệu chỉnh, gía trị ME tốt hơn rõ rệt. Có thể thấy, hầu hết các tháng trong năm, ME có giá trị khá bé (phổ biến –4–4 mm); riêng các tháng 5, 7, 9 có nơi ≤12 mm.

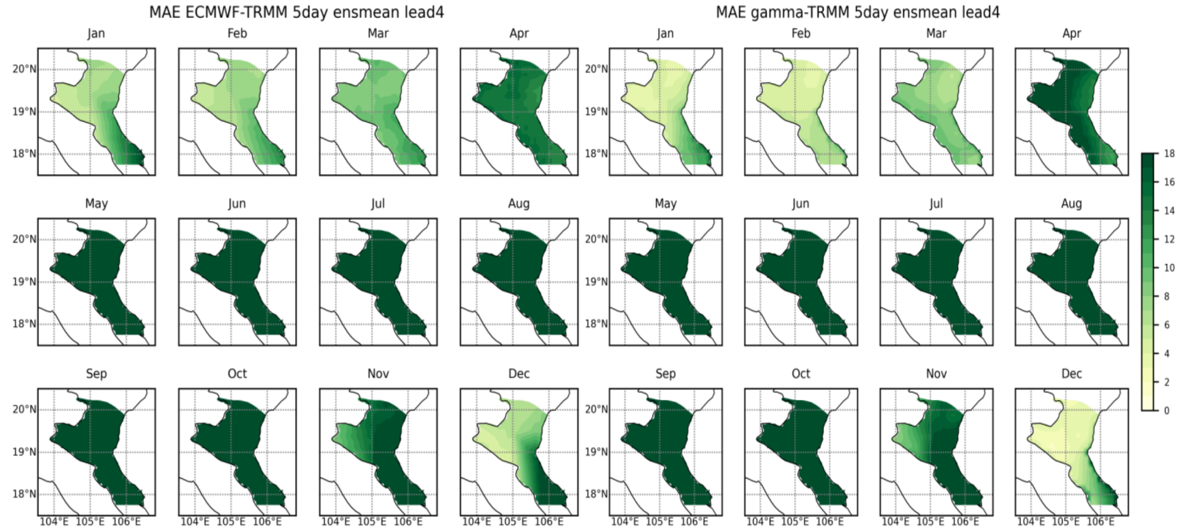
3.2. Kết quả sai số trung bình tuyệt đối MAE, sai số quân phương RMSE đối với tổng lượng mưa dự báo 5 ngày

Từ bảng 3 có thể thấy, sai số MAE và RMSE sau hiệu chỉnh từ tháng 12–4 bé hơn so với tháng 5–11. Trong đó, tháng 2 có giá trị sai số bé nhất, tháng 9 có giá trị sai số lớn nhất. Các sai số trong giai đoạn từ tuần thứ 3 đến tuần thứ 6 (tương ứng với leadtime 4–9) có giá trị lớn hơn so với leadtime 1–3. Các giá trị của RMSE của hầu hết các tháng trong năm không sát với giá trị của MAE chứng tỏ, sai số của mô hình không ổn định, nhiều giá trị tản mạn lớn, đặc biệt là từ tháng 9–10.

Bảng 3. Kết quả đánh giá MAE, RMSE tổng lượng mưa 5 ngày giữa TRMM và S2S sau hiệu chỉnh cho khu vực Bắc Trung Bộ trong 20 năm từ 2000–2019 đối với các leadtime khác nhau.

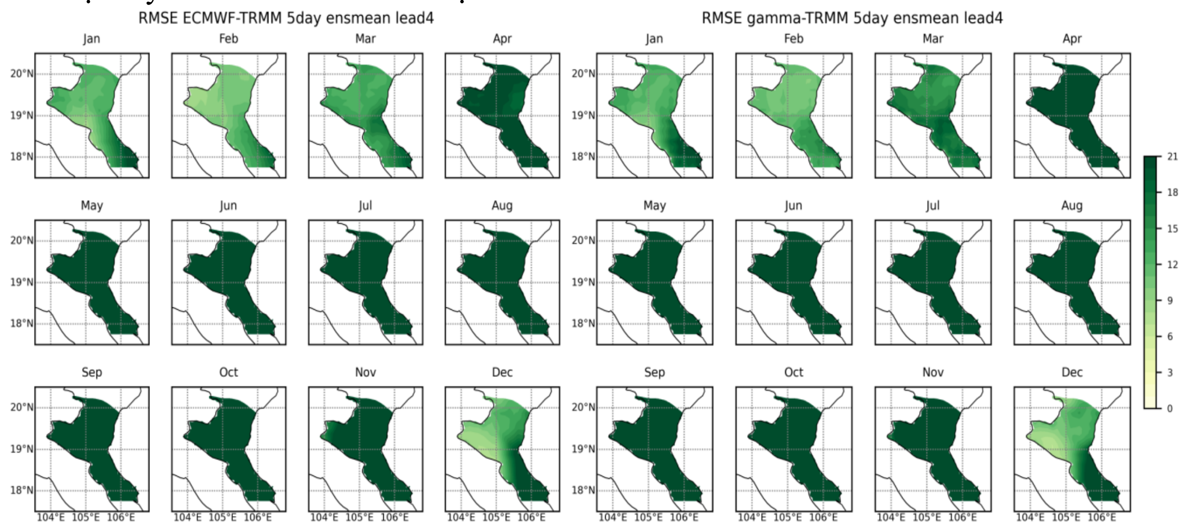
Leadtime		**Tháng 1**	**Tháng 2**	**Tháng 3**	**Tháng 4**	**Tháng 5**	**Tháng 6**	**Tháng 7**	**Tháng 8**	**Tháng 9**	**Tháng 10**	**Tháng 11**	**Tháng 12**
	MAE	6,026	3,092	6,8	15,499	28,811	26,134	28,641	38,404	54,763	40,57	16,903	6,086
1	RMSE	14,371	7,836	14,688	24,796	44,154	41,922	43,826	56,816	83,504	80,372	39,562	17,421
	MAE	5,909	3,655	6,729	16,031	34,275	30,523	33,582	42,825	65,906	55,785	22,663	5,789
2	RMSE	13,518	8,73	14,528	25,02	51,665	46,984	51,992	62,368	101,37	99,161	54,996	19,645
	MAE	5,962	4,793	9,13	17,531	34,518	36,463	36,158	45,826	70,885	57,815	24,777	5,975
3	RMSE	14,496	11,598	18,354	26,499	50,066	55,642	53,813	68,934	105,058	98,78	54,36	17,981
	MAE	5,808	4,99	8,109	16,832	33,085	35,524	38,669	47,29	75,077	63,062	24,986	4,959
4	RMSE	13,808	12,017	16,284	25,919	47,203	53,057	59,29	68,537	110,749	108,103	52,969	15,069
	MAE	5,734	5,465	9,125	17,078	30,051	35,271	42,491	50,024	77,799	63,398	21,201	5,153
5	RMSE	14,11	12,417	17,858	26,365	44,036	52,772	63,756	72,876	114,432	105,293	48,914	15,96
	MAE	5,899	5,303	8,415	17,272	30,878	35,498	42,129	47,871	77,478	56,703	20,142	4,736
6	RMSE	14,47	12,678	16,954	26,61	43,893	53,458	63,54	69,267	115,88	97,267	42,602	13,534
7	MAE	5,687	5,648	9,091	17,349	30,458	33,707	43,222	48,933	78,635	58,662	21,656	4,807

Leadtime		**Tháng 1**	**Tháng 2**	**Tháng 3**	**Tháng 4**	**Tháng 5**	**Tháng 6**	**Tháng 7**	**Tháng 8**	**Tháng 9**	**Tháng 10**	**Tháng 11**	**Tháng 12**
	RMSE	13,489	12,536	17,712	26,606	44,267	50,277	64,039	70,254	112,999	96,143	49,125	14,531
	MAE	5,887	5,328	8,635	16,722	31,22	36,124	42,567	48,755	81,998	60,412	20,532	5,65
8	RMSE	14,029	12,471	17,455	25,807	44,674	54,811	63,67	71,193	120,525	103,209	44,553	15,22
	MAE	5,437	5,566	8,584	17,301	30,665	33,443	43	48,425	81,149	61,792	21,299	4,651
9	RMSE	13,617	12,854	17,2	26,58	43,838	50,058	66,573	69,32	117,639	101,163	46,743	13,459



Hình 5. Kết quả đánh giá MAE tổng lượng mưa 5 ngày giữa TRMM và S2S trước và sau hiệu chỉnh cho khu vực Bắc Trung Bộ trong 20 năm từ 2000–2019 đối với leadtime = 4.

Từ hình 5 ta có thể thấy, đối với mưa dự báo trước hiệu chỉnh, từ tháng 12 đến tháng 3, MAE có giá trị nhỏ hơn các tháng còn lại (6–12 mm), sai số bé. Từ tháng 4 đến tháng 11, MAE có giá trị lớn hơn (>18 mm). Sau khi hiệu chỉnh, gía trị MAE của các tháng 12–3 tốt hơn rõ rệt, sai số giảm xuống còn 4–8 mm. Tuy nhiên từ tháng 4–11, sai số sau hiệu chỉnh ít có sự thay đổi so với trước khi hiệu chỉnh.

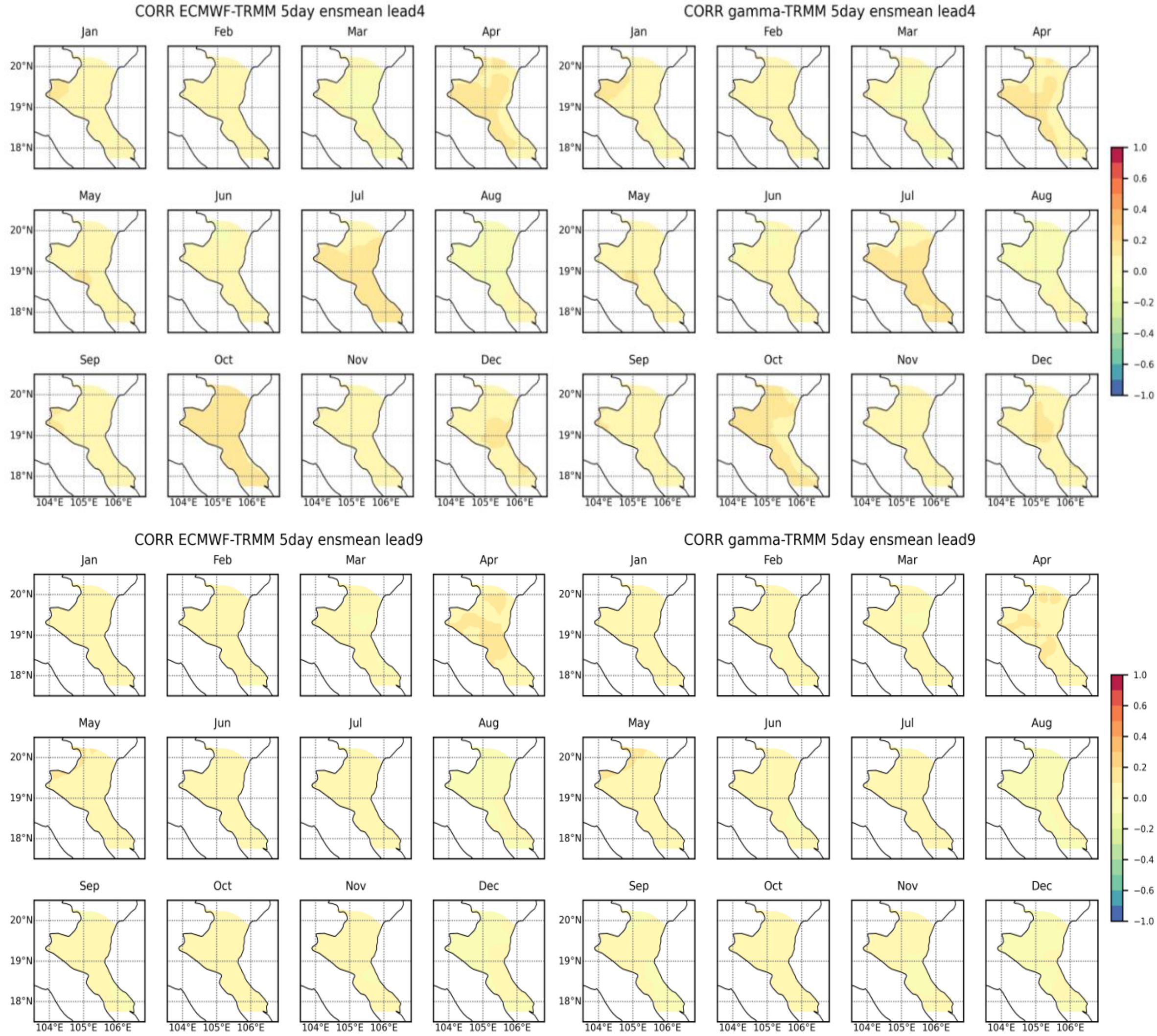


Hình 6. Kết quả đánh giá RMSE tổng lượng mưa 5 ngày giữa TRMM và S2S trước và sau hiệu chỉnh cho khu vực Bắc Trung Bộ trong 20 năm từ 2000–2019 đối với leadtime = 4.

Từ hình 6 ta có thể thấy, đối với mưa dự báo trước hiệu chỉnh, tương tự như MAE, từ tháng 12 đến tháng 3, RMSE có giá trị nhỏ hơn các tháng còn lại (12–18 mm), sai số bé. Từ tháng 4 đến tháng 11, MAE có giá trị lớn hơn (>21 mm). Sau khi hiệu chỉnh, gía trị RMSE của các tháng 12–3 tốt hơn, sai số giảm xuống còn 9–15 mm. Tuy nhiên, từ tháng 4–11, sai số sau hiệu chỉnh ít có sự thay đổi so với trước khi hiệu chỉnh.

3.3. Kết quả hệ số tương quan CORR đối với tổng lượng mưa dự báo 5 ngày

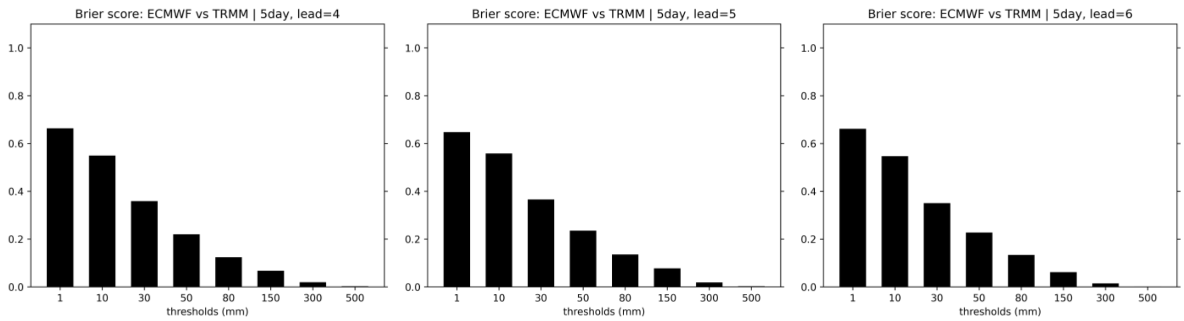
Từ hình 7 có thể thấy, hệ số tương quan của leadtime = 4 tốt hơn một chút so với leadtime = 9. Trước và sau khi hiệu chỉnh, hệ số tương quan của các leadtime khá thấp (phổ biến 0–0,4), chứng tỏ quan trắc và dự báo có quan hệ đồng biến nhưng ít tuyến tính với nhau. Sau khi hiệu chỉnh, hệ số tương quan hầu như không cải thiện hơn so với trước khi hiệu chỉnh.

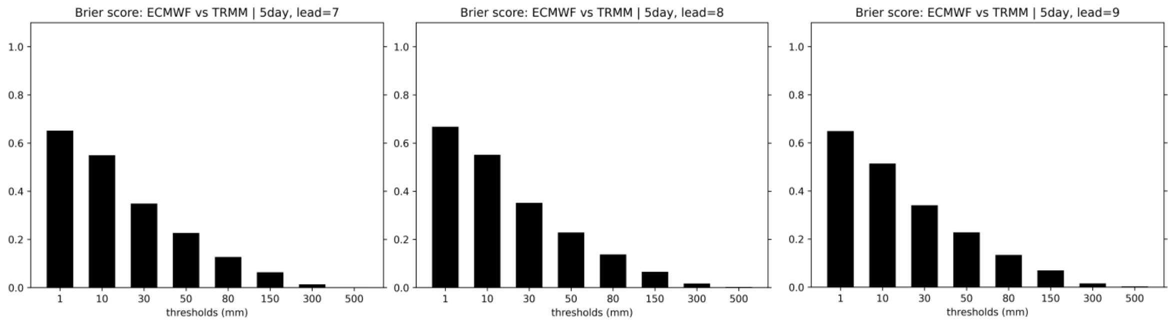


Hình 7. Kết quả đánh giá CORR tổng lượng mưa 5 ngày giữa TRMM và S2S trước và sau hiệu chỉnh cho khu vực Bắc Trung Bộ trong 20 năm từ 2000–2019 đối với leadtime = 4 và 9.

3.4. Kết quả chỉ số BS đối với tổng lượng mưa dự báo 5 ngày

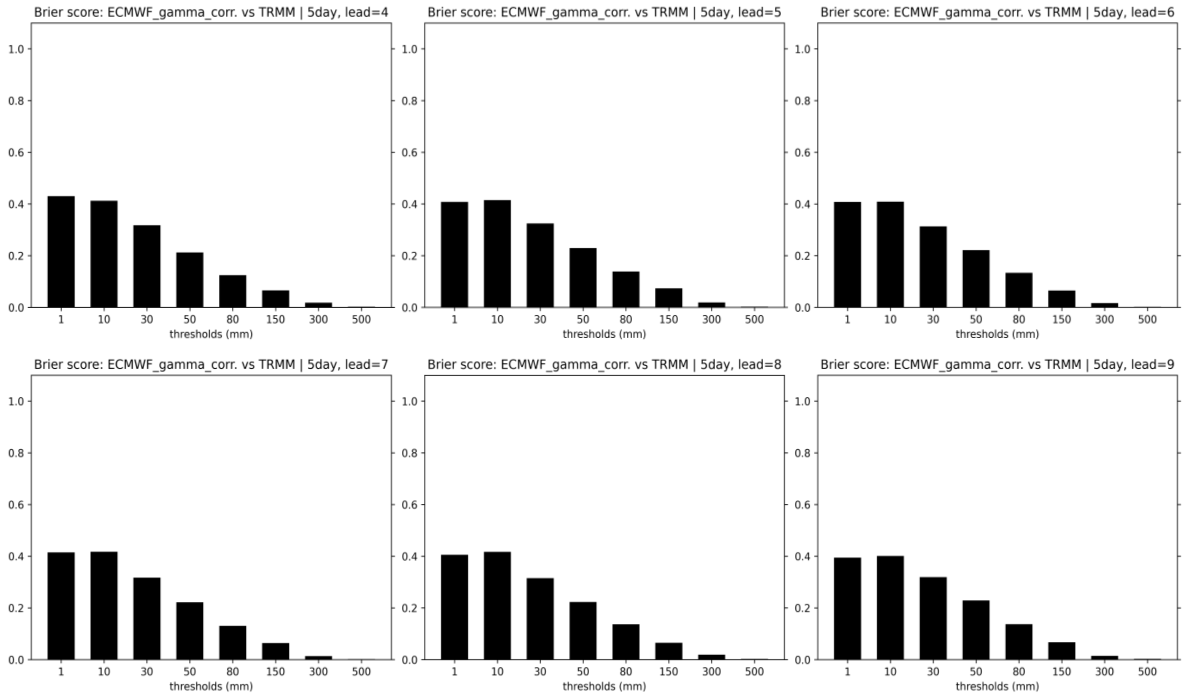
Dưới đây là kết quả đánh giá bằng chỉ số BS đối với tổng lượng mưa dự báo 5 ngày từ leadtime = 4 đến leadtime = 9 trước và sau khi hiệu chỉnh.





Hình 8. Kết quả đánh giá BS đối với tổng lượng mưa 5 ngày giữa TRMM và S2S trước hiệu chỉnh cho khu vực Bắc Trung Bộ trong 20 năm từ 2000–2019 từ leadtime = 4 đến leadtime = 9.

Ta thấy, kết quả của các leadtime từ 4 đến 9 gần như tương đồng nhau, tổng lượng mưa càng cao khả năng phát hiện càng tốt. Cụ thể, đối với lượng mưa thấp (R = 1mm, 10mm, 30mm) thì chỉ số BS đều cho giá trị cao, điều nay cho thấy khả năng phát hiện lượng mưa, mưa nhỏ của mô hình là tương đối kém. Tuy nhiên, với lượng mưa R = 50 mm, giá trị BS đã cải thiện trông thấy khi chỉ lớn hơn 0,2 một chút, cho thấy kỹ năng cải thiện hơn đối với việc phát hiện các lượng mưa vừa và mưa rào. Với các lượng mưa tăng dần R= 80mm, 150mm và 300mm, giá trị của chỉ số BS giảm dần từ 0,1, cho thấy sai số đối với dự báo pha ở các ngưỡng mưa to đến rất to là rất thấp.

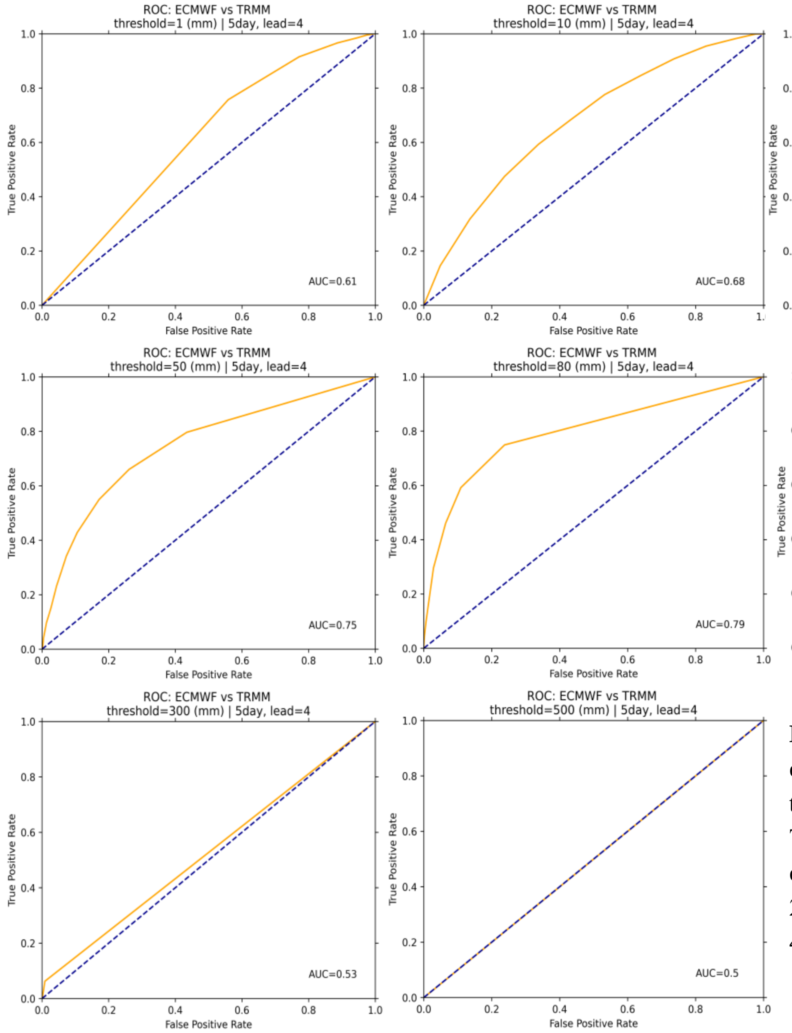


Hình 9. Kết quả đánh giá BS đối với tổng lượng mưa 5 ngày giữa TRMM và S2S sau hiệu chỉnh cho khu vực Bắc Trung Bộ trong 20 năm từ 2000–2019 từ leadtime = 4 đến leadtime = 9.

Ta thấy, kết quả của các leadtime từ 4 đến 9 sau hiệu chỉnh cũng giống như trước hiệu chỉnh, gần như tương đồng nhau, tổng lượng mưa càng cao khả năng phát hiện càng tốt. Tuy nhiên, trong khi đối với lượng mưa lớn (R > 30 mm), chỉ số BS sau hiệu chỉnh gần như không thay đổi so với trước hiệu chỉnh thì đối lượng lượng mưa bé (R = 1mm, 10mm) cho thấy khả năng phát hiện tốt hơn đáng kể, nhất là R = 1 mm.

3.5. Kết quả đánh giá đường cong ROC đối với tổng lượng mưa 5 ngày

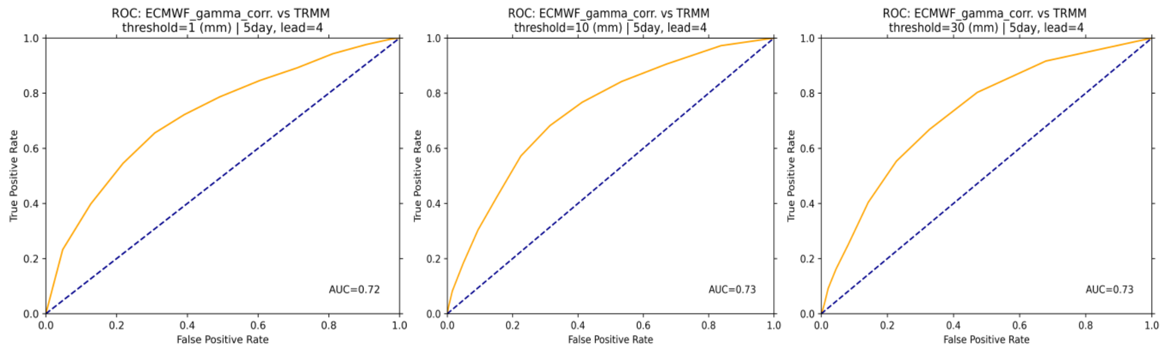
Dưới đây là kết quả đánh giá đường cong ROC đối với các ngưỡng mưa khác nhau của leadtime = 4.

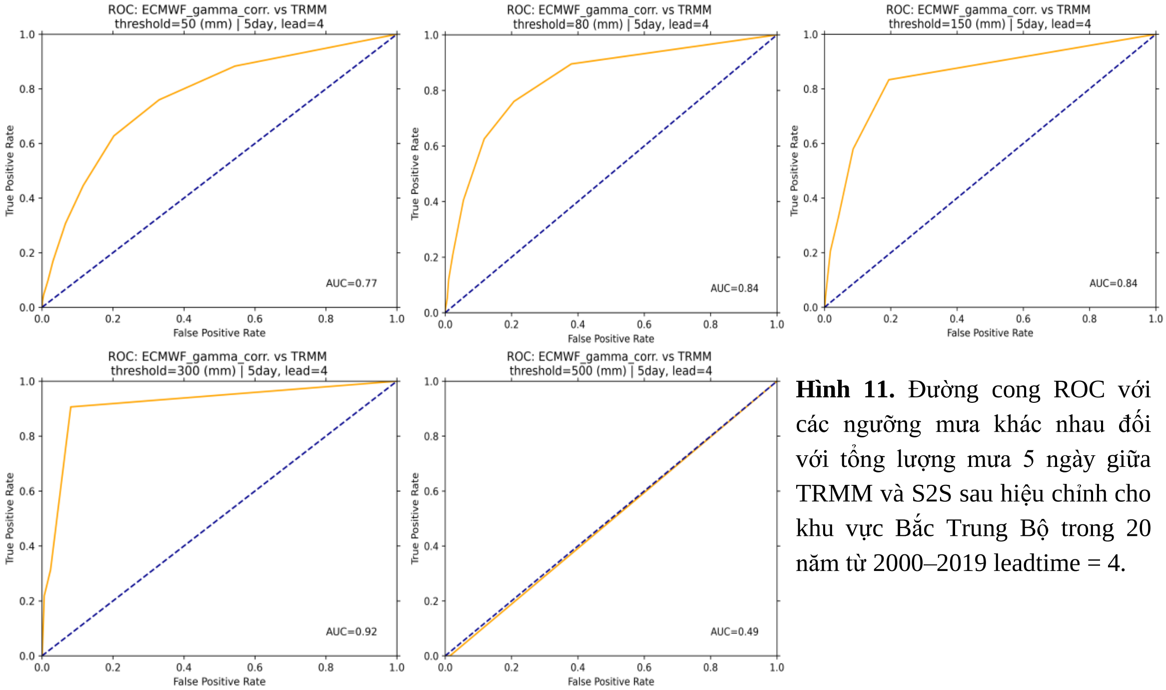


Hình 10. Đường cong ROC với các ngưỡng mưa khác nhau đối với tổng lượng mưa 5 ngày giữa TRMM và S2S trước hiệu chỉnh cho khu vực Bắc Trung Bộ trong 20 năm từ 2000–2019 leadtime = 4.

Trước hiệu chỉnh, có thể thấy, với lượng mưa từ 1 đến 150mm, ROC curve chủ yếu nằm ở phía trên bên trái của đường chéo no skill, điều nay cho thấy mô hình có kỹ năng trong việc dự báo các lượng mưa này, tuy nhiên với giá trị AUC từ 0,63–0,75, cho thấy kỹ năng dự báo ở mức trung bình. Với các lượng mưa lớn, từ 300–500 mm, ta có thể thấy mô hình thể hiện không có kỹ năng với đường cong ROC gần như trùng với đường chéo no skill và giá trị AUC = 0,5–0,53.

Sau hiệu chỉnh, cũng giống như trước hiệu chỉnh, với lượng mưa từ 1 đến 300 mm, ROC curve chủ yếu nằm ở phía trên bên trái của đường chéo no skill, tuy nhiên với giá trị AUC đã tăng lên đáng kể từ 0,72–0,92, đặc biệt là với R = 300 mm, giá trị của AUC = 0,92. Với lượng mưa 500 mm, gần như không có sự thay đổi sau hiệu chỉnh, thể hiện không có kỹ năng với đường cong ROC gần như trùng với đường chéo no skill và giá trị AUC = 0,49.





Hình 11. Đường cong ROC với các ngưỡng mưa khác nhau đối với tổng lượng mưa 5 ngày giữa TRMM và S2S sau hiệu chỉnh cho khu vực Bắc Trung Bộ trong 20 năm từ 2000–2019 leadtime = 4.

4. Kết luận

Qua kết quả bước đầu nghiên cứu hiệu chỉnh dự báo mưa hạn nội mùa của mô hình ECMWF với các hạn dự báo khác ở khu vực Bắc Trung Bộ, nhóm nghiên cứu rút ra một số kết luận như sau:

Sau khi hiệu chỉnh, các giá trị sai số như ME, MAE, RMSE giảm đi rõ rệt, đặc biệt là từ tháng 12 đến tháng 3 năm sau. Riêng tháng 7–9, tuy sai số đã giảm so với trước hiệu chỉnh, tuy nhiên vẫn còn nhiều vùng (chủ yếu là phía Nam khu vực) vẫn còn tồn tại sai số lớn

Các sai số với hạn dự báo từ tuần thứ 3 đến tuần thứ 6 có giá trị lớn hơn so với hạn dự báo từ tuần 1 đến tuần 2. Các giá trị của RMSE của hầu hết các tháng trong năm không sát với giá trị của MAE chứng tỏ, sai số của mô hình không ổn định, nhiều giá trị tản mạn lớn, đặc biệt là từ tháng 9–10.

Trước và sau khi hiệu chỉnh, hệ số tương quan khá thấp, chứng tỏ quan trắc và dự báo có quan hệ đồng biến nhưng ít tuyến tính với nhau. Sau khi hiệu chỉnh, hệ số tương quan không cải thiện hơn, thậm chí có tháng còn giảm so với trước khi hiệu chỉnh.

Kết quả đánh giá chỉ số BS của các leadtime trước và sau hiệu chỉnh gần như tương đồng nhau, tổng lượng mưa càng cao khả năng phát hiện càng tốt. Tuy nhiên, trong khi đối với lượng mưa lớn (R = 30mm, 50mm, 80mm, 300mm, 500mm), chỉ số BS sau hiệu chỉnh gần như không thay đổi so với trước hiệu chỉnh thì đối lượng lượng mưa bé (R = 1mm, 10mm) cho thấy khả năng phát hiện tốt hơn đáng kể, nhất là R = 1mm.

Trước và sau khi hiệu chỉnh, với lượng mưa từ 1 đến 300mm, ROC curve chủ yếu nằm ở phía trên bên trái của đường chéo no skill, tuy nhiên với giá trị AUC sau hiệu chỉnh đã tăng lên đáng kể. Với lượng mưa 500mm, gần như không có sự thay đổi sau hiệu chỉnh, thể hiện không có kỹ năng với đường cong ROC gần như trùng với đường chéo no skill.

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

Lời cảm ơn: Tập thể tác giả xin trân trọng cảm ơn Đề tài "Nghiên cứu xây dựng công nghệ dự báo mưa hạn nội mùa từ một số mô hình toàn cầu kết hợp với mô hình thủy văn để phục vụ quản lý nguồn nước, điều tiết hồ chứa cho khu vực Bắc Trung Bộ", mã số TNMT.2022.02.17 đã hỗ trợ nhóm nghiên cứu về số liệu và phương pháp luận.

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. De Andrade, F.M.; Young, M.P.; Macleod, D.; Hirons, L.C.; Woolnough, S.J.; Black, E. Subseasonal Precipitation Prediction for Africa: Forecast Evaluation and Sources of Predictability, National Centre for Atmospheric Science, University of Reading, Reading, United Kingdom, School of Geographical Sciences, University of Bristol, Bristol, United Kingdom, 2020.
2. Liu, X.; Li, J.; Jie, W.; Huang, L.; Gu, W. (), Subseasonal Predictions of Regional Summer Monsoon Rainfall over Tropical Asian Oceans and Land. *J. Clim.* **2015**, 28(24), 9583–9605. https://doi.org/10.1175/JCLI-D-14-00853.1
3. Li, S.; Robertson, A.W. Evaluation of Submonthly Precipitation Forecast Skill from Global Ensemble Prediction Systems. *Mon. Weather Rev.* **2015**, 143(7), 2871–2889. https://doi.org/10.1175/MWR-D-14-00277.1.
4. Enayati, M.; et al. Bias correction capabilities of quantile mapping methods for rainfall and temperature variables. *J. Water Clim. Change* **2021**, 401–419.
5. Ines, A.V.; Hansen, J.W.; Robertson, A.W. Enhancing the utility of daily GCM rainfall for crop yield prediction. *Int. J. Climatol.* **2011**, *31*, 2168–2182.
6. Zhao, T.; Bennett, J.C.; Wang, Q.J.; Schepen, A.; Wood, A.W.; Robertson, D.E.; Ramos, M.H. How Suitable is Quantile Mapping For Postprocessing GCM Precipitation Forecasts? *J. Clim.* **2017**, *30*, 3185–3196.
7. Khiêm, M.V. Nghiên cứu phương pháp hiệu chỉnh sản phẩm mưa dự báo hạn mùa cho khu vực Việt Nam. *Tạp chí Khoa học ĐHQGHN: Các Khoa học Trái đất và Môi trường* 2018, 34(1S), 33–40.
8. Yến, N.T.H.; Nam, P.Q.; Nam, N.Đ.; Hiệp, N.V. Dự tính số ngày nắng nóng cho Việt nam bằng tổ hợp số liệu hiệu chỉnh phân vị đa mô hình động lực. *Tạp chí Khí tượng Thủy văn* **2017**, *683*, 12–20.
9. Linh, L.N.; Khiêm, M.V. Nghiên cứu phương pháp thống kê hiệu chỉnh lượng mưa mô hình. *Tạp chí Khí tượng Thủy văn* **2016**, *664*, 43–49.
10. Thom, H.C. A note on the gamma distribution. *Mon. Weather Rev.* **1958,** *86*, 117–122.
11. Fang, G.H.; Yang, J.; Chen, Y.N.; Zammit, C. Comparing bias correction methods in downscaling meteorological variables for a hydrologic impact study in an arid area in China. *Hydrol. Earth Syst. Sci.* **2015**, *19*, 2547.
12. Stockdale, T. An overview of techniques for season forecasting. *Stochastic Eviron. Res. Assess.* **2000**, *14*, 305–318.
13. Vitarta, F.; Robertson, A.W. Why Sub–seasonal to Seasonal Prediction (S2S)?. Sub–seasonal to Seasonal Prediction, 2018, pp. 3–15.
14. Hui, D.T.W.; Shum, K.K.Y.; Chen, J.; Chen, S.C.; Ritchie, J.; Roads, J.O. Case studies of seasonal rainfall forecasts for Hong Kong and its vicinity using a regional climate model. Springer Science+Business Media B.V, 2007.
15. Argüeso, D.; Evans, J.P.; Fita, L. Precipitation bias correction of very high resolu–tion regional climate models. *Hydrol. Earth Syst. Sci.* **2013**, *17*, 4379–4388. Doi:10.5194/hess-17-4379-2013.
16. Bennett, J.C.; Grose, M.R.; Corney, S.P.; White, C.J.; Holz, G.K.; Katzfey, J.J.; Post, D.A.; Bindoff, N.L. Performance of an empirical bias–correction of a high–resolution cli–mate dataset. *Int. J. Climatol.* **2014**, *34*, 2189–2204. Doi:10.1002/joc.3830.
17. Christensen, J.H.; Boberg, F.; Christensen, O.B.; Lucas–Picher, P. On the need for bias correction of regional climate change projections of temperature and

precipitation. *Geophys. Res. Lett.* **2008**, *35*, L20709. Doi:10.1029/2008GL035694.
18. Brier, W.; Panofsky, H. Some applications of statistics to meteorology. Mineral In–ustries Extension Services", School of Mineral Industries, Pennsylvania State College, 1968.
19. Gao, Y.; Fu, J.S.; Drake, J.B.; Liu, Y.; Lamarque, J.F. Projected changes of extreme weath–events in the eastern United States basedon a high resolution climate modelingsystem, 2012.
20. Gutjahr, O.; Heinemann, G. Comparing precipitation bias correction methods for high– resolution regional climate simulations using COSMO–CLM. *Theor. Appl. Climatol.* **2013**, *114*, 511–529.

Application of the quantile mapping method with gamma approximation (QM–G) for sub–seasonal to seasonal rainfall forecast over North central region

Hoang Thi Thu Huong[1]*, Nguyen Van Luong[1], Phan Van Vinh[1], Pham Thanh Ha[2]

[1] North Central regional hydro–meteorology center; hoanghuong.btb@gmail.com; luongnvkttv@gmail.com; Vinhpv@gmail.com@gmail.com
[2] VNU University of Science–Vietnam National University, Hanoi; phamthanhha5693@gmail.com

Abstract: The timescale from 10 days to 2 months (sub–seasonal to seasonal) is important in water resource management, reservoir regulation, both irrigation and hydroelectricity, as well as agricultural scheduling in the North Central region. However, it is also impossible to deny the fact that, in Vietnam in general and the North Central region in particular, the sub season forecasting skills are still low. The authors corrected the sub–seasonal to seasonal rainfall forecast of the ECMWF model based on the quantile mapping method with gamma approximation (QM–G) for the total rainfall of 5 days with each forecast period (3–6 weeks). Two main data sets include: the ECMWF model's sub–seasonal rainfall forecast data set for 20 years in the period 2000–2019 and the TRMM satellite rain data set. The results show that: After correcting, the ME, MAE, RMSE values decreased significantly, especially from December to March next year. Besides, in July–September, although the bias has decreased compared to before the correction, there are still many areas (mainly in the southern region) that still have large bias. However, the correction by QM–G did not improve the correlation, even some month it decreased compared to before correction.

Keywords: Sub–seasonal to seasonal; Bias correction; ECMWF; QM–G, North Central Region.