

# SO SÁNH, ĐÁNH GIÁ CÁC MÔ HÌNH THỐNG KÊ HIỆU CHỈNH DỰ BÁO LƯỢNG MƯA SỬ DỤNG CÁC CHỈ SỐ HOÀN LƯU VÀ NHIỆT ĐỘNG LỰC KHÍ QUYỂN

Hoàng Phúc Lâm<sup>1</sup>, Nguyễn Đức Nam<sup>2</sup>, Lê Minh Đức<sup>3</sup>

**Tóm tắt:** Trong bài báo này, nhóm nghiên cứu đánh giá và so sánh chất lượng hiệu chỉnh dự báo của các mô hình thống kê sử dụng các chỉ số hoàn lưu khí quyển với các yếu tố dự báo là các loại số liệu lượng mưa khác nhau, bao gồm: mưa quan trắc, mưa tái phân tích ERA và mưa ước lượng từ vệ tinh GPCP. Nghiên cứu đã đánh giá được tương quan của lượng mưa với các giá trị tốc độ thẳng đứng, độ phân kỳ và giá trị xoáy thế trên các mực khác nhau ở các khu vực Bắc Bộ. Bên cạnh đó, nghiên cứu đã xây dựng được 9 phương trình hồi quy tuyến tính đa biến cho 3 loại số liệu mưa và cho 3 khu vực ở Bắc Bộ. Kết quả đánh giá độc lập cho thấy khi so sánh lượng mưa tính toán với số liệu mưa tái phân tích ERA cho sai số nhỏ nhất, giá trị tương quan cao, có điểm lưới đạt giá trị tương quan trên 0,9.

**Từ khóa:** Xoáy thế, định lượng mưa, thống kê sau mô hình, hiệu chỉnh.

Ban Biên tập nhận bài: 08/12/2018 Ngày phản biện xong: 18/02/2019 Ngày đăng bài: 25/02/2019

## 1. Mở đầu

Ở vùng nhiệt đới, Berry và cs (2012) [1], Lam Hoang và cs (2019) [6] đã sử dụng xoáy thế trên các mặt đẳng nhiệt độ thế vị để phân tích và theo dấu các nhiễu động xoáy thuận trên khu vực Châu Úc và chỉ ra rằng mưa vùng nhiệt đới thường liên quan đến xoáy thế trên các mực đẳng nhiệt độ thế vị. Mỗi quan hệ giữa xoáy thế và dông hay các hệ thống đối lưu cũng được Hertenstein [8] và Nielsen-Gammon [3]. Ở khu vực Châu Á, năm 1996, Du Jun và cs (1996) [4] đã nghiên cứu về các hệ thống đối lưu quy mô vừa dọc front Baiu và chỉ ra rằng, nguyên nhân tạo ra những ổ mây đối lưu này là sự bất ổn định của xoáy thế cực đại tầng thấp dọc front. Tác giả cũng chỉ ra rằng sự di chuyển của các hệ thống đối lưu quy mô vừa từ tây sang đông có dạng sóng với bước sóng 1700 - 2100 km và ít phụ thuộc vào kích thước của dị thường xoáy thế.

Ở Việt Nam, Hoàng Phúc Lâm và cs [5] cũng

đã sử dụng xoáy thế, tốc độ thẳng đứng, độ phân kỳ trên các mực đẳng nhiệt độ thế vị và mực đẳng áp là nhân tố dự báo trong bài toán xây dựng mô hình hồi quy từng bước dự báo mưa cho khu vực Bắc Bộ với yếu tố dự báo là lượng mưa quan trắc 24 giờ. Kết quả đánh giá và so sánh lượng mưa hiệu chỉnh với lượng mưa tái phân tích cho thấy hệ số tương quan đạt 0,6 đến 0,8, tuy nhiên sai số RMSE còn lớn, từ 4,6 đến 6,1 mm/ngày.

Bài toán hiệu chỉnh dự báo mưa của mô hình số ra đời đã giúp cải thiện đáng kể sai số của mô hình. Hiện tại, thống kê sau mô hình (MOS) đang được tiếp tục phát triển theo hai hướng mới là GMOS (dự báo thống kê sau mô hình trên lưới) và EMOS (dự báo thống kê sau mô hình tổ hợp). Nếu như theo phương pháp MOS thông thường ta chỉ xây dựng phương trình dự báo tại trạm, sau đó cũng đưa ra dự báo tại điểm trạm thì GMOS hướng đến sử dụng MOS dự báo trên

<sup>1</sup>Trung tâm Dự báo khí tượng thủy văn quốc gia, Tổng cục KTTV

<sup>2</sup>Viện Vật lý Địa cầu, Viện Hàn lâm Khoa học và Công nghệ Việt Nam

<sup>3</sup>Học viên Trường Đại học Tài nguyên và Môi trường

Email: lamhpvn@gmail.com

lưới lưới (Dallavalle và Glahn 2005) [2]. Nếu như thám sát được cho trên lưới thì GMOS chính là phương pháp MOS thông thường. Nếu các thám sát phân bố không đều, sau khi có dự báo MOS tại trạm, sử dụng các kỹ thuật phân tích hai chiều để chuyển các dự báo này về lưới. Với EMOS, thay vì sử dụng MOS cho một mô hình người ta sử dụng các phương pháp thống kê dựa trên dự báo từ nhiều mô hình. Trong nghiên cứu này sẽ sử dụng phương pháp GMOS để dự báo mưa trên lưới. Trong nỗ lực tăng cường chất lượng dự báo, đã có một vài nghiên cứu trong nước thử nghiệm lọc Kalman vào hiệu chỉnh dự báo từ mô hình như của Võ Văn Hòa và cs [9]. Mặc dù phương trình thống kê nhóm tác giả này sử dụng còn đơn giản, kết quả hiệu chỉnh đã cho thấy những cải tiến đáng kể cho thấy khả năng phát triển dự báo thống kê sau mô hình ở Việt Nam. Từ những cơ sở đó, nghiên cứu này đánh giá, so sánh các mô hình thống kê khác nhau hiệu chỉnh lượng mưa dự báo của mô hình số trị căn cứ vào mối quan hệ giữa các chỉ số hoàn lưu và nhiệt động lực khí quyển trên các mặt đẳng nhiệt độ thế và lượng mưa.

## 2. Số liệu và phương pháp

Phương pháp nghiên cứu và hầu hết các loại số liệu trong nghiên cứu này giống với số liệu trong nghiên cứu của Hoàng Phúc Lâm và cs [5], hai bộ số liệu mưa từ GPCP và số liệu tái phân tích ERA được bổ sung sử dụng làm yếu tố dự báo trong các mô hình thống kê để so sánh, đánh giá với mô hình xây dựng trên số liệu quan trắc. Các số liệu ở dạng mảng 2 chiều [3 x 4383], chiều thứ nhất là chiều không gian gồm 3 khu vực Tây Bắc, Việt Bắc - Đông Bắc và Đồng bằng Bắc Bộ; chiều thứ hai là thời gian theo từng ngày từ năm 1997 - 2008 (4383 ngày). Số liệu đều được đưa về định dạng NetCDF để thuận tiện trong xử lý tính toán.

## 3. Kết quả và thảo luận

### 3.1 Kết quả tính tương quan giữa các yếu tố

Trong phần này, giá trị tương quan của lượng mưa tương ứng trong ba loại số liệu mưa là mưa quan trắc ( $R_{obs}$ ), mưa tái phân tích ( $Tp_{era}$ ) và mưa vệ tinh GPCP ( $pcp$ ) lần lượt với các yếu tố

nhiệt động lực cho trong Bảng 1, từ thứ 4 đến thứ 18 của ba phân vùng trên khu vực Bắc Bộ. Kết quả được trình bày trong bảng 1.

Xét tương quan giữa các loại số liệu mưa, qua phân tích tổng quát bảng 1 cho thấy số liệu tái phân tích ERA có tương quan tốt nhất với mưa quan trắc, giá trị tương quan ở các khu vực B1, B2, B3 lần lượt là 0,54 và 0,6. Trong khi đó, tương quan với mưa quan trắc của mưa vệ tinh GPCP kém hơn, với giá trị tương ứng là 0,45, 0,49 và 0,52. Điều này có thể do số liệu mưa vệ tinh GPCP có độ phân giải thô hơn so với số liệu mưa tái phân tích là  $1 \times 1$  độ so với  $0,25 \times 0,25$  nên khi tính toán lượng mưa trung bình khu vực sẽ dẫn đến chênh lệch giá trị so với khi tính toán lượng mưa trung bình khu vực với mưa tái phân tích.

Về tương quan của mưa với các yếu tố khác, trong các yếu tố được liệt kê ở trên thì lượng mưa tương quan tốt nhất với tốc độ thẳng đứng ở mực 925mb và 850mb, trong đó tốc độ thẳng đứng trên mực 925mb có tương quan tốt hơn ở khu vực B1 và B2, nhưng riêng ở khu vực B3 thì mưa lại có tương quan tốt với tốc độ thẳng đứng mực 850mb hơn là ở mực 925mb. Ngoài ra ta cũng có thể thấy tương quan của tốc độ thẳng đứng mực 700mb với lượng mưa ở vùng B3 luôn lớn hơn so với tương ứng ở hai vùng B1 và B2. Điều này có thể cho thấy mưa ở khu vực B3 thường liên quan đến đối lưu mạnh hơn so với khu vực B1 và B2. Xét mối tương quan của lượng mưa với độ phân kỳ thì tương quan âm và kém hơn, giá trị tuyệt đối nhỏ hơn 0,5.

Xét tương quan của lượng mưa với các giá trị xoáy thế, có thể thấy rõ tương quan của lượng mưa tốt nhất là với xoáy thế trên mực đẳng nhiệt độ thế 330K sau đó là đến xoáy thế trên mực 315K, và tương quan ở khu vực B3 cũng tốt hơn ở khu vực B1 và B2; điều này một lần nữa khẳng định mưa ở khu vực B3 thường do đối lưu sâu hơn so với khu vực B1 và B2. Với xoáy thế trên mực 350K, đại diện cho giá trị xoáy thế trên tầng cao, thì hầu như không có tương quan với lượng mưa có giá trị tương quan tuyệt đối nhỏ hơn 0,1. Tương quan của chuỗi giá trị xoáy thế

cực đại trong ngày thường có tương quan thấp hơn so với giá trị xoáy thể trung bình ngày.

Khi so sánh tương quan ba loại số liệu mưa với các yếu tố hoàn lưu và nhiệt động lực khí quyển, có thể thấy tương quan của lượng mưa tái phân tích ERA thường có giá trị tương quan tuyệt đối lớn hơn so với lượng mưa quan trắc và

lượng mưa vệ tinh GPCP. Nhưng đối với tương quan của mưa và các giá trị xoáy thể thì chênh lệch giá trị tương quan giữa ba loại số liệu mưa là không nhiều. Ví dụ đối với xoáy thể trung bình ngày trên mực 330 (pv\_330\_tbdays) ở khu vực B2 thì giá trị tương quan với ba loại số liệu mưa là 0,35, 0,35 và 0,32.

Bảng 1. Kết quả tính tương quan giữa các yếu tố

Khu vực Tây Bắc (B1)																		
	r_obs	tp_era	pcp	w_925	w_850	w_700	w_500	div_315	div_330	div_350	pv_315_max	pv_330_max	pv_350_max	pv_tbdays3lv	pv_315_tbdays	pv_330_tbdays	pv_350_tbdays	pv_tbdays3lv
R_obs	1.00	0.54	0.45	-0.46	-0.36	-0.01	0.23	-0.20	-0.07	0.09	0.15	0.26	-0.01	0.12	0.17	0.32	0.02	0.18
T_era	0.54	1.00	0.54	-0.77	-0.68	-0.20	0.19	-0.32	-0.05	0.21	0.09	0.26	0.00	0.12	0.08	0.31	0.03	0.16
Pcp	0.45	0.54	1.00	-0.46	-0.27	0.06	0.31	-0.18	-0.08	-0.04	0.09	0.23	-0.01	0.10	0.12	0.30	0.02	0.16
Khu vực Việt Bắc - Đông Bắc (B2)																		
	r_obs	tp_era	pcp	w_925	w_850	w_700	w_500	div_315	div_330	div_350	pv_315_max	pv_330_max	pv_350_max	pv_tbdays3lv	pv_315_tbdays	pv_330_tbdays	pv_350_tbdays	pv_tbdays3lv
R_obs	1.00	0.60	0.49	-0.46	-0.45	-0.12	0.13	-0.28	0.15	0.01	0.14	0.28	-0.02	0.12	0.17	0.35	0.00	0.18
T_era	0.60	1.00	0.55	-0.70	-0.67	-0.27	0.14	-0.40	0.15	0.10	0.09	0.29	0.00	0.13	0.11	0.35	0.03	0.19
Pcp	0.49	0.55	1.00	-0.42	-0.31	0.03	0.25	-0.26	0.05	-0.06	0.11	0.27	0.01	0.13	0.14	0.32	0.03	0.18
Khu vực Đồng Bằng Bắc Bộ (B3)																		
	r_obs	tp_era	pcp	w_925	w_850	w_700	w_500	div_315	div_330	div_350	pv_315_max	pv_330_max	pv_350_max	pv_tbdays3lv	pv_315_tbdays	pv_330_tbdays	pv_350_tbdays	pv_tbdays3lv
R_obs	1.00	0.60	0.52	-0.39	-0.46	-0.36	-0.17	-0.29	0.22	-0.06	0.20	0.33	0.01	0.19	0.22	0.38	0.02	0.23
T_era	0.60	1.00	0.61	-0.62	-0.76	-0.54	-0.23	-0.43	0.31	-0.04	0.22	0.42	0.05	0.25	0.23	0.47	0.07	0.30
Pcp	0.52	0.61	1.00	-0.38	-0.41	-0.28	-0.07	-0.39	0.21	-0.15	0.18	0.35	0.01	0.19	0.20	0.40	0.04	0.25

**3.2 Lựa chọn mô hình hồi quy đa biến**

Nghiên cứu tiến hành lựa chọn các thành phần của mô hình hồi quy đa biến bằng phương pháp hồi quy từng bước. Từ đó ta sẽ chọn được

các biến tốt nhất để xây dựng mô hình tối ưu nhất. Sau đây là kết quả sau khi sử dụng phương pháp hồi quy từng bước.

*Bảng 2. Kết quả hồi quy từng bước, các yếu tố được chọn cho mỗi phương trình được đánh dấu [ x ], không chọn không đánh dấu [ ]*

STT	Y	r_obs			tp_era			pcp		
	Khu vực	B1	B2	B3	B1	B2	B3	B1	B2	B3
	X1..n									
1	w 925	x	x	x	x	x	x	x	x	x
2	w 850	x	x	x	x	x	x	x	x	
3	w 700	x	x	x	x	x	x	x		x
4	w 500	x	x	x	x		x	x	x	x
5	div 315		x	x	x	x	x	x	x	x
6	div 330		x	x	x		x			x
7	div 350	x	x	x		x	x	x	x	x
8	pv 315 tbdays	x	x	x	x			x	x	x
9	pv 330 tbdays	x	x	x	x	x	x	x	x	x
10	pv 350 tbdays						x	x	x	
11	pv tbdays3lv									x
12	pv 315 max				x	x	x	x	x	
13	pv 330 max	x	x	x	x				x	
14	pv 350 max				x					
15	pv tbmax3lv	x	x	x	x		x	x	x	x

Bảng 2 đưa ra danh sách 15 nhân tố dự báo, tên mô tả các biến đã được liệt kê ở bảng 1. Với mỗi biến Y (Y1 là mưa quan trắc (r\_obs), Y2 là mưa tái phân tích (tp\_era), Y3 là mưa GPCP (pcp)) ta xây dựng được phương trình hồi quy đa biến cho ba khu vực Tây Bắc (B1), Việt Bắc - Đông Bắc (B2), Đồng bằng Bắc Bộ (B3). Như vậy sẽ có 9 phương trình hồi quy đa biến được xây dựng từ các yếu tố được chọn khác nhau.

Sau khi tiến hành xong các bước nêu trên ta được kết quả mỗi vùng có 3 phương trình hồi quy, tính toán lượng mưa với các yếu tố được lựa chọn theo từng loại số liệu mưa và từng vùng. Từ ba phương trình này, nghiên cứu sẽ lần lượt đưa ra lượng mưa tính toán từ các yếu tố tốc độ thẳng đứng, xoáy thế... lấy từ số liệu độc lập của năm 2009. Mỗi phương trình sẽ được áp dụng để tính toán mưa tại tất cả các điểm lưới thuộc từng khu vực. Tính toán mưa với các yếu tố từ phương trình sẽ thu được kết quả lượng mưa tại mỗi điểm lưới, với những lượng mưa nhỏ hơn 0 sẽ được gán bằng không. Tiếp theo, tiến hành đánh giá các sai số và tương quan của lượng mưa tính toán từ mô hình hồi quy lần lượt với lượng

mưa tái phân tích và lượng mưa VnGP [6], kết quả được trình bày trong phần sau. Với 3 loại số liệu mưa: mưa tái phân tích ERA, mưa quan trắc, mưa vệ tinh GPCP; ta sẽ xây dựng được ba mô hình hồi quy với ba loại số liệu mưa cho 3 vùng. Trong phần tiếp theo, mục 3.3 đến 3.5 sẽ trình bày kết quả đánh giá sai số và tương quan với số liệu mưa tái phân tích và số liệu mưa VnGP của từng mô hình với ba loại số liệu mưa kể trên: Mục 3.3 là đối với mưa tái phân tích Era, mục 3.4 là đối với mưa trên lưới VnGP.

**3.3 Đánh giá kết quả mô hình hồi quy có yếu tố dự báo Y là mưa tái phân tích ERA**

Hình 1 mô tả sai số trung bình ME, sai số tuyệt đối trung bình MAE, sai số bình phương trung bình RMSE và tương quan CORR được tính trên bộ số liệu độc lập năm 2009 tại mỗi điểm lưới. Ở đây ba vùng B1, B2, B3 sẽ được áp dụng ba phương trình khác nhau để đưa ra tính toán mưa tại điểm lưới thuộc khu vực, từ đó tính được sai số và tương quan so với mưa tái phân tích ERA. Các giá trị sai số, tương quan trung bình theo thời gian và tính trung bình cho khu vực cũng được thể hiện trên hình với ký hiệu

[“Tên chỉ số”\_ “vùng” = “giá trị”], với “Tên chỉ số” là: ME, MAE, RMSE, CORR; “vùng” gồm: B1 (Tây Bắc), B2 (Việt Bắc- Đông Bắc), B3 (Đồng bằng Bắc Bộ).

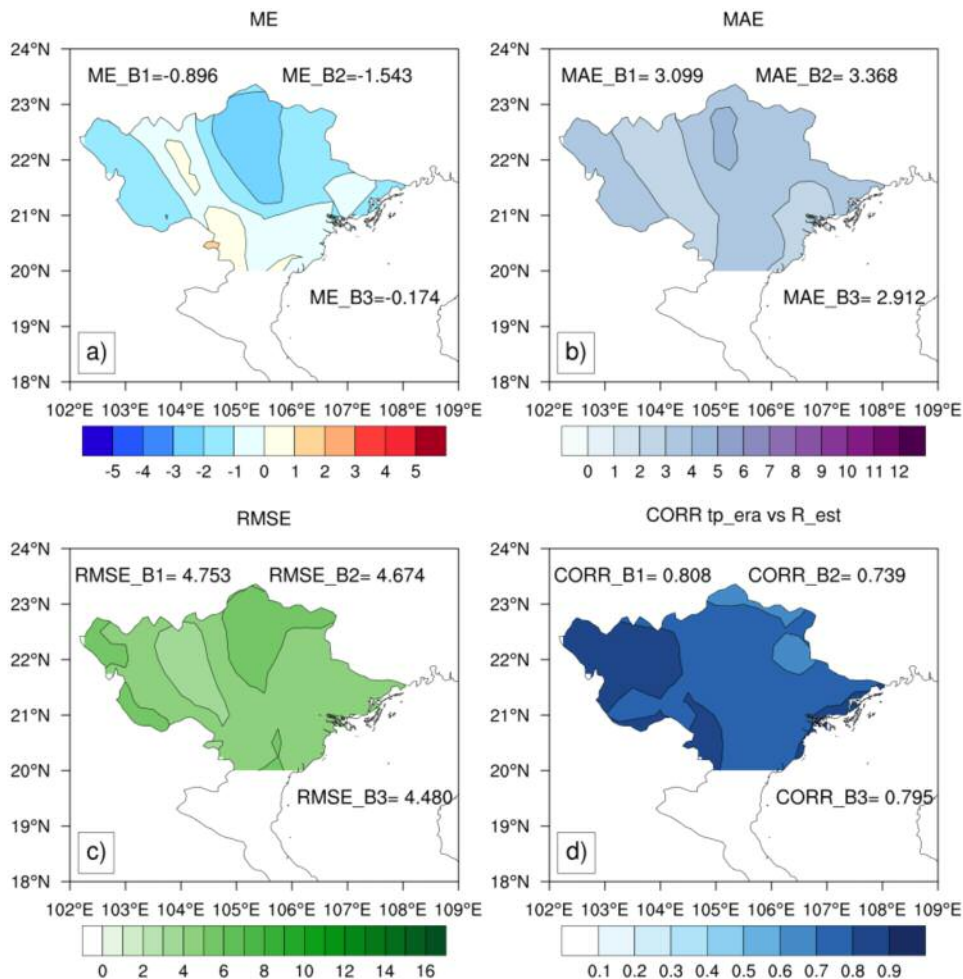
Trên Hình 1a có thể thấy sai số ME chủ yếu mang giá trị âm và có giá trị tuyệt đối nhỏ phổ biến từ 0 đến 2 mm, sai số trung bình vùng có giá trị tuyệt đối nhỏ nhất là ở khu vực B3, lớn nhất là ở khu vực B2. Có thể nhận xét rằng giá trị lượng mưa hiệu chỉnh sử dụng phương trình hồi quy thường có lượng mưa nhỏ hơn so với lượng mưa quan trắc trong bộ số liệu VnGP.

So với sai số ME, sai số MAE (Hình 1b) là sai số tuyệt đối trung bình phổ biến từ 2 đến 4 mm, giá trị MAE trung bình vùng B3 vẫn là nhỏ nhất (2.912).

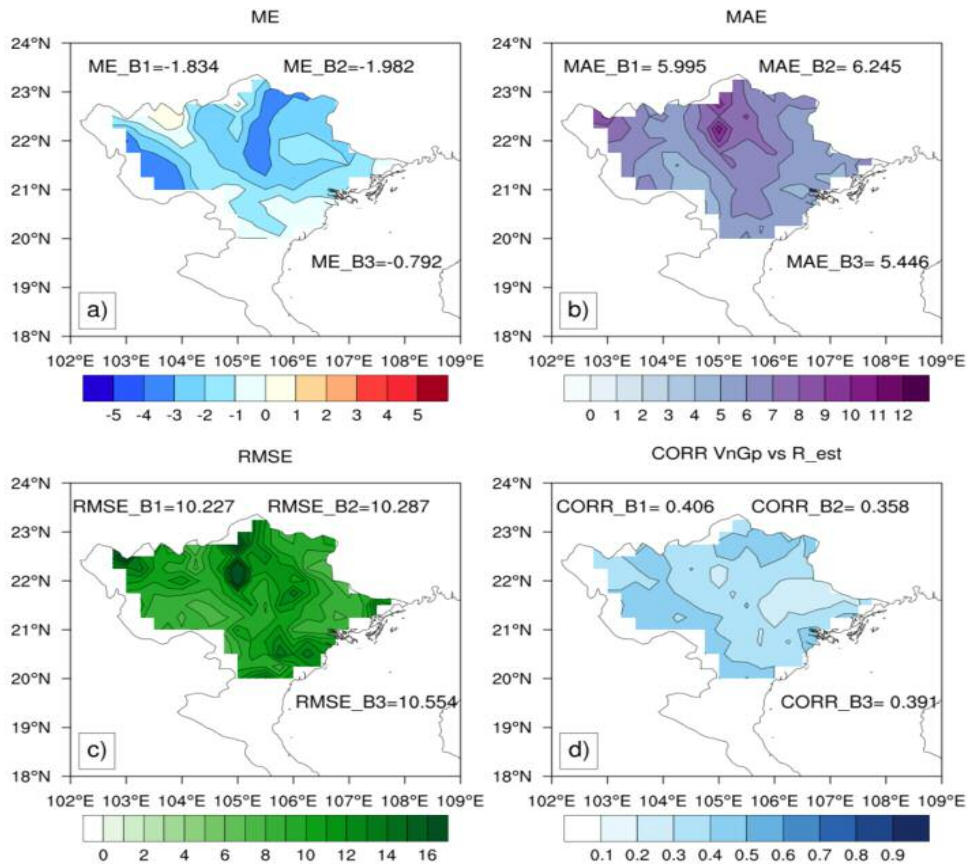
Giá trị sai số RMSE (Hình 1c) phân bố từ 3 đến 7. Giá trị RMSE lớn nhất ở khu vực B1 và nhỏ nhất ở khu vực B3. Tương quan (năm 2009) (Hình 1d) của số liệu mưa tính toán từ mô hình

thống kê với mưa tái phân tích phổ biến từ 0,7 - 0,9, giá trị tương quan tốt nhất ở vùng B1.

Hình 2 tương tự như Hình 1 nhưng là so sánh mưa tính toán từ mô hình thống kê với mưa VnGP. Hình 2a có giá trị ME đều là âm giá trị từ 0 đến -4 mm, vùng B1 và B2 có giá trị ME trung bình vùng tuyệt đối đều có trị số xấp xỉ 2 mm, trong khi đó ở khu vực B3 là 0.79 mm. Hình 2b cho thấy giá trị MAE dao động trong khoảng từ 4 - 11 mm, giá trị trung bình ba vùng B1, B2, B3 lần lượt là 5,9; 6,2 và 5,4 mm. Hình 2c giá trị RMSE từ khoảng 10 đến 16. Giá trị RMSE trung bình vùng tương đương nhau và đều trong khoảng 10 - 10,5. Hình 2d giá trị tương quan từ 0,2 đến 0,6. Tương quan trung bình vùng dao động quanh giá trị 0,4. Có thể thấy được một vùng có sai số lớn và tương quan rất kém ở khu vực gần 22N, 105E.



Hình 1. Sai số (a) ME, (b) MAE, (c) RMSE và (d) tương quan của mưa tính toán từ mô hình có yếu tố dự báo là  $tp\_era$ , so sánh với mưa tái phân tích  $tp\_era$  năm 2009



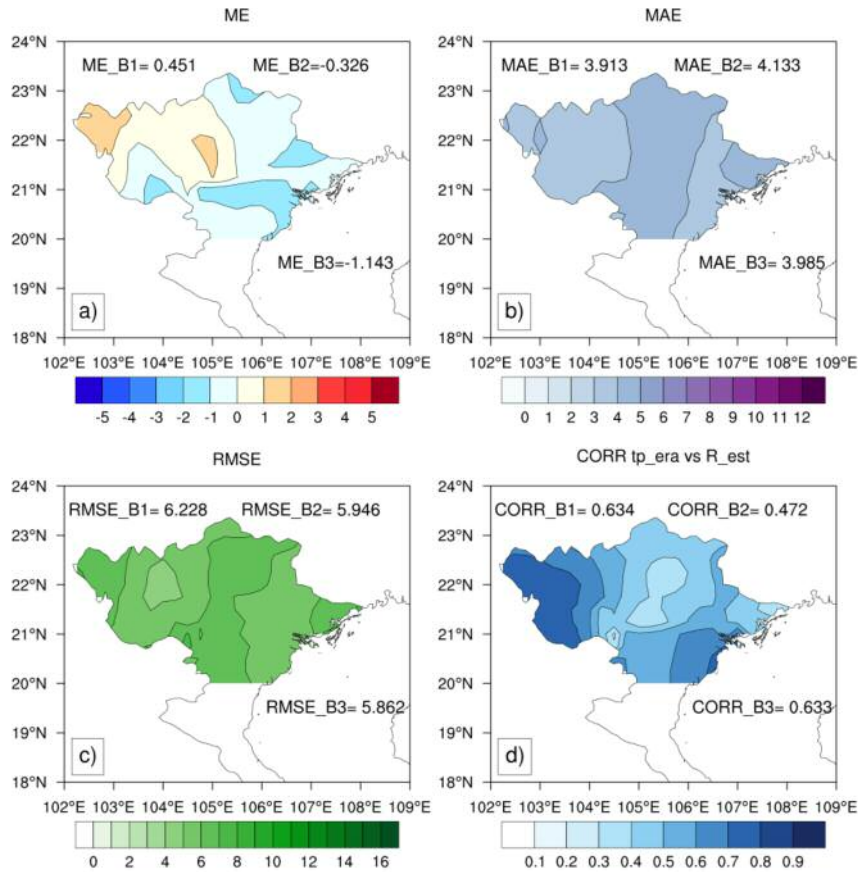
Hình 2. Sai số (a) ME, (b) MAE, (c) RMSE và (d) tương quan của mưa tính toán từ mô hình có yếu tố dự báo là  $tp\_era$ , so sánh với mưa VnGP năm 2009

**3.4 Đánh giá kết quả mô hình hồi quy có yếu tố dự báo Y là mưa vệ tinh (GPCP)**

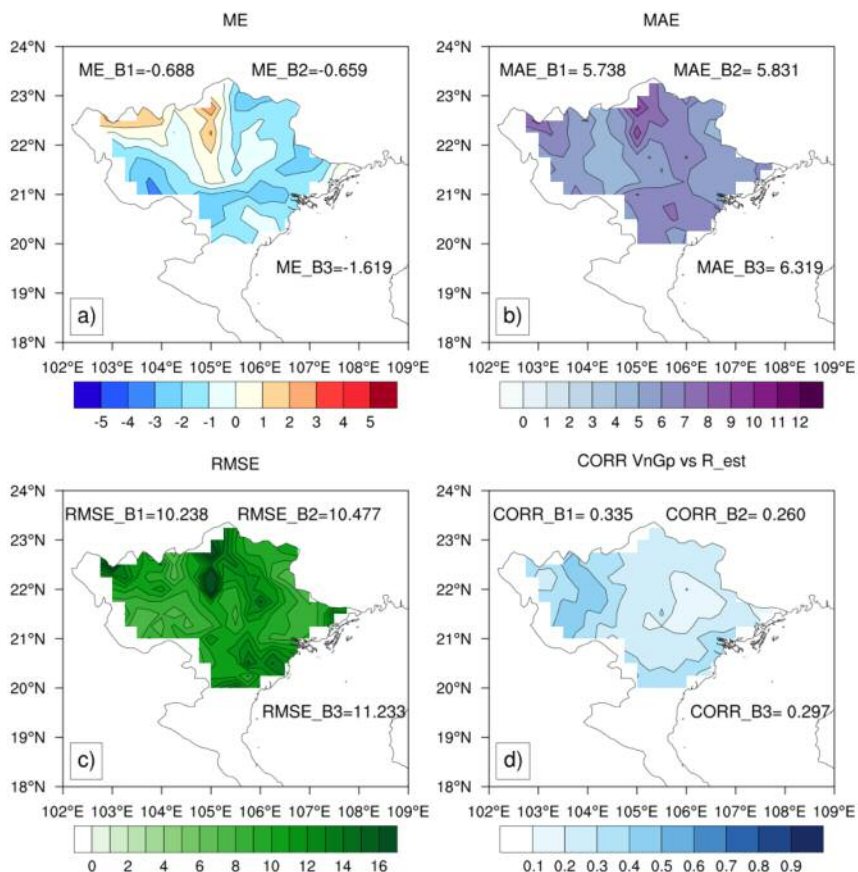
Hình 3 tương tự hình 1, thể hiện sai số và tương quan của mưa tính toán từ phương trình hồi quy sử dụng yếu tố dự báo là mưa GPCP, hình 3a thể hiện phân bố giá trị ME, vùng B3 và vùng B2 giá trị ME chủ yếu mang giá trị âm, trong khi ở vùng B1 giá trị này dương nhiều hơn. Ở hình 3b, MAE ở khoảng 3 - 5 mm, MAE trung bình ở ba vùng ở khoảng 4 mm. Giá trị RMSE (Hình 3c) trung bình khu vực B1, B2, B3 là 6,22; 5,94; 5,86 mm. Tương quan thể hiện ở hình 3d cho thấy tương quan trung bình vùng khoảng 0,4 - 0,6 kém hơn so với tương quan trên hình 1.

Tiếp tục so sánh với lượng mưa VnGP, sai số ME (Hình 4a) chủ yếu mang giá trị âm, sai số trung bình ba vùng B1, B2, B3 lần lượt là -0,65; -0,68; -1,61 mm. Giá trị MAE (hình 4b) trung bình khoảng 6 mm. Trên hình 4c có thể thấy xuất hiện vùng có RMSE lớn ở khoảng 22N, 105E. Giá trị tương quan ở hình 4d lớn nhất ở vùng B1 là 0,33, vùng B2 là 0,26, vùng B3 là 0,29, nhìn chung những giá trị này kém hơn so với tương quan ở hình 2.

Các kết quả trên được so sánh với kết quả đánh giá trong công trình của Hoàng Phúc Lâm và cs [4] và được tổng hợp trong phần dưới đây.



Hình 3. Sai số (a) ME, (b) MAE, (c) RMSE và (d) tương quan của mưa tính toán từ mô hình có yếu tố dự báo là mưa GPCP (Pcp), so sánh với mưa tái phân tích (Tp\_Era) năm 2009



Hình 4. Sai số (a) ME, (b) MAE, (c) RMSE và (d) tương quan của mưa tính toán từ mô hình có yếu tố dự báo là mưa GPCP (Pcp), so sánh với mưa VnGP năm 2009

### 3.6 Tổng hợp kết quả đánh giá mô hình hồi quy

Bảng 3 và bảng 4 thể hiện các chỉ số đánh giá tính cho từng khu vực. Với ba mô hình hồi quy với yếu tố dự báo là mưa quan trắc ( $r_{obs}$ ), mưa tái phân tích ( $tp_{era}$ ) và mưa GPCP ( $pcp$ ). Kết quả đánh giá sai số của lượng mưa tính toán với mưa tái phân tích ERA (Bảng 3) cho thấy giá trị tương quan cao và giá trị sai số nhỏ hơn so các giá trị tương quan và giá trị sai số với lượng mưa VnGP (Bảng 4).

Tương quan khi đánh giá sử dụng số liệu mưa tái phân tích phổ biến khoảng 0,6 đến 0,8; trong khi tương quan với số liệu VnGP chỉ ở khoảng 0,3 đến 0,4. Giá trị sai số ME ở cả hai trường hợp đều thấp và chủ yếu mang giá trị âm. Giá trị sai

số MAE ở trong khoảng 3 đến 4 mm; ở Bảng 4 là khoảng 5 đến 6 mm. Với giá trị RMSE, đánh giá trên số liệu tái phân tích phổ biến ở khoảng 4,5 đến 6 mm, trên số liệu VnGP là khoảng 10 đến 11 mm. Điều này có thể được lý giải là do lượng mưa tái phân tích ERA là cùng bộ số liệu với các đặc trưng hoàn lưu và nhiệt động lực (vì đều trích xuất từ số liệu tái phân tích ERA) nên các đặc trưng này đã chứa đựng những mối quan hệ động lực với nhau. Số liệu VnGP là số liệu tính toán từ số liệu quan trắc, thông qua phương pháp nội suy Spheremap đưa về các điểm lưới. Như vậy, để khách quan thì khi xem xét ứng dụng các sản phẩm hiệu chỉnh, ta nên sử dụng các sai số và tương quan khi so với số liệu mưa VnGP.

Bảng 3. Tổng hợp kết quả tương quan và sai số của mưa tính toán với mưa tái phân tích ERA

STT	Y	r_obs			tp_era			pcp		
	Khu vực	B1	B2	B3	B1	B2	B3	B1	B2	B3
	Chỉ số									
1	CORR	0.73	0.61	0.78	0.81	0.74	0.8	0.63	0.47	0.63
2	ME	0.85	-2.86	-0.02	-0.9	-1.54	-0.17	0.45	-0.33	-1.14
3	MAE	3.43	4.58	3.01	3.09	3.37	2.91	3.91	4.13	3.99
4	RMSE	5.69	6.14	4.59	4.75	4.67	4.48	6.23	5.95	5.86

Bảng 4. Tổng hợp kết quả tương quan và sai số của mưa tính toán với mưa VnGP

STT	Y	r_obs			tp_era			pcp		
	Khu vực	B1	B2	B3	B1	B2	B3	B1	B2	B3
	Chỉ số									
1	CORR	0.41	0.31	0.4	0.41	0.36	0.39	0.34	0.26	0.29
2	ME	-0.25	-3.37	-0.63	-1.83	-1.98	-0.79	-0.69	-0.66	-1.62
3	MAE	5.25	7.38	5.37	5.99	6.25	5.45	5.74	5.83	6.32
4	RMSE	9.75	11.05	10.49	10.23	10.29	10.55	10.24	10.48	11.23



### 3. Kết luận và kiến nghị

Nghiên cứu đã đánh giá được giá trị tương quan giữa các loại số liệu lượng mưa: mưa quan trắc, mưa tái phân tích ERA và mưa vệ tinh GPCP. Kết quả cho thấy lượng mưa tái phân tích ERA tương quan tốt hơn với lượng mưa quan trắc so với số liệu mưa vệ tinh GPCP, nhưng mức độ tương quan còn nhỏ. Cụ thể, giá trị tương quan của mưa quan trắc với mưa tái phân tích ở khoảng 0,6, trong khi giá trị tương quan giữa mưa quan trắc với mưa tính toán vệ tinh GPCP ở khoảng 0,5.

Nghiên cứu đã xây dựng được 9 phương trình hồi quy tuyến tính đa biến cho 3 loại số liệu mưa và cho 3 khu vực Tây Bắc (B1), Việt Bắc - Đông Bắc (B2) và Đồng bằng Bắc Bộ. Với mỗi phương trình bao gồm các biến khác nhau được lựa chọn thông qua phương pháp hồi quy từng bước. Thử nghiệm áp dụng 9 phương trình với số liệu độc lập năm 2009 và so sánh lượng mưa tính toán với số liệu mưa tái phân tích ERA cho sai số

tương đối nhỏ, giá trị tương quan cao, có điểm lưới đạt giá trị tương quan trên 0,9. Tuy nhiên, khi đánh giá với số liệu lượng mưa VnGP, sai số đều tăng và giá trị tương quan giảm.

Có thể nhận thấy nhược điểm của phương pháp là cho mưa tính toán nhỏ hơn thực tế khi xảy ra mưa lớn. Trong khi nếu thực tế mưa nhỏ thì mưa tính toán từ phương trình cho giá trị lớn hơn so với thực tế. Tiêu biểu là sự khác biệt giữa sai số ME chủ yếu là âm vào những tháng mùa đông, và có giá trị ME dương lớn vào mùa hè. Do vậy, có thể để khắc phục được vấn đề trên định hướng tiếp theo của nhóm nghiên cứu là sẽ xây dựng hồi quy cho từng mùa khác nhau. Tiếp tục thử nghiệm áp dụng phương trình hồi quy đã xây dựng vào số liệu dự báo của mô hình và đánh giá sai số. Từ những sai số này có thể dùng phương pháp trung bình có trọng số hoặc xây dựng phương trình hiệu chỉnh cho từng mùa để đưa ra hiệu chỉnh dự báo mưa tốt hơn trong tương lai.

**Lời cảm ơn:** Nhóm nghiên cứu xin chân thành cảm ơn sự hỗ trợ của các đề tài nghiên cứu khoa học KC.08.06/16-20 và TNMT.2017.05.01 trong việc thực hiện và công bố nghiên cứu này.

### Tài liệu tham khảo

1. Berry, Reeder, Jakob (2012), *Coherent Synoptic Disturbances in the Australian Monsoon*;
2. Dallavalle, J.P. and Glahn, B. (2005), *Toward a gridded MOS system. Preprints, 21th Conference on Weather Analysis and Forecasting, Washington, DC, Amer. Meteor. Soc., 13B.2.*
3. John, W. Nielsen-Gammon and Gold, D.A. (2008), *Potential Vorticity Diagnosis of the Severe Convective Regime.*
4. Du, J. and Cho, H.R. (1996), *Potential Vorticity Anomaly and Mesoscale Convective Systems on the Baiu (Mei-Yu) Front.*
5. Hoang Phuc Lam, Nguyen Duc Nam, Do Thi Thanh Thuy, Hoang Duc Cuong (2019), *Develop a statistical model to correct the model rainfall forecast for Bac Bo using coherent potential vorticity*, Vietnam Journal of Hydrometeorology, 697, 26-32.
6. Hoang, P.L., Reeder, M.J., Berry, G.J. and Schwendike, J. (2017), *Coherent Potential Vorticity Maxima and Their Relationship to Extreme Summer Rainfall in the Australian and North African Tropics*, Journal of Southern Hemisphere Earth System Sciences, 66, 424-441.
7. Nguyen-Xuan, T., Ngo-Duc, T., Kamimera, H., Trinh-Tuan, L., Matsumoto, J., Inoue, T. and Phan-Van, T. (2016), *The Vietnam gridded precipitation (VnGP) dataset: Construction and validation*. Scientific Online Letters on the Atmosphere (SOLA)
8. Rolf, F., Hertenstein, A. and Wayne Schubert, H. (1991), *Potential Vorticity Anomalies Asso-*

ciated with Squall Lines.

9. Võ Văn Hòa, Nguyễn Thị Hằng và Mai, N.C. (2007), *Nghiên cứu ứng dụng lọc Kalman hiệu chỉnh một số trường dự báo bề mặt từ mô hình HRM*. Tạp chí Khí tượng Thủy văn, 557, 49-57.

## VERIFICATION AND COMPARISON OF RAINFALL FORECAST STATISTICAL MODELS USING ATMOSPHERIC CIRCULATION AND THERMODYNAMIC INDICES

Hoang Phuc Lam<sup>1</sup>, Nguyen Duc Nam<sup>2</sup>, Le Minh Duc<sup>3</sup>

<sup>1</sup>National Meteorological Forecasting Center, General Department of Hydrometeorology

<sup>2</sup>Institute of Geophysics, Vietnam Academy of Science and Technology

<sup>3</sup>Hanoi University of Natural Resources and Environment

**Abstract:** In this paper, the research team verify and compare rainfall forecast skills of different output statistical models using different predictors (independent variable): observed ERA re-analysis and GPCP rainfall. The research built 9 stepwise regression models predicting rainfall for the 3 different areas in Bac Bo region using 3 different predictors. The independent evaluation results show that the model used ERA rainfall as predictor has smallest RMSE errors, highly correlated of 0.6 to 0.8 with exceptional value of more than 0.9.

**Keywords:** Potential vorticity, quantitative precipitation forecast, MOS.