

# HIỆU CHÍNH DỰ BÁO MƯA LỚN TỪ MÔ HÌNH S2S TRÊN KHU VỰC HÀ NỘI BẰNG PHƯƠNG PHÁP HỒI QUY PHÂN VỊ

Phạm Thị Minh<sup>(1)</sup>, Lê Thị Tú Như<sup>(1)</sup>, Trịnh Minh Ngọc<sup>(2)</sup>, Nguyễn Thị Phương Chi<sup>(1)</sup>,  
Phan Vũ Hoàng Phương<sup>(1)</sup>, Huỳnh Thị Minh Sương<sup>(1)</sup>

<sup>(1)</sup>Trường Đại học Tài nguyên và Môi trường TP Hồ Chí Minh

<sup>(2)</sup>Trường Đại học Khoa học tự nhiên, Đại học Quốc gia Hà Nội

Ngày nhận bài: 17/10/2024; ngày chuyển phản biện: 18/10/2024; ngày chấp nhận đăng: 8/11/2024

**Tóm tắt:** Bài báo này trình bày kết quả ứng dụng phương pháp hồi quy phân vị dự báo mưa lớn hạn từ 3 đến 10 ngày trên khu vực Hà Nội. Nghiên cứu sử dụng chuỗi số liệu mưa ngày quan trắc tại trạm Láng và số liệu mưa ngày dự báo từ mô hình S2S (the sub-seasonal to seasonal forecast model) của tháng 6, tháng 7 và tháng 8 trong 20 năm (2001-2020) để xây dựng phương trình dự báo bằng phương pháp hồi quy tuyến tính (OLS) và hồi quy phân vị, trong đó số liệu mưa ngày quan trắc là số liệu phụ thuộc, số liệu mưa ngày dự báo từ mô hình S2S là số liệu độc lập. Kết quả kiểm định các phương trình dự báo cho thấy phương trình OLS, q75 và q90 thỏa mãn mức tin cậy thống kê và có thể sử dụng để dự báo lượng mưa ở hạn dự báo 3 ngày, 5 ngày, 7 ngày và 10 ngày tại trạm Láng. Ngoài ra, nghiên cứu thử nghiệm dự báo 5 đợt mưa lớn tại khu vực Hà Nội bằng các phương trình hồi quy tuyến tính và hồi quy phân vị q75 và q90 ở các hạn dự báo 3 ngày, 5 ngày, 7 ngày và 10 ngày. Kết quả đánh giá sai số dự báo lượng mưa cho thấy, phương trình ở phân vị 90 cải thiện đáng kể sai số dự báo lượng mưa so với mô hình S2S ở tất cả các hạn dự báo. Trong đó sai số tương đối của phương trình hồi quy phân vị 90 cải thiện sai số dự báo lượng mưa từ 20 đến 29% so với kết quả dự báo lượng mưa từ mô hình S2S ở hạn dự báo từ 3 ngày đến 10 ngày.

**Từ khóa:** Hồi quy phân vị, hồi quy tuyến tính, mưa lớn, dự báo mưa.

## 1. Mở đầu

Mưa lớn xuất hiện có thể gây lũ lụt, sạt lở đất... Hậu quả của mưa lớn còn kéo dài sau đó nhiều tháng, bao gồm các vấn đề về nước sinh hoạt, dịch bệnh, môi trường, ... Ngược lại, thiếu hụt lượng mưa kéo dài có thể gây nên hạn hán. Nghiên cứu trên thế giới chỉ ra các trận mưa cực đoan có thể gây ảnh hưởng nghiêm trọng đến đời sống của con người [1], bởi lượng mưa lớn tập trung trong một khoảng thời gian ngắn (lượng mưa >100 mm/ngày). Trong khi đó lượng mưa ngày lớn nhất năm có xu hướng tăng trên hầu hết các khu vực trên toàn cầu (2/3 số trạm) [2].

Việt Nam nằm trong khu vực nhiệt đới gió mùa và mưa lớn thường xuyên xuất hiện trên

nhiều vùng, miền khác nhau. Mưa lớn gây ra những thiệt hại to lớn không chỉ về cơ sở vật chất mà còn thiệt hại về người. Điển hình như đợt mưa lớn kéo dài sau cơn bão số 3 năm 2024 đã kéo theo các hậu quả về sạt lở và lũ lụt xảy ra ở một số tỉnh phía Bắc Việt Nam. Do đó việc nghiên cứu và dự báo mưa lớn là cần thiết để giảm thiệt hại do thiên tai bão lũ.

Cuối những năm 1990 và đầu những năm 2000, tại Việt Nam các nghiên cứu sự xuất hiện mưa lớn nhờ việc phân tích số lượng lớn các hình thể synop liên quan đến sự xuất hiện mưa sau đó đúc kết thành quy luật [3], [4]. Sau những năm 2002, các nghiên cứu tập trung nghiên cứu và khai thác các sản phẩm dự báo của mô hình số [5-16]. Trong đó một số nghiên cứu dự báo tổ hợp bằng cách sử dụng nhiều mô hình động lực, tăng độ phân giải của mô hình, đồng hóa số liệu, và phương pháp thống kê để tăng độ chính xác

Liên hệ tác giả: Phạm Thị Minh  
Email: ptminh@hcmunre.edu.vn

trong mô phỏng và dự báo mưa lớn. Kết quả các nghiên cứu cho thấy cải thiện đáng kể khả năng dự báo mưa trong hạn 3 ngày.

Tuy nhiên, có rất nhiều yếu tố khiến cho khả năng dự báo của mô hình số giảm rất nhanh sau 3 ngày. Các yếu tố này bao gồm: Sai số trong trường số liệu đầu vào, phép đơn giản hóa tính toán trong mô hình, sự mô phỏng chưa chính xác của các quá trình khí quyển bởi mô hình... Đặc biệt đối với một quốc gia nằm trong khu vực nhiệt đới gió mùa và có địa hình phức tạp như Việt Nam, mưa lớn thường gây ra bởi sự tương tác của các quá trình ở nhiều quy mô khác nhau. Các nghiên cứu cho thấy, việc tăng độ phân giải từ 100 km xuống 5 km có thể cho sự cải thiện rất lớn trong việc dự báo mưa, tuy nhiên khi tăng độ phân giải dưới 5 km, độ chính xác của mô hình tăng lên không đáng kể. Việc tăng độ phân giải cũng làm tăng tài nguyên và thời gian tính toán, do đó, cũng có nhiều hạn chế trong việc áp dụng dự báo nghiệp vụ. Trong khi đó, sản phẩm dự báo của trung tâm hạn vừa và hạn dài Châu Âu cung cấp sản phẩm dự báo với độ phân giải là 9 km đến 13 km. Nếu tận dụng được nguồn số liệu này dự báo cho địa phương thì khá hữu ích và giảm chi phí rất nhiều.

Trên thế giới, một số tác giả đã sử dụng mô hình hồi quy tuyến tính và mô hình máy học dự báo nhiệt độ cực trị và nhiệt độ [17-21]. Hầu hết các nghiên cứu này đều chỉ ra rằng kỹ năng dự báo nhiệt độ  $T_x$  thấp và do đó không thể phù hợp cho dự báo nắng nóng. Đối với dự báo mưa lớn, phương pháp thống kê phổ biến được sử dụng là dự báo mưa phân vị áp dụng kết hợp giữa đầu ra của mô hình động lực và phương pháp thống kê để tạo ra các dự báo mưa hữu ích [22], [23]. Các phương pháp này được phát triển dựa trên cơ sở dự báo hình thể synop tốt hơn so với dự báo mưa. Các hình thể hoàn lưu quy mô synop như là nhân tố dự báo cho dự báo mưa trong dự báo mưa bằng hồi quy phân vị. Thuật toán khởi điểm cho phương pháp hồi quy phân vị như phương pháp tương tự [24], phương pháp prog hoàn hảo [25] và kỹ thuật dấu vân tay [26], hồi quy phân vị rừng [27] đã được ứng dụng khá hiệu quả trong dự báo các hiện tượng cực đoan.

Do đó, nghiên cứu ứng dụng phương pháp

hồi quy phân vị hiệu chỉnh dự báo mưa lớn cho khu vực Hà Nội hạn từ 3 đến 10 ngày, trong đó số liệu mưa quan trắc tại trạm là biến phụ thuộc và mưa tích lũy 24 giờ từ mô hình (s2s) được lấy từ sản phẩm dự báo của trung tâm hạn vừa và hạn dài Châu Âu là số liệu độc lập.

## 2. Số liệu và phương pháp nghiên cứu

### 2.1. Số liệu

Bộ số liệu để xây dựng các phương trình dự báo bao gồm: Số liệu mưa ngày quan trắc tại trạm Láng (đại diện cho khu vực Hà Nội) trong các tháng 6, 7 và 8 từ 2001-2020; số liệu mưa tích lũy 24 giờ dự báo hạn 3 ngày, 5 ngày, 7 ngày và 10 ngày tại trạm Láng từ mô hình S2S được tải về từ trang web <https://apps.ecmwf.int/datasets/data/s2s-forecasts-daily-averaged-ecmf/levtype=sfc/type=cf/>. Trong đó, số liệu dự báo của ECMWF bao gồm 12 thành phần (member dự báo) trên khắp thế giới đóng góp có sẵn trong cơ sở dữ liệu. Số liệu này được tạo nên hai lần mỗi tuần và mỗi sản phẩm dự báo có độ dài (hạn dự báo) là 46 ngày. Nghiên cứu sử dụng dự báo của ECMWF cho khu vực Bắc Bộ với độ phân giải  $0,125^\circ \times 0,125^\circ$  với 241 điểm lưới.

Nghiên cứu sử dụng số liệu mưa ngày quan trắc tại trạm Láng thuộc khu vực Hà Nội với thông tin vị trí, giai đoạn lấy số liệu được liệt kê trong Bảng 1 và minh họa trong Hình 1. Ngoài ra, nghiên cứu thử nghiệm dự báo lượng mưa trong 5 đợt mưa lớn của các năm 2004, 2006, 2011, 2012 và 2013, số liệu mưa tích lũy 24 giờ (mưa ngày) hạn 3 ngày, 5 ngày, 7 ngày và 10 ngày dự báo từ mô hình S2S liệt kê trong Bảng 2.

### 2.2. Phương pháp hồi quy phân vị

Phương pháp hồi quy phân vị được Koenker & Bassett giới thiệu lần đầu tiên năm 1978 [28]. Phương pháp hồi quy phân vị là một sự mở rộng của hồi quy tuyến tính (OLS). Thay vì ước lượng các tham số của hàm hồi quy trên giá trị trung bình trong phương pháp OLS, phương pháp hồi quy phân vị ước lượng các tham số hồi quy trên từng phân vị của biến phụ thuộc để sao cho tổng chênh lệch tuyệt đối của hàm hồi quy tại phân vị của biến phụ thuộc là nhỏ nhất. Nghĩa là, thay vì xác định tác động biên của biến độc lập đến giá trị trung bình của biến phụ thuộc, hồi quy phân

vị sẽ giúp xác định tác động biên của biến độc lập đến biến phụ thuộc trên từng mức phân vị của biến phụ thuộc đó. Chi tiết về phương pháp hồi quy phân vị có thể tìm trong [29].

### 2.3. Phương pháp đánh giá

Để đánh giá kết quả thử nghiệm trong nghiên cứu này, chúng tôi sử dụng các chỉ số đánh giá thống kê như: Sai số trung bình (ME), sai số tuyệt đối trung bình (MAE), sai số quân phương (RMSE) (Wilks, 1997) [30] và sai số tương đối.

Sai số trung bình ME:

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

Trong đó:  $n$  tổng số dự báo;  $F_i$  là giá trị dự báo thứ  $i$ ;  $O_i$  là giá trị quan trắc thứ  $i$ .

Sai số tuyệt đối trung bình (MAE):

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

Trong đó:  $n$  tổng số dự báo;  $F_i$  là giá trị dự báo thứ  $i$ ;  $O_i$  là giá trị quan trắc thứ  $i$ .

Sai số quân phương (RMSE):

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

Trong đó:  $n$  tổng số dự báo;  $F_i$  là giá trị dự báo thứ  $i$ ;  $O_i$  là giá trị quan trắc thứ  $i$ .

Sai số quân phương dùng để biểu thị độ lớn trung bình của sai số. Cũng giống như MAE, RMSE không phản ánh xu hướng lệch giữa giá trị dự báo và giá trị quan trắc. Giá trị RMSE tối ưu là  $RMSE = 0$  tức là khi giá trị dự báo bằng giá trị quan trắc tại mọi điểm trong không gian đánh giá (Wilks, 1997).

Sai số tương đối

$$SSTD = \left| \frac{fcs - OBS}{OBS} \right| \times 100\% \quad (4)$$

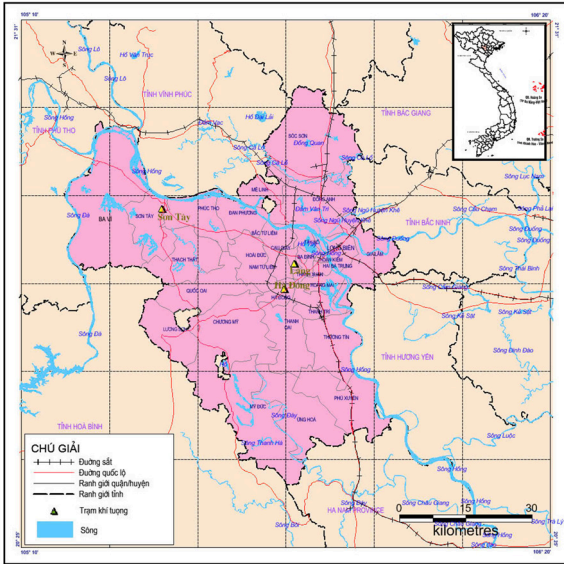
Trong đó:  $SSTD$  là sai số tương đối;  $fcs$  là lượng mưa ngày dự báo;  $OBS$  là lượng mưa ngày quan trắc.

Bảng 1. Danh sách các trạm quan trắc khí tượng tại khu vực Hà Nội

TT	Tên trạm	Tọa độ		Thời gian quan trắc
		Kinh độ	Vĩ độ	
1	Láng	105,8000°E	21,0167°N	2001-2020

Bảng 2. Lượng mưa 24 giờ quan trắc và dự báo hạn 3 ngày, 5 ngày, 7 ngày và 10 ngày tại trạm Láng trong 5 đợt mưa (Đơn vị: mm/ngày)

STT	Ngày	Hạn 3 ngày	Hạn 5 ngày	Hạn 7 ngày	Hạn 10 ngày	Lượng mưa 24 giờ quan trắc (OBS)
1	21/7/2004	19,0	7,0	21,0	3,0	83,0
2	22/7/2004	22,0	23,0	24,0	9,0	82,0
3	23/7/2004	16,0	24,0	11,0	16,0	69,0
4	17/8/2006	20,0	54,0	2,0	2,0	53,7
5	18/8/2006	11,0	10,0	25,0	11,0	99,3
6	24/6/2011	21,0	11,0	7,0	6,0	65,8
7	25/6/2011	149,0	18,0	18,0	23,0	65,4
8	17/8/2012	3,0	6,0	14,0	4,0	90,8
9	18/8/2012	5,0	5,0	3,0	2,0	126,7
10	19/8/2012	17,0	2,0	8,0	10,0	53,0
11	8/8/2013	8,0	3,0	2,0	3,0	171,2
12	9/8/2013	21,0	25,0	4,0	5,0	62,3



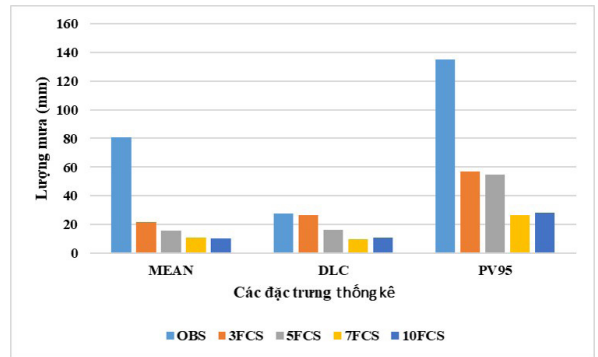
Hình 1. Vị trí các trạm khí tượng trên khu vực Hà Nội

### 3. Kết quả và thảo luận

#### 3.1. Đánh giá sai số dự báo lượng mưa > 50 mm/ngày của mô hình S2S

Trong phần này nghiên cứu sử dụng số liệu mưa ngày  $\geq 50$  mm/ngày quan trắc trong các đợt mưa lớn tại trạm Láng (OBS) trong tháng 6, 7 và 8 từ năm 2001 đến năm 2020 (20 năm), với tổng số ngày mưa lớn khoảng 89 ngày. Ngoài ra, số liệu dự báo của mô hình S2S được thu thập cho các hạn 3 ngày, 5 ngày, 7 ngày và 10 ngày (3FCS, 5FCS, 7FCS và 10FCS) cùng thời điểm với số liệu mưa quan trắc ở trên. Nghiên cứu tính toán một số đặc trưng thống kê của 5 chuỗi số liệu trên (Hình 2) và các chỉ số sai số dự báo lượng mưa bao gồm sai số trung bình (ME), sai số tuyệt đối trung bình (MAE) và sai số quân phương (RMSE) (Hình 3).

Từ Hình 2 cho thấy, số liệu quan trắc có giá trị trung bình (MEAN) khoảng 80 mm, trong khi đó giá trị MEAN dự báo từ S2S hạn 3 ngày (3FCS), 5 ngày (5FCS), 7 ngày (7FCS) và 10 ngày (10FCS) dưới 20 mm. Trong khi đó, độ lệch chuẩn (DLC) phản ánh sự dao động số liệu xung quanh giá trị trung bình thì OBS và 3FCS có sự dao động mạnh quanh giá trị trung bình, tiếp theo là đến 5FCS, còn 7FCS và 10FCS ít biến động hơn. Phân vị 95 (PV95) là giá trị số liệu gần cuối chuỗi số liệu đã

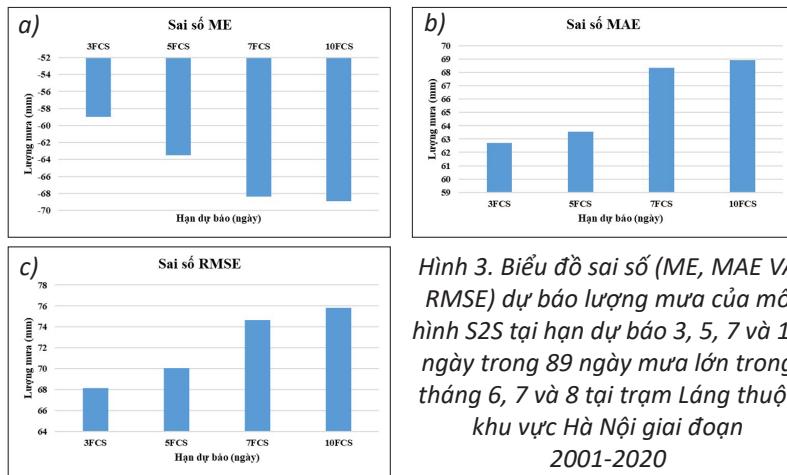


Hình 2. Biểu diễn trung bình, độ lệch chuẩn và phân vị 95 của chuỗi số liệu lượng mưa quan trắc và mô hình S2S trong 89 ngày mưa lớn (tháng 6, 7 và 8) xảy ra tại trạm Láng trên khu vực Hà Nội giai đoạn 2001-2020

sắp xếp theo thứ tự, OBS có giá trị gần 140 mm, 3FCS và 5FCS có PV95 gần 58 mm, còn 7FCS và 10FCS có PV95 dưới 30 mm. Gần như số liệu dự báo bởi mô hình S2S không nắm bắt tốt các đặc trưng thống kê của chuỗi số liệu mưa tại trạm Láng. Để đánh giá sai số dự báo lượng mưa của mô hình S2S, nghiên cứu tính các chỉ số sai số ME, MAE và RMSE (Hình 3) trong chuỗi số liệu 89 ngày mưa lớn của tháng 6, 7 và 8 trong 20 năm (2001-2020).

Từ Hình 3a, cho thấy giá trị ME lần lượt ở các hạn dự báo 3 ngày, 5 ngày, 7 ngày và 10 ngày đều nhỏ hơn 0, hay mô hình S2S luôn dự báo lượng mưa thấp hơn so với quan trắc. Trong khi đó giá trị sai số dự báo lượng mưa (Hình 3b) dao động từ 62 mm đến 69 mm cho các hạn dự báo trên, với hạn dự báo càng tăng thì sai số dự báo lượng mưa cũng tăng lên. Biên độ dao động của sai số dự báo lượng mưa cũng tăng dần theo hạn dự báo từ 68 mm đến 78 mm ở hạn dự báo 10 ngày (Hình 3c).

Như vậy, kết quả dự báo mưa lớn của của mô hình S2S có sai số khá lớn khi áp dụng trực tiếp dự báo mưa lớn tại khu vực Hà Nội. Trong phần tiếp theo nghiên cứu xây dựng phương trình hồi quy tuyến tính và hồi quy phân vị để dự báo mưa lớn cho khu vực Hà Nội.



Hình 3. Biểu đồ sai số (ME, MAE VÀ RMSE) dự báo lượng mưa của mô hình S2S tại hạn dự báo 3, 5, 7 và 10 ngày trong 89 ngày mưa lớn trong tháng 6, 7 và 8 tại trạm Láng thuộc khu vực Hà Nội giai đoạn 2001-2020

### 3.2. Xây dựng phương trình

Trong phần này nghiên cứu sử dụng chuỗi số liệu mưa ngày (quan trắc - OBS, và mô hình S2S) của các tháng 6, 7 và 8 trong 20 năm (2001-2020) đã loại 5 đợt mưa làm số liệu dự báo kiểm tra (Bảng 2) để xây dựng phương trình hồi quy tuyến tính (OLS) và các phương trình hồi quy phân vị (q10, q25, q50, q75, q90) cho dự báo

lượng mưa hạn từ 3 đến 10 ngày tại trạm Láng thuộc khu vực Hà Nội. Kết quả dạng các phương trình dự báo và mức tin cậy thống kê được thể hiện trong các bảng từ Bảng 3 đến Bảng 6. Trong đó, đối với OLS chỉ cần giá trị P của hệ số góc nhỏ hơn 0,05 là phương trình có độ tin cậy thống kê (nhận) đối với các phương trình hồi quy phân vị thì giá trị P của hệ số góc cùng hệ số tự do đồng thời thỏa mãn đều nhỏ hơn 0,05 thì nhận.

Bảng 3. Phương trình hồi quy tuyến tính và hồi quy phân vị ứng với độ tin cậy tại trạm Hà Nội cho hạn 3 ngày

STT	Phương pháp	Phương trình	P - hệ số tự do	P - hệ số góc	Độ dài chuỗi	Ghi chú
1	OLS	$y = 0,4310x + 5,2022$	0,00	0,00	1,828	Nhận
2	Median	$y = 0,2043x - 28,7734$	0,00	1,00		Không nhận
3	q10	$y = -4,2300x - 16,5357$	1,00	0,00		Không nhận
4	q25	$y = 0,0071x - 0,0143$	0,027	0,816		Không nhận
5	q50	$y = 0,2043x - 28,7734$	0,00	1,00		Không nhận
6	q75	$y = 1,014x + 0,686$	0,00	0,00		Nhận
7	q90	$y = 1,4571x + 16,2143$	0,00	0,00		Nhận

Bảng 4. Phương trình hồi quy tuyến tính và hồi quy phân vị ứng với độ tin cậy tại trạm Hà Nội cho hạn 5 ngày

STT	Phương pháp	Phương trình	P - hệ số tự do	P - hệ số góc	Độ dài chuỗi	Ghi chú
1	OLS	$y = 0,4020x + 5,4731$	0,00	0,00	1,828	Nhận
2	Median	$y = 0,1483x - 20,9475$	0,00	1,00		Không nhận
3	q10	$y = 5,3250x - 30,0888$	1,00	0,00		Không nhận
4	q25	$y = -0,0986x - 17,6696$	1,00	0,00		Không nhận
5	q50	$y = 0,1483x - 20,9475$	0,00	1,00		Không nhận
6	q75	$y = 0,8895x + 1,7421$	0,00	0,00		Nhận
7	q90	$y = 1,2x + 17,6$	0,00	0,00		Nhận

Bảng 5. Phương trình hồi quy tuyến tính và hồi quy phân vị ứng với độ tin cậy tại trạm Hà Nội cho hạn 7 ngày

STT	Phương pháp	Phương trình	P - hệ số tự do	P - hệ số góc	Độ dài chuỗi	Ghi chú
1	OLS	$y = 0,2354x + 7,1504$	0,00	0,00	1,828	Nhận
2	Median	$y = 0,1000x - 13,6571$	0,00	1,00		Không nhận
3	q10	$y = -4,0017x - 11,6782$	1,00	0,00		Không nhận
4	q25	$y = -0,9821x - 12,0754$	1,00	0,00		Không nhận
5	q50	$y = 0,1000x - 13,6571$	0,00	1,00		Không nhận
6	q75	$y = 0,5679x + 4,2000$	0,00	0,00		Nhận
7	q90	$y = 0,69x + 23,29$	0,00	0,00		Nhận

Bảng 6. Phương trình hồi quy tuyến tính và hồi quy phân vị ứng với độ tin cậy tại trạm Hà Nội cho hạn 10 ngày

STT	Phương pháp	Phương trình	P - hệ số tự do	P - hệ số góc	Độ dài chuỗi	Ghi chú
1	OLS	$y = 0,1389x + 8,0357$	0,00	0,00	1,828	Nhận
2	Median	$y = 0,0444x + 6,5479$	0,00	1,00		Không nhận
3	q10	$y = 10,4571x - 15,6927$	1,00	0,00		Không nhận
4	q25	$y = -2,9637x - 33,8947$	1,00	0,00		Không nhận
5	q50	$y = 0,0444x + 6,5479$	0,00	1,00		Không nhận
6	q75	$y = 0,345x + 6,700$	0,00	0,00		Nhận
7	q90	$y = 0,3642x + 26,8358$	0,014	0,00		Nhận

Từ các Bảng 3 đến Bảng 6 cho thấy, tại trạm Láng các phương trình OLS và phương trình ở phân vị q75 và q90 thỏa mãn độ tin cậy thống kê cho các hạn dự báo lượng mưa lần lượt là 3 ngày, 5 ngày, 7 ngày và 10 ngày.

Trong phần tiếp theo, nghiên cứu sử dụng các phương trình thỏa mãn ý nghĩa thống kê trong các Bảng 3, 4, 5 và 6 để dự báo một số đợt mưa lớn tại trạm Láng thuộc khu vực Hà Nội. Số liệu các đợt mưa này đã được tách khỏi chuỗi số liệu xây dựng phương trình ở trên để đảm bảo tính độc lập.

### 3.3. Hiệu chỉnh mưa lớn tại Hà Nội

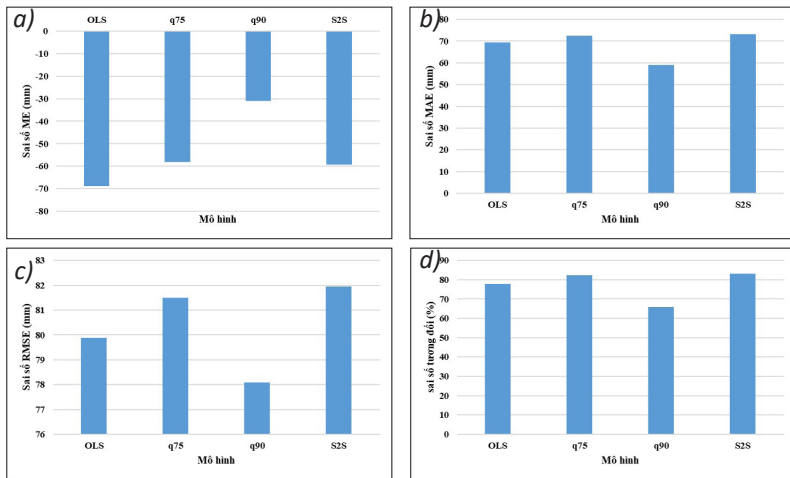
Trong phần này, nghiên cứu thử nghiệm dự báo 5 đợt mưa lớn tại Hà Nội (2004, 2006, 2011, 2012 và 2013) với tổng số ngày mưa lớn là 12 ngày có lượng mưa ngày trên 50 mm. Kết quả dự báo lượng mưa của mô hình S2S, phương trình OLS, phương trình q75 và phương trình q90 đợt mưa năm 2013 được thể hiện trong Bảng 7. Từ Bảng 7 cho thấy lượng mưa quan trắc lần lượt là 171,2 mm và 62,3 mm/ngày, tuy nhiên

mô hình S2S dự báo dưới 10 mm/ngày đối với ngày 8/8/2013 ở tất cả các hạn dự báo (3 ngày, 5 ngày, 7 ngày và 10 ngày), trong ngày 9/8/2013 mô hình S2S dự báo lượng mưa khoảng 21 mm -25 mm/ngày trong hạn dự báo 3 ngày và 5 ngày, còn ở hạn dự báo 7 ngày và 10 ngày mô hình S2S chỉ dự báo lượng mưa là 4-5 mm/ngày. Đối với phương trình OLS và q75 kết quả dự báo lượng mưa khá tương đồng nhau trong 2 ngày 8 và ngày 9 tháng 8 năm 2013, nhưng lượng mưa dự báo có thể cao hơn hoặc thấp hơn so với lượng mưa dự báo từ mô hình S2S tùy từng hạn dự báo. Đối với phương trình q90, lượng mưa dự báo trong 2 ngày kể trên đều trên 20 mm/ngày, riêng hạn 3 ngày và 5 ngày lượng mưa dự báo được trên 40 mm/ngày.

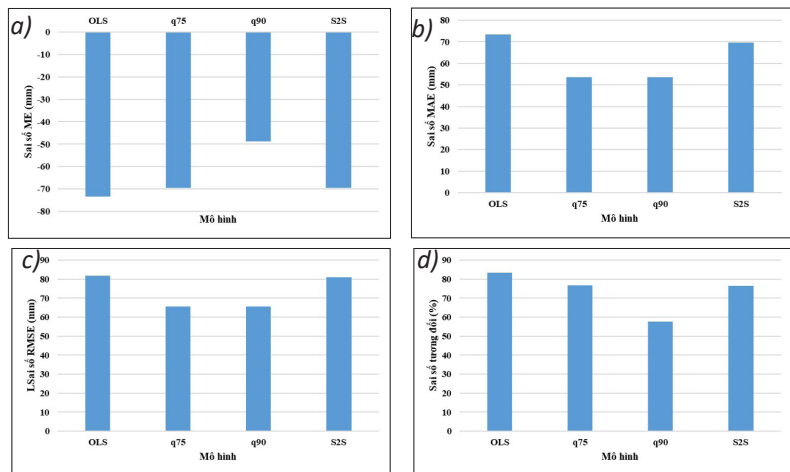
Để có nhận xét cụ thể về tính hiệu quả của từng phương trình dự báo (OLS, q75 và q90) được xây dựng trong phần 3.2, nghiên cứu tính toán các chỉ số sai số dự báo lượng mưa trung bình trong 5 đợt mưa kể trên với 12 ngày mưa lớn (Hình 4 đến Hình 7).

Bảng 7. Kết quả dự báo lượng mưa (mm) của mô hình S2S, OLS, q75 và q90 tại trạm Láng thuộc khu vực Hà Nội cho hạn dự báo 3 ngày, 5 ngày, 7 ngày và 10 ngày đợt mưa 8/8/2013 đến 9/8/2013

STT	Phương pháp	Hạn dự báo	08/8/2013	09/8/2013	Hạn dự báo	08/8/2013	09/8/2013
	OBS		171,2	62,3		171,2	62,3
1	S2S	3 ngày	8,00	21,00	7 ngày	2,00	4,00
2	OLS		8,65	14,25		7,62	8,09
3	q75		8,80	21,98		5,34	6,47
4	q90		27,87	46,81		24,67	26,05
5	S2S	5 ngày	3,00	25,00	10 ngày	3,00	5,00
6	OLS		6,68	15,52		8,45	8,73
7	q75		4,41	23,98		7,74	8,43
8	q90		21,20	47,60		27,93	28,66



Hình 4. Biểu đồ sai số ME, MAE, RMSE và sai số tương đối dự báo lượng mưa của mô hình S2S, mô hình hồi quy tuyến tính và mô hình hồi quy phân vị tại trạm Láng với hạn dự báo 3 ngày trong 5 đợt mưa

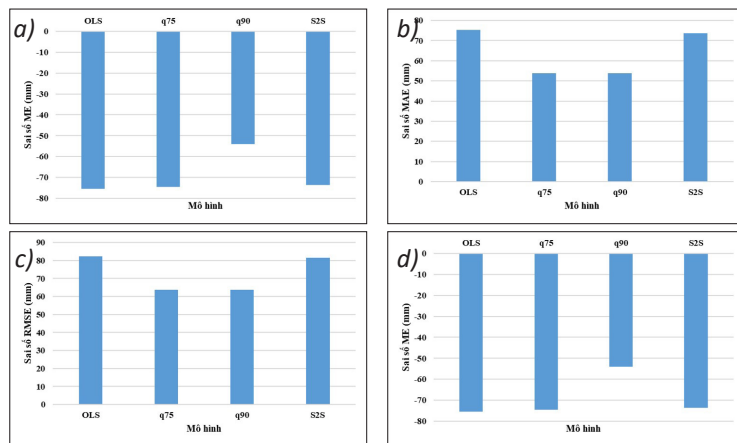


Hình 5. Biểu đồ sai số ME, MAE, RMSE và sai số tương đối dự báo lượng mưa của mô hình S2S, mô hình hồi quy tuyến tính và mô hình hồi quy phân vị tại trạm Láng với hạn dự báo 5 ngày trong 5 đợt mưa

Đối với hạn dự báo 3 ngày (Hình 4), sai số ME (Hình 4a) phản ánh dự báo thiên cao hoặc thiên thấp so với quan trắc, từ Hình 4a cho thấy mô hình S2S, OLS, q75 và q90 có giá trị ME nhỏ hơn 0, hay các mô hình này đều dự báo lượng mưa thấp hơn so với quan trắc. Sai số MAE phản ánh độ lớn của sai số dự báo lượng mưa trung bình tại trạm Láng (Hình 4b), kết quả cho thấy giá trị MAE của OLS và q90 nhỏ hơn so với sai số dự báo lượng mưa của mô hình S2S, riêng q75 có giá trị MAE tương đồng với S2S khoảng 72 mm. Đối với biên độ dao động của sai số thể hiện qua chỉ số RMSE (Hình 4c), biên độ dao động sai số dự báo lượng mưa của S2S rộng hơn so với OLS, q75 và q90, trong đó biên độ dao động sai số dự báo lượng mưa của q90 trong khoảng (-78; 78) hẹp hơn so với biên độ dao động sai số của OLS và q75. Xét về sai số tương đối, q90 cải thiện khoảng 16% sai số dự báo lượng mưa so

với S2S, OLS cải thiện khoảng 5% sai số dự báo lượng mưa so với sai số tương đối của S2S, riêng q75 hầu như không cải thiện sai số tương đối so với S2S (Hình 4d).

Đối với hạn dự báo 5 ngày (Hình 5), tương tự ở hạn 3 ngày sai số ME (Hình 5a) có giá trị nhỏ hơn) ở hầu hết các thử nghiệm dự báo (S2S, OLS, q75 và q90) hay các mô hình này đều dự báo lượng mưa thấp hơn so với quan trắc. Sai số MAE của q75 và q90 thấp hơn so với giá trị MAE của S2S và OLS, riêng giá trị MAE của OLS còn cao hơn giá trị MAE của S2S (Hình 5b). Biên độ dao động sai số của q75 và q90 hẹp hơn so với biên độ dao động sai số của S2S và OLS. Còn sai số tương đối của q90 cải thiện khoảng 20% so với sai số tương đối của S2S, trong khi đó OLS và q75 dường như không cải thiện sai số tương đối so với S2S ở hạn dự báo 5 ngày.



Hình 6. Biểu đồ sai số ME, MAE, RMSE và sai số tương đối dự báo lượng mưa của mô hình S2S, mô hình hồi quy tuyến tính và mô hình hồi quy phân vị tại trạm Láng với hạn dự báo 7 ngày trong 5 đợt mưa

Kết quả đánh giá các chỉ số sai số ME, MAE, RMSE và sai số tương đối của S2S, OLS, q75 và q90 cho dự báo lượng mưa ở hạn 7 ngày (Hình 6) khá tương đồng với dự báo lượng mưa ở hạn 5 ngày. Cụ thể, sai số ME (Hình 6a) có giá trị nhỏ hơn 0 ở hầu hết các thử nghiệm dự báo (S2S, OLS, q75 và q90) hay các mô hình này đều dự báo lượng mưa thấp hơn so với lượng mưa quan trắc. Sai số MAE của q75 và q90 thấp hơn so với giá trị MAE của S2S và OLS, riêng giá trị MAE của OLS còn cao hơn giá trị MAE của S2S (Hình 6b). Biên độ dao động sai số của q75 và q90 hẹp hơn so với biên độ dao động sai số của S2S và OLS (Hình 6c). Còn sai số tương đối của

q90 cải thiện khoảng 25% so với sai số tương đối của S2S, trong khi đó OLS và q75 dường như không cải thiện sai số tương đối so với S2S ở hạn dự báo 5 ngày (Hình 6d).

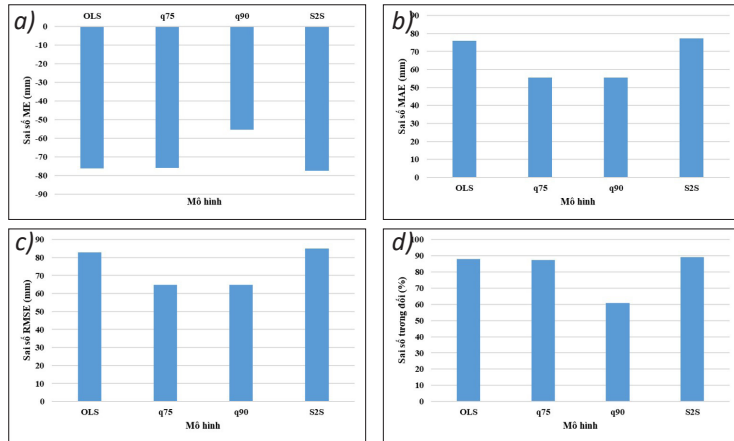
Tương tự như kết quả dự báo lượng mưa hạn 7 ngày, ở hạn 10 ngày (Hình 7), các thử nghiệm đều dự báo lượng mưa thấp hơn quan trắc khá nhiều, trong khi giá trị sai số dự báo lượng mưa dao động từ 55 mm (q90) đến 78 mm (S2S), hay q90, q75 có sai số dự báo lượng mưa thấp hơn so với S2S khá nhiều khoảng 13 mm, trong khi đó OLS có MAE thấp hơn S2S khoảng 1-2 mm. Biên độ dao động sai số của S2S rộng nhất, sau đó đến OLS và biên độ dao động sai số của q75



và q90 hẹp hơn khoảng (-65; +65) mm. Tương tự như ở hạn dự báo 7 ngày sai số tương đối của q90 giảm khoảng 29% so với sai số tương đối của S2S, còn q75 và OLS chỉ giảm khoảng 1% so với sai số tương đối của S2S.

Như vậy, ở hầu hết các hạn dự báo, phương

trình ở phân vị 90 cho kết quả dự báo lượng mưa tại trạm Láng thuộc khu vực Hà Nội khá khả quan so với số liệu gốc từ mô hình S2S, với giá trị sai số dự báo lượng mưa giảm đáng kể so với kết quả dự báo của S2S, còn sai số tương đối của q90 cải thiện từ 20 đến 29% so với sai số tương đối của S2S.



Hình 7. Biểu đồ sai số ME, MAE, RMSE và sai số tương đối dự báo lượng mưa của mô hình S2S, mô hình hồi quy tuyến tính và mô hình hồi quy phân vị tại trạm Láng với hạn dự báo 10 ngày trong 5 đợt mưa

#### 4. Kết luận

Sau khi phân tích các kết quả tính toán và đánh giá chất lượng các phương trình trong 5 đợt mưa tại Hà Nội, nghiên cứu rút ra một số kết luận sau:

Sai số lượng mưa dự báo trực tiếp từ mô hình S2S cho khu vực Hà Nội là khá lớn từ 63 mm đến 69 mm/ngày (tính trung bình cho cả các đợt mưa rơi vào tháng 6, 7, và 8 trong 20 năm). Do đó việc sử dụng trực tiếp sản phẩm từ mô hình S2S để dự báo mưa lớn cho khu vực Hà Nội là không khả thi.

Nghiên cứu dựa vào đặc điểm giá trị cực trị trong lượng mưa để xây dựng các phương trình dự báo lượng mưa dựa trên phương pháp hồi quy tuyến tính và hồi quy phân vị. Kết quả cho

thấy, phương trình OLS, q75 và q90 thỏa mãn mức tin cậy thống kê và có thể sử dụng để dự báo lượng mưa trong các hạn dự báo 3 ngày, 5 ngày, 7 ngày và 10 ngày.

Kết quả đánh giá chất lượng phương trình dự báo của OLS, q75 và q90 ở hạn 3, 5, 7 và 10 ngày thông qua 5 đợt mưa điển hình ở Hà Nội cho thấy, phương trình dự báo lượng mưa ở phân vị 90 cho kết quả khả quan nhất so với phương trình dự báo lượng mưa OLS và q75, đồng thời cải thiện sai số từ 20 đến 29% so với kết quả dự báo mưa của mô hình S2S.

Trong nghiên cứu tiếp theo, nhóm nghiên cứu sẽ áp dụng phương pháp máy học khác để dự báo mưa lớn cho khu vực Hà Nội dựa trên chuỗi số liệu là các tháng bao quát mùa mưa ở khu vực Hà Nội.

**Đóng góp của từng tác giả trong bài báo:** Xây dựng ý tưởng: Phạm Thị Minh, Lê Thị Tú Như; Xử lý số liệu: Phạm Thị Minh, Lê Thị Tú Như; Viết bản thảo: Phạm Thị Minh, Trịnh Minh Ngọc, Nguyễn Thị Phương Chi, Phan Vũ Hoàng Phương.

**Lời cảm ơn:** Các kết quả trong nghiên cứu này được tài trợ từ kinh phí của đề tài nghiên cứu khoa học cấp sinh viên năm 2024 mã số SV.2024.10 và đề tài nghiên cứu khoa học cấp trường năm 2024 mã số CT.2024.13 của trường Đại học Tài nguyên và Môi trường Thành phố Hồ Chí Minh.

**Lời cam đoan:** Tập thể tác giả cam đoan bài báo này là công trình nghiên cứu của mình, chưa từng công bố trước đó, không sao chép, đạo văn; không có sự tranh chấp lợi ích trong nhóm tác giả.

## Tài liệu tham khảo

1. Eric M. Masereka et al. (2018), "Statistical analysis of annual maximum daily rainfall for Nelspruit and its environs", *Jambá - Journal of Disaster Risk Studies*, 499. DOI: 10.4102/jamba.v10i1.499.
2. Seth Westra et al. (2013), "Global Increasing Trends in Annual Maximum Daily Precipitation", *Journal of Climate*, 26(11), 3904-3918.
3. Nguyễn Ngọc Thục, Lương Tuấn Minh (1990), "Các hình thể synop gây mưa lớn ở miền bắc Việt nam", *Tuyển tập các báo cáo khoa học tại Hội nghị Khoa học về dự báo KTTV lần thứ III 1986-1990*, tr. 110-120.
4. Vũ Đình Hải (2000), "Các dạng hình thể thời tiết gây mưa lớn ở miền Trung", *Tuyển tập về dự báo KTTV lần thứ V, 1996-2000*.
5. Trần Tân Tiến (2004), *Xây dựng mô hình dự báo các trường khí tượng thủy văn Biển Đông*, Đề tài cấp Nhà nước, mã số KC09-04.
6. Kiều Thị Xin và các cộng sự (2005), *Nghiên cứu dự báo mưa lớn diện rộng bằng công nghệ hiện đại phục vụ phòng chống lũ lụt ở Việt Nam*, Báo cáo kết quả thực hiện đề tài KHCN độc lập cấp Nhà nước, Mã số: ĐTĐL-02/2002.
7. Hoàng Đức Cường và các cộng sự (2008), *Nghiên cứu thử nghiệm dự báo mưa lớn ở Việt Nam bằng mô hình MM5*, Báo cáo tổng kết đề tài NCKH cấp Bộ TN&MT.
8. Hoàng Đức Cường và các cộng sự (2016), *Nghiên cứu xây dựng hệ thống nghiệp vụ dự báo mưa lớn cho khu vực Bắc Bộ Việt Nam*, thuộc Chương trình KC08.
9. Lê Đức và cộng sự (2007), "Thử nghiệm khai thác số liệu vệ tinh địa tĩnh bổ sung trường ẩm cho mô hình nghiệp vụ HRM. Phần II: Phương pháp thực hiện và Kết quả nghiên cứu", *Tạp chí Khí tượng thủy văn*, số 558.
10. Chu Thị Hường (2006), "Nghiên cứu thử nghiệm dự báo mưa thời hạn từ 1 đến 3 ngày cho khu vực Trung Bộ Việt Nam bằng mô hình WRF", *Tạp chí Khí tượng thủy văn*, 10/2006, 34-42.
11. Võ Văn Hòa và cộng sự (2007), *Nghiên cứu ứng dụng dự báo tổ hợp một số trường dự báo bão*, Báo cáo tổng kết đề tài NCKH cấp Bộ.
12. Vũ Thanh Hằng và Kiều Thị Xin (2007), "Dự báo mưa lớn khu vực Trung Bộ sử dụng sơ đồ tham số hóa đối lưu Heise trong mô hình HRM", *Tạp chí Khí tượng Thủy văn*, số 560, trang 49-54.
13. Đỗ Lệ Thủy và các cộng sự (2002), *Nghiên cứu ứng dụng phương pháp dự báo lượng mưa dựa trên sản phẩm mô hình dự báo số của Nhật Bản*, Báo cáo tổng kết đề tài NCUĐ cấp Tổng cục.
14. Bùi Minh Tăng và các cộng sự (2009), *Nghiên cứu, thử nghiệm dự báo định lượng mưa từ sản phẩm mô hình HRM và GSM*, Báo cáo tổng kết đề tài NCKH cấp Bộ.
15. Bùi Minh Tăng và cộng sự (2014), *Nghiên cứu xây dựng công nghệ dự báo mưa lớn thời hạn 2-3 ngày phục vụ công tác cảnh báo sớm lũ lụt khu vực Trung Bộ Việt Nam*, Báo cáo tổng kết đề tài NCKH cấp Nhà nước, 304tr.
16. Vũ Văn Thăng (2015), *Nghiên cứu cơ chế nhiệt động lực gây mưa lớn và khả năng dự báo mưa lớn mùa hè khu vực Nam Bộ và Nam Tây Nguyên do tương tác gió mùa Tây Nam-Bão trên Biển Đông*, Đề tài cấp Bộ.
17. Gao, M., et al. (2018), "Are peak summer sultry heat wave days over the Yangtze-Huaihe River basin predictable", *J. Clim.*, 31, 2185-2196.
18. Hjort, J. et al. (2011), "Spatial prediction of urban-rural temperatures using statistical methods", *Theor. Appl. Climatol.*, 106, 139-152.
19. Liu, X., Yuan, S., Li, L., (2012a), "Prediction of temperature time series based on wavelet transform and support vector machine", *JCP*, 7, 1911-1918.
20. Naing, W.Y.N., Htike, Z.Z., (2015), "Forecasting of monthly temperature variations using random forests APRN", *J. Eng. Appl. Sci.*, 10, 10109-10112.
21. Zhu, Z., Li, T., (2018), "Extended-range forecasting of Chinese summer surface air temperature and heat waves", *Clim. Dyn.*, 50, 2007-2021.

22. Hart, R. E., and R. H. Grumm (2001), "Using normalized climatological anomalies to rank synoptic-scale events objectively", *Mon. Weather Rev.*, 129(9), 2426-2442.
23. Panziera, L., and U. Germann (2010), "The relation between airflow and orographic precipitation on the southern side of the Alps as revealed by weather radar", *Q. J. Roy. Meteorol. Soc.*, 136(646), 222-238.
24. Lorenz, E. N. (1969), "Atmospheric predictability as revealed by naturally occurring analogues", *J. Atmos. Sci.*, 26(4), 636-646.
25. Klein, W. H. (1971), "Computer prediction of precipitation probability in the United States", *J. Appl. Meteorol.*, 10(5), 903-915.
26. Root, B., et al. (2007), "A fingerprinting technique for major weather events", *J. Appl. Meteorol. Climatol.*, 46(7), 1053-1066.
27. Shastri, H. et al. (2017), "Improving Global Forecast System of extreme precipitation events with regional statistical model: Application of quantile-based probabilistic forecasts", *J. Geophys. Res. Atmos.*, 122, 1617-1634, doi:10.1002/2016JD025489.
28. Koenker, R., & Bassett Jr, G. (1978), "Regression quantiles", *Econometrica: journal of the Econometric Society*, 33-50. <https://doi.org/10.2307/1913643>
29. Phạm Lê Mỹ (2016), *Phân tích đầu tư chứng khoán trên thị trường chứng khoán Việt Nam bằng phương pháp thống kê phân vị*, Luận án Tiến sĩ Kinh tế, Trường Đại học Kinh tế Quốc dân, Hà Nội.
30. Wilks Daniel S. *Statistical Methods in the Atmospheric Sciences*. Ithaca New York. 1997; 59: 255.

## ADJUSTMENT OF HEAVY RAINFALL FORECAST FROM S2S MODEL IN HANOI REGION USING SPATIAL REGRESSION METHOD

Pham Thi Minh<sup>(1)</sup>, Le Thi Tu Nhu<sup>(1)</sup>, Trinh Minh Ngoc<sup>(2)</sup>, Nguyen Thi Phuong Chi<sup>(1)</sup>,  
Phan Vu Hoang Phuong<sup>(1)</sup>, Huynh Thi Minh Suong<sup>(1)</sup>

<sup>(1)</sup>Ho Chi Minh University of Natural Resources and Environment

<sup>(2)</sup>VNU Hanoi university of science

*Received: 17/10/2024; Accepted: 8/11/2024*

**Abstract:** *The research presents the results of applying the quantile regression method to forecast heavy rainfall over a 3- to 10-day period in the Hanoi area. The study uses a time series of daily observed rainfall data at Lang station and daily forecasted rainfall data from the S2S model (the sub-seasonal to seasonal forecast model) for June, July, and August over 20 years (2001-2020) to develop forecasting equations using linear regression (OLS) and quantile regression, where the observed daily rainfall data is the dependent variable and the forecasted daily rainfall data from the S2S model is the independent variable. The evaluation results of the forecasting equations show that the OLS, q75, and q90 equations meet the statistical confidence level and can be used to forecast rainfall over 3-day, 5-day, 7-day, and 10-day forecast periods at Lang station. Additionally, the study tests the forecasting of five heavy rainfall events in the Hanoi area using the linear regression and quantile regression equations at q75 and q90 quantiles for 3-day, 5-day, 7-day, and 10-day forecast periods. The results of the forecast error assessment indicate that the q90 quantile regression equation significantly improves the rainfall forecast error compared to the S2S model across all forecast periods. Specifically, the relative error of the q90 quantile regression equation reduces the rainfall forecast error by 20 to 29% compared to the S2S model for forecast periods from 3 to 10 days.*

**Keywords:** *Quantile regression, linear regression, heavy rainfall, rainfall forecast.*