

ỨNG DỤNG KỸ THUẬT SOM THỬ NGHIỆM DỰ BÁO MƯA CỰC ĐOAN TRONG THỜI KỲ MÙA XUÂN TRÊN KHU VỰC ĐỒNG BẰNG BẮC BỘ

Nguyễn Bình Phong^{1,*}, Lê Đức Việt², Trịnh Hoàng Dương³, Lê Văn Phong³

¹Trường Đại học Tài nguyên và Môi trường Hà Nội

² Trung tâm Khí tượng hàng không, Tổng công ty Quản lý bay Việt Nam

³ Viện khoa học Khí tượng thủy văn và Biến đổi khí hậu

Tóm tắt

Các sự kiện mưa cực đoan đặt ra thách thức lớn cho công tác quản lý thiên tai, nhất là tại khu vực đồng bằng Bắc Bộ, nơi có mật độ dân cư cao. Nghiên cứu này nhằm cải thiện kỹ năng dự báo các sự kiện mưa cực đoan thông qua việc sử dụng kỹ thuật Self-Organizing Maps (SOM) và phân cụm K-means. Sử dụng số liệu khí quyển của ERA5 để học SOM, sau đó sản phẩm dự báo toàn cầu (GFS) được sử dụng cho thử nghiệm dự báo mưa cực đoan. Ngưỡng lượng mưa cực đoan được xác định là 22,3 mm/ngày, tương ứng với xác suất tích lũy 90 %. Bốn hình thái khí tượng gây mưa cực đoan trong mùa xuân đã được phân loại dựa trên các kỹ thuật nhận dạng hình thái khách quan từ số liệu ERA5 bao gồm: (i) Rãnh áp thấp qua Bắc Bộ kết hợp với hoàn lưu xoáy thấp lên đến độ cao 1.500 m; (ii) Không khí lạnh hoặc không khí lạnh lệch đông nén rãnh áp thấp kết hợp với hoàn lưu xoáy thấp lên đến độ cao 1.500 m; (iii) Hội tụ gió từ bề mặt lên đến độ cao 5.000 m; (iv) Không khí lạnh kèm front mạnh ở bề mặt kết hợp với dòng xiết cận nhiệt đới tầng cao (500 mb); (v) Bão nhiệt đới. Phương pháp SOM tiếp tục được áp dụng để hiệu chỉnh dự báo hạn 1 - 3 ngày của GFS, kết quả cho thấy độ cải thiện kỹ năng dự báo được phân nào cải thiện, nhất là khi chỉ số khoảng cách trung bình (md) nhỏ hơn 2,5. Những kết quả này cho có thể sử dụng SOM trong việc cải thiện kỹ năng dự báo mưa cực đoan hạn ngắn.

Từ khóa: SOM; Mưa cực đoan mùa Xuân; Đồng bằng Bắc Bộ.

Abstract

Application of SOM technique for extreme rainfall forecasting during the Spring season in the Northern Vietnam delta region

Extreme rainfall events pose significant challenges to disaster management, particularly in the densely populated Northern Vietnam delta region. This study aims to improve the forecasting skills of extreme rainfall events through the use of Self-Organizing Maps (SOM) and K-means clustering. Atmospheric data from ERA5 was used to train the SOM, while global forecasting system (GFS) products were utilized to test extreme rainfall predictions. The threshold for extreme rainfall was defined as 22.3 mm/day, corresponding to 90 % of cumulative probability. Four synoptic patterns responsible for extreme rainfall during the spring season were classified using objective pattern recognition techniques from ERA5 data, including: (i) A low-pressure trough over Northern Vietnam combined with a low-level cyclonic circulation extending up to 1500 m; (ii) Cold air or an eastward-slanted cold air mass compressing a low-

pressure trough, combined with a low-level cyclonic circulation up to 1500 m; (iii) Wind convergence from the surface to 5000 m; (iv) Cold air accompanied by a strong surface front, combined with a high-altitude subtropical jet stream (500 mb); (v) Tropical cyclones. The SOM method was applied to bias-correct GFS 1 - 3 day lead forecasts, showing improved forecast accuracy, especially when the mean distance index (md) was less than 2.5. These results demonstrate the potential of SOM in enhancing the quality of short-term extreme rainfall forecasts.

Keywords: SOM; Extreme spring rainfall; Northern Vietnam delta.

BBT nhận bài: 20/12/2024; Phản biện xong: 31/12/2024; Chấp nhận đăng: 26/3/2025

*Tác giả liên hệ, Email: nbphong@hunre.edu.vn

DOI: <http://doi.org/10.63064/khtnmt.2025.659>

1. Mở đầu

Các hệ thống thời tiết có thể gây mưa cực đoan ở đồng bằng Bắc Bộ (ĐBBS) vào tháng 3 và tháng 4 như: 1) Trong tháng 3 - 4, không khí lạnh vẫn có thể hoạt động, tuy với cường độ yếu hơn so với mùa Đông. Khi không khí lạnh kết hợp với hội tụ gió ở mực 1.500 - 5.000 m, có thể gây ra mưa lớn, mưa dông diện rộng ở khu vực ĐBBS; 2) Rãnh áp thấp hình thành trên khu vực phía Bắc Việt Nam, đặc biệt trong tháng 4 - 5. Khi gió Đông Nam mang theo hơi ẩm từ Biển Đông hoặc vịnh Bắc Bộ bị hội tụ dọc rãnh áp thấp, có thể gây mưa lớn kéo dài; 3) Gió Tây Nam, khi tương tác với luồng khí ẩm từ Biển Đông, có thể gây ra các trận mưa cực đoan. Đặc biệt, trong giai đoạn tháng 4 - 5, hiện tượng hội tụ nhiệt đới bắt đầu hình thành và di chuyển dần về phía Bắc, làm tăng cường sự bất ổn định trong khí quyển, dẫn đến các trận mưa lớn và dông mạnh; 4) Áp thấp nóng phía Tây thường hình thành và hoạt động mạnh dần trong tháng 4. Khi kết hợp với rãnh gió Tây từ phía Tây Bắc Bộ, có thể gây mưa dông cực đoan ở khu vực ĐBBS [4, 5, 6].

SOM, hay bản đồ tự tổ chức, là một kỹ thuật học máy không giám sát được

phát triển bởi tác giả Kohonen [10, 11], sử dụng để giảm chiều dữ liệu và phân cụm các mẫu dữ liệu có tính chất tương đồng. SOM có khả năng trực quan hóa dữ liệu phức tạp và nhận diện khách quan các mẫu (patterns) trong tập dữ liệu đa chiều, giúp phát hiện các mối quan hệ tiềm ẩn giữa các biến khí quyển, vì vậy các nghiên cứu trước đây đã đề xuất sử dụng kỹ thuật SOM để dự báo lượng mưa [7, 8, 9, 12, 13, 14, 15]. Tác giả [12] đã thực hiện dự báo mưa lớn trên lưu vực thượng nguồn Sông Nan, phía Tây Bắc Thái Lan. Tác giả đã sử dụng SOM để phân loại khách quan các hình thái thời tiết liên quan đến lượng mưa lớn tháng 5 - 9 từ năm 1979 - 2007 trên thượng nguồn Sông Nan, Tây Bắc Thái Lan. Kết quả cho thấy tháng 5 - 6 mưa lớn chủ yếu do gió mùa mùa Hè hướng Tây tăng cường đáng kể.

Ở Việt Nam, tác giả Nguyễn Văn Hưởng (2012) [1] đã sử dụng K-Means để phân loại các hình thái mưa lớn cho khu vực Miền Trung và Tây Nguyên. Tác giả đã xác định được 6 nhóm hình thái gây mưa lớn. Mặc dù các kết quả đạt được là rất khả quan, nhưng nhược điểm lớn nhất của nghiên cứu này là việc xác định 6 nhóm được thực hiện bởi chính tác giả dựa trên các thông kê synop có trước.

Nghiên cứu

Hay nói khác, việc áp dụng phương pháp K-Means thực tế là kiểm nghiệm lại các tổng kết của dự báo viên trước đây, do đó nghiên cứu này vẫn mang tính chủ quan. Để khắc phục nhược điểm của SOM, tác giả Vũ Anh Tuấn (2015) [2] đã nghiên cứu kết hợp giữa kỹ thuật SOM kết hợp K-Mean để tạo ra một hệ thống phân tích khách quan các hình thể thời tiết gây ra mưa lớn ở hầu hết các khu vực Việt Nam. Tác giả đã sử dụng bộ số liệu JRA25 trong 33 năm (1979 - 2012) phân loại các hình thể gây mưa lớn (>50 mm/24h) và chỉ ra được các hình thể thời tiết gây mưa lớn cho các khu vực Việt Nam trong mùa mưa. Gần đây, tác giả Trịnh Hoàng Dương và cộng sự (2023) [3] đã nghiên cứu dự báo rét đậm, rét hại và mưa lớn trong mùa Đông khu vực Bắc Bộ. Nghiên cứu đã áp dụng SOM nhận dạng các hình thể cho từng nút SOM để dự báo mưa lớn trong mùa Đông. Quá trình học SOM dựa trên các yếu tố khí quyển từ số liệu ERA5 của Trung tâm dự báo hạn vừa Châu Âu (ECMWF) giai đoạn 1983 - 2016. Cấu trúc SOM khi dự báo tương tự như quá trình luyện SOM, nhưng dựa trên số liệu dự báo của ECMWF, nút chiến thắng dựa trên khoảng cách Euclid để xác định hình thể tương tự với nút SOM trong quá trình học. Kết quả dự báo mưa lớn sẽ là khả năng có mưa lớn (tần suất) tương tự như nút SOM khi luyện.

2. Số liệu sử dụng và phương pháp nghiên cứu

2.1. Số liệu sử dụng

Nghiên cứu sử dụng: (1) Số liệu ERA5 (1983 - 2015) với độ phân giải $0,25^\circ$ với phạm vi từ $0 - 35^\circ\text{N}$; $90 - 130^\circ\text{E}$. Các trường khí quyển gồm gió mực 850 hPa (uv 850),

khí áp mực biển trung bình (mslp), độ cao địa thế vị mực 500 hPa (hgt 500) được sử dụng làm đầu vào cho SOM học nhằm phân loại hình thể thời tiết khách quan; (2) Số liệu khí quyển với độ phân giải tương tự như quá trình SOM học từ dự báo GFS (2016 - 2023) hạn dự báo 1 - 3 ngày của Trung tâm Dự báo Môi trường (NCEP) được sử dụng để thử nghiệm dự báo dựa trên SOM; (3) Số liệu mưa tại 15 trạm khí tượng và số liệu mưa vệ tinh PERSIANN (Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks) độ phân giải $0,25^\circ$ được nghiên cứu sử dụng cho phân tích phân bố mưa; (4) Tài liệu về “Đặc điểm khí tượng thủy văn hàng năm” trong giai đoạn 1994 - 2017 cho tổng hợp hình thể thời tiết trong mùa Xuân.

2.2. Phương pháp thử nghiệm dự báo mưa cực đoan dựa trên SOM

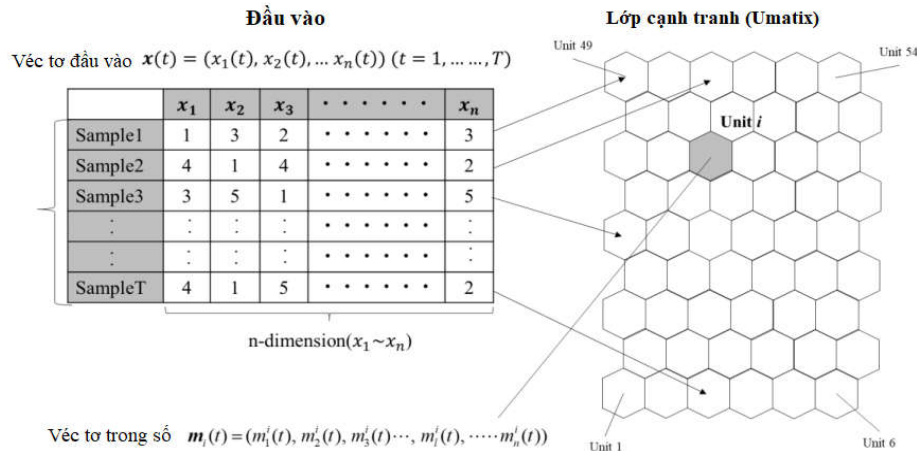
Nghiên cứu đã sử dụng cách tiếp cận kỹ thuật SOM cho dự báo lượng mưa cực đoan ở ĐBBB trong mùa Xuân (cụ thể về kỹ thuật SOM của các tác giả [2, 3, 12, 13], có thể tóm tắt như sau:

- *Kỹ thuật SOM*: SOM được cấu hình như một mạng ANN hai lớp, sử dụng thuật toán học không giám sát (tự phân tích dữ liệu), bao gồm các quá trình học lặp đi lặp lại để cập nhật dần các nút dữ liệu trong bản đồ đầu ra cho đến khi hội tụ ổn định đại diện cho không gian đầu vào. Mỗi bước học của SOM bắt đầu bằng việc chọn ngẫu nhiên một vector trọng số đầu vào. Các vector đầu vào cho lớp đầu ra có thể được tính theo công thức . Quá trình học SOM theo các bước: (1) Tìm nơ-ron chiến thắng (Best Matching Unit - BMU), nơ-ron có trọng số gần nhất với mẫu dữ liệu đầu vào được chọn làm

BMU = $\text{argmin}_{ij}(x, w_{ij})$. Khoảng cách được tính bằng khoảng cách Euclid; (2) Cập nhật trọng số: Trọng số của BMU và các nơ-ron lân cận được cập nhật để tiến gần hơn đến giá trị của mẫu đầu vào: $W_{(t+1)} = W_{(t)} + \eta_{(t)} \cdot h(\text{BMU}, t) \cdot (X - W_{(t)})$. Trong đó $W_{(t)}$ là trọng số hiện tại; $\eta_{(t)}$ là tốc độ học; $h(\text{BMU}, t)$ là hàm lân cận (neighborhood function), xác định mức độ ảnh hưởng của BMU lên các nơ-ron lân cận; X là mẫu số liệu đầu vào.

Trong SOM, các vector đầu vào có đặc điểm tương tự sẽ được ánh xạ và phân loại vào cùng một đơn vị (nút/nơ-ron),

được đại diện bởi một vector trọng số trong lớp cạnh tranh. Vector trọng số này biểu thị đặc điểm trung bình của tất cả các vector đầu vào được ánh xạ vào nút đó. Cả vector đầu vào và vector trọng số đều có cùng số chiều để đảm bảo tính tương đồng trong tính toán. Trong SOM, các nút có thể có nhiều hình dạng khác nhau tùy thuộc vào nghiên cứu, như hình vuông hoặc hình lục giác. Nghiên cứu này sử dụng các nút có hình lục giác đều, vì đặc điểm của nó cho phép khoảng cách giữa một nút và sáu nút lân cận gần nhất luôn bằng nhau trên toàn bộ bản đồ, giúp duy trì tính nhất quán trong việc tính toán và biểu diễn dữ liệu.



Hình 1: Mô hình khái niệm của SOM

- *Phương pháp PCA*: Cho tập dữ liệu gồm n quan trắc của m điểm lưới x_1, \dots, x_m , một ma trận (X) kích thước $(n \times m)$. PCA giảm (X) bằng cách chiếu hình học nó lên kích thước thấp hơn được gọi là các thành phần chính, với mục tiêu khái quát X tốt nhất khi giới hạn số lượng (k) thành phần chính (PC_k). X có thể được viết dưới dạng hàm ngẫu nhiên của $X(t,s)$ như sau: . Trong đó tất cả (k) véc tơ EOF_i là trực giao với nhau và bằng giá trị riêng thứ i^{th} .

- *Phương pháp Kmean*: (1) Khởi tạo ngẫu nhiên hoặc chọn K tâm cụm (centroids); (2) Tính khoảng cách giữa

từng điểm dữ liệu và các tâm cụm (t khoảng cách Euclid/ Mahalanobis); (3) Tính toán lại tâm cụm mới bằng cách lấy trung bình của tất cả các điểm dữ liệu trong cụm: . Trong đó C_i là tâm cụm i; S_i là tập hợp các điểm trong cụm i.

- *Phương pháp Bootstrap*: Giả sử có một tập dữ liệu gồm 10 mẫu $x_1, x_2, x_3, \dots, x_{10}$, trong một lần bootstrap sẽ tạo ra một tập dữ liệu ngẫu nhiên như $x_2, x_5, x_8, x_1, x_3, x_{10}, x_7, x_2, x_4$, SOM sẽ được huấn luyện trên tập mẫu ngẫu nhiên và quy trình này có thể lặp lại 1.000 lần với các tập bootstrap khác nhau.

Nghiên cứu

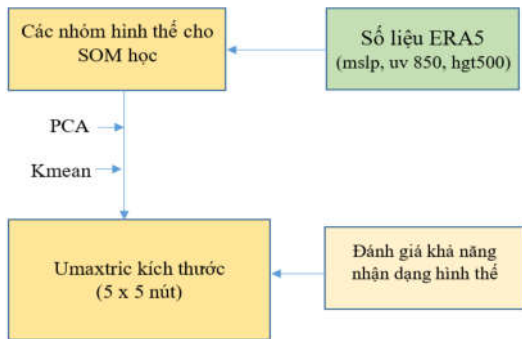
Trong giai đoạn huấn luyện: Các trường khí quyển từ ERA5 với đầu vào gồm 4 biến khí quyển được sử dụng. Áp dụng giảm chiều bằng PCA, giữ lại các thành phần chính (PC) có độ tin cậy với phương sai trên 90 %, phương pháp bootstrap (B = 1.000) tạo dữ ổn định véc tơ đầu vào, và sau đó là áp dụng K-means cho phân cụm khách quan và hiển thị kết quả trực quan bằng U-Matrix (Unified Distance Matrix) (Hình 1). U-Matrix là một công cụ trực quan mạnh mẽ trong SOM, được sử dụng để biểu diễn khoảng cách giữa các nơ-ron lân cận trên bản đồ. Hình 3 biểu diễn khoảng cách Euclid/ Mahalanobis trung bình giữa các nơ-ron lân cận. Các ô màu tối (màu xanh đậm hơn), biểu thị khoảng cách lớn giữa các nơ-ron, thường là ranh giới giữa các cụm khác nhau. Các ô màu sáng (màu xanh nhạt) biểu thị khoảng cách nhỏ giữa các nơ-ron, thường đại diện cho các nơ-ron trong cùng một cụm. Trong hình minh họa, các điểm dữ liệu (kết quả của K-means) được gán vào các nơ-ron trên bản đồ và được phân biệt bằng các màu khác nhau cho 4 cụm tương ứng (từ 1 - 4) cho 4 hình thể trong Hình 5.

Trong giai đoạn dự báo: Cấu trúc của SOM tương tự quá trình đào tạo, các biến khí quyển giai đoạn 2016 - 2023 được sử dụng từ dự báo của mô hình GFS với hạn dự báo từ 1 đến 3 ngày và làm đầu vào cho SOM. Bài báo tính toán khoảng cách md_j , về cơ bản là khoảng cách giữa vector đầu vào mới $P = [P_1, P_2, \dots, P_d]$ và centroid $c_j = [c_{1j}, c_{2j}, \dots, c_{dj}]$ của cụm j ($1 \leq j \leq K$), được chuẩn hóa theo độ lệch chuẩn của cụm trong mỗi chiều. Do đó, khoảng cách đo độ tương đồng giữa vector mới (dự báo) và tâm cụm (centroid) trong không gian PCA sẽ là:

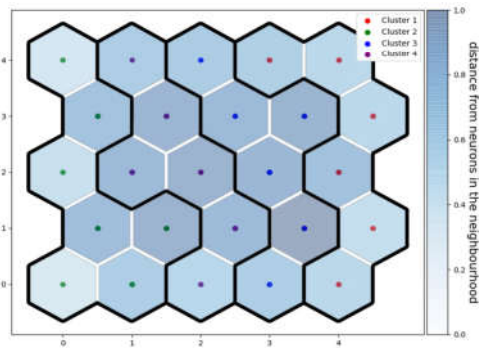
$$md_j = \sqrt{\sum_{i=1}^d \left(\frac{p_i - c_{ij}}{\sigma_{ij}} \right)^2}$$

trong đó: p_i : Thành phần thứ i của vector dự báo mới P ; c_{ij} : Thành phần thứ i của tâm cụm j ; σ_{ij} : Độ lệch chuẩn của cụm j tại thành phần i ;

Nếu khoảng cách nhỏ nhất $\min(md_j) \leq z$ (với thử nghiệm z khác nhau, vector dự báo được gán cho cụm có khoảng cách md_j nhỏ nhất ($c = \text{argmin}_j md_j$). Điều này nghĩa là hình thể thời tiết dự báo đủ tương đồng với hình thể đã xảy ra trong lịch sử thuộc cụm đó. Dựa trên phương pháp tương tự, lượng mưa cực đoan được dự báo sẽ xảy ra trên khu vực nghiên cứu là tương tự với lượng mưa của cụm được dự báo. Ngưỡng của $\min(md_j)$ được xác định sau khi thực hiện một số thử nghiệm độ nhạy, trong đó giá trị của nó được thay đổi từ 0,5 đến 4.



Hình 2: Sơ đồ SOM học

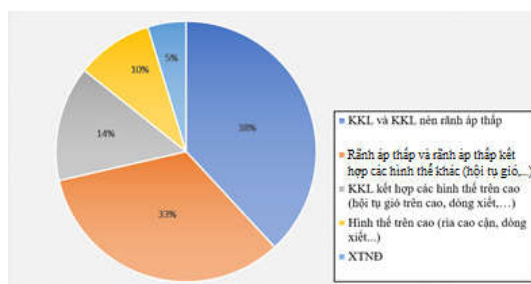


Hình 3: Bản đồ (U-matrix) phân cụm SOM

3. Thử nghiệm dự báo mưa cực đoan trong thời kỳ mùa Xuân dựa trên SOM

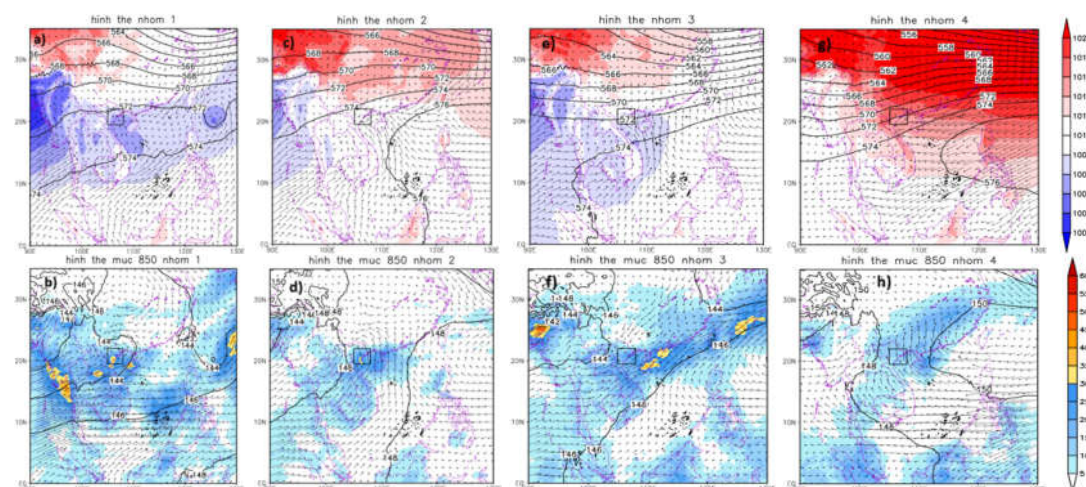
3.1. Ứng dụng kỹ thuật SOM trong phân loại hình thể

Thống kê và phân loại 21 đợt mưa cực đoan diện rộng từ năm 1994 - 2017 (tổng hợp các hình thể từ đặc điểm khí tượng thủy văn hàng năm), trên khu vực ĐBBB nhận thấy có 05 nhóm hình thể chủ đạo gây mưa trong mùa Xuân (Hình 4): 1) Không khí lạnh (KKL) và KKL nén rãnh áp thấp; 2) Rãnh áp thấp và rãnh áp thấp kết hợp các hình thể khác (hội tụ gió,...); 3) KKL kết hợp các hình thể trên cao (hội tụ gió trên cao, dòng xiết,...); 4) Hình thể trên cao (rìa cao cận, dòng xiết,...); 5) Xoáy thuận nhiệt đới (XTNĐ).



Hình 4: Phân loại hình thể gây mưa cực đoan ở đồng bằng Bắc Bộ

Dựa trên chuỗi số liệu ERA5 giai đoạn 1983 - 2015 trong các ngày xảy ra mưa cực đoan diện rộng, ngưỡng mưa cực đoan được xác định là ngày có lượng mưa trên 22,3 mm/ngày tương ứng xác suất tích lũy 90 %, bài báo đã tiến hành phân cụm SOM và được dẫn ra dưới dạng ma trận U-Matrix gồm 25 nút (5 × 5) (Hình 3). Các ngày mưa cực đoan được gán vào các cụm dữ liệu, trích xuất các ngày này trong mỗi cụm này từ đó xây dựng bản đồ synop đặc trưng cho từng cụm theo thứ tự (từ 1 đến 4) tương ứng trong Hình 3 và cụ thể (Hình 5): Nhóm 1/cụm 1: Rãnh áp thấp qua khu vực Bắc Bộ kết hợp với hoàn lưu xoáy thấp lên đến độ cao 1.500 m; Nhóm 2/cụm 2: KKL hoặc KKL lệch đông nén rãnh áp thấp kết hợp với hoàn lưu xoáy thấp lên đến độ cao 1.500 m; Nhóm 3/cụm 3: Hội tụ gió từ bề mặt lên đến độ cao 5.000 m; Nhóm 4/cụm 4: KKL kèm front mạnh ở bề mặt kết hợp với dòng xiết cận nhiệt đới tầng cao (500 mb). Có thể nhận thấy, kết quả phân tích 4 hình thể trong mùa Xuân dựa trên SOM phù hợp với 4 trong 5 hình thể chủ đạo đã được tổng hợp hình thể từ đặc điểm khí tượng thủy văn [4] hàng năm từ năm 1994 - 2017.

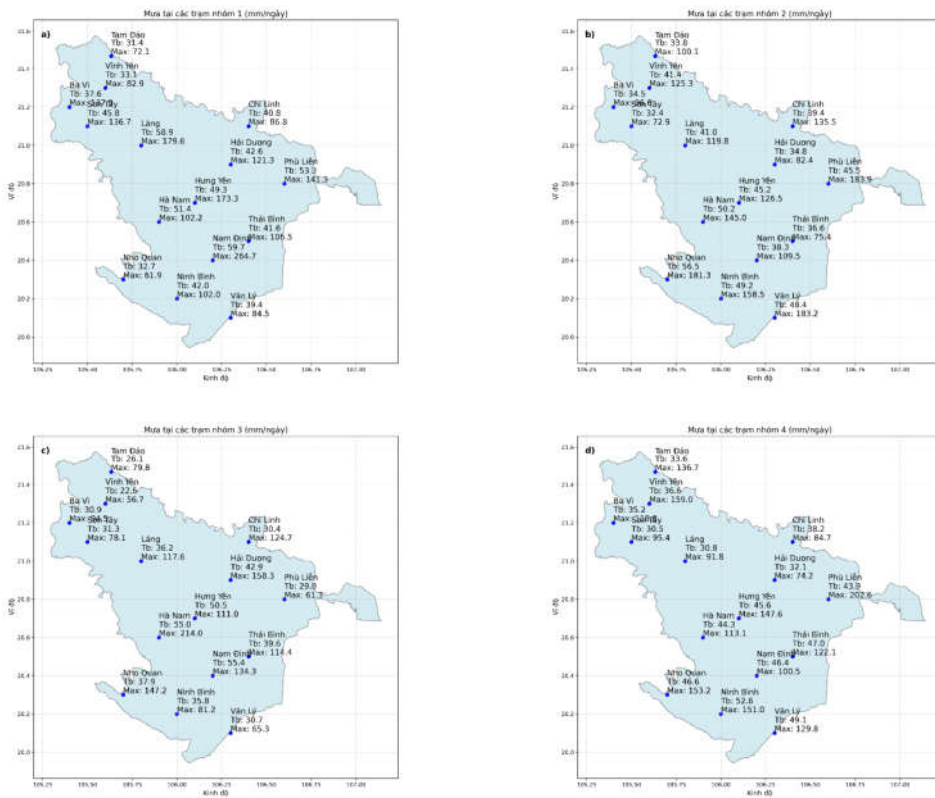


Hình 5: Phân loại hình thể gây mưa trên khu vực đồng bằng Bắc Bộ. Trong hình (a, c, e, g), đường contour là hgt500 hpa, véc tơ gió mực 500 hpa và tô màu cho mslp. Trong hình (b, d, f, h), đường contour là hgt850 hpa véc tơ gió mực 850 hpa, và mưa vệ tinh PERSIANN

Nghiên cứu

Hình 6 phân tích lượng mưa trung bình (Tb) và cực đại (Max) tại các trạm ở ĐBBB theo các nhóm: Đối với nhóm 1 (Hình 6a): Phía Bắc (Tam Đảo, Ba Vì) có Tb khoảng 32 - 45 mm/ngày, Max khoảng 72 - 137 mm/ngày. Khu vực trung tâm (Láng, Hưng Yên) có Tb tăng mạnh, cao nhất tại Láng (179 mm/ngày). Phía Nam có mức cực đoan lớn nhất, Nam Định đạt Max cao nhất toàn khu vực (264 mm/ngày), Tb cao nhất (59,7 mm/ngày). Nhóm 2 (Hình 6b): Phía Bắc có Tb khoảng 30 - 34 mm/ngày, Max 56 - 100 mm/ngày. Trung tâm có Tb khoảng 36 - 45 mm/ngày, Max tại Hưng Yên đạt 126 mm/ngày. Phía Nam (Hà Nam, Nam Định) ghi nhận Max cao nhất (145 mm/ngày), phản ánh mưa cực đoan cục bộ. Nhóm 3 (Hình 6c): Phía Bắc Tb thấp (22 - 37 mm/ngày),

Max không quá 137 mm/ngày. Trung tâm Tb khoảng 36 - 50 mm/ngày, Max khoảng 111 - 158 mm/ngày. Phía Nam có Max cao hơn (134 - 158 mm/ngày), trạm Hà Nam đạt 214 mm/ngày. Nhóm 4 (Hình 6d): Phía Bắc có Tb thấp nhất (30 - 36 mm/ngày), Max 136 mm/ngày. Trung tâm Tb ổn định (30 - 46 mm/ngày), Max tại Hưng Yên đạt 147 mm/ngày. Phía Nam Max 100 - 153 mm/ngày, Phù Lãng đạt cao nhất (202 mm/ngày). Như vậy, phía Bắc có lượng mưa thấp nhất, ít mưa cực đoan (Max cao nhất 137 mm/ngày). Khu vực trung tâm duy trì mức Tb và Max ổn định hơn. Phía Nam có lượng mưa cực đoan nhất, đặc biệt tại Nam Định (Max 264 mm/ngày), cho thấy ảnh hưởng mạnh mẽ của các hiện tượng mưa lớn



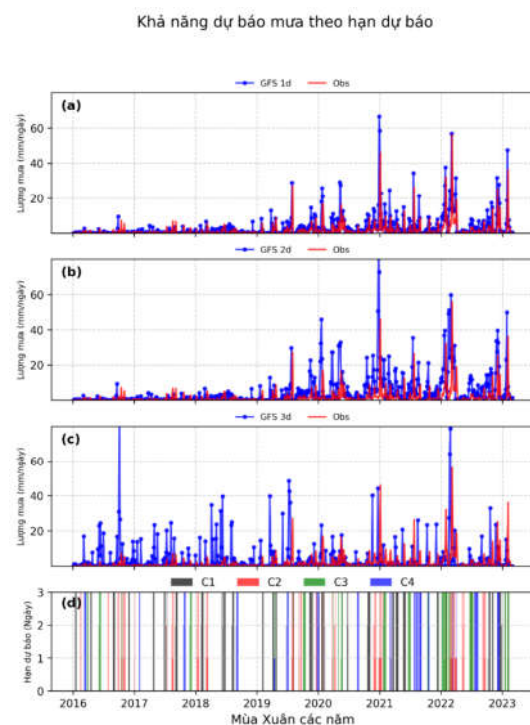
Hình 6: Lượng mưa trung bình tại trạm (Tb, mm/ngày) và cực đại (Max, mm/ngày) theo 4 nhóm trong giai đoạn học SOM: (a) C1, (b) C2, (c) C3, và (d) C4 ở đồng bằng Bắc Bộ

3.2. Kết quả thử nghiệm dự báo mưa cực đoan dựa trên kỹ thuật SOM

Bài báo sử dụng SOM để dự báo mưa cực đoan diện rộng, dựa trên 4 nhóm hình thể đã học từ dữ liệu quá khứ, với kết quả “có” hoặc “không” xảy ra mưa cực đoan. Dự báo “có” nghĩa là điều kiện hiện tại rơi vào một trong 4 nhóm và “không” khi không thuộc các nhóm này. Các kịch bản mưa trong quá khứ được dùng làm tham chiếu, ví dụ nếu dự báo rơi vào nhóm 2, SOM đại diện cho hình thể KKL và KKL nén rãnh áp thấp, phân bố mưa cực đoan có thể tương tự như Hình 5d và Hình 6b và cùng theo đó là những tác động đi kèm như ngập lụt trong quá khứ. Thông tin này góp phần trong nhận định dự báo mưa cực đoan và tác động của chúng, tuy nhiên tính phù hợp của giả định cần được kiểm tra, vì biến đổi khí hậu có thể làm thay đổi hình thể thời tiết và cần tham khảo nhiều kết quả dự báo trong quá trình nhận định, bởi vì dự báo mưa cực đoan là bài toán khó bởi tính biến động của mưa vùng nhiệt đới là rất lớn.

Có thể nhận thấy kết quả dự báo của GFS thông qua Hình 7, dự báo hạn 1 ngày có độ tin cậy cao nhất, với thời gian mưa lớn được dự đoán chính xác, nhưng thường phóng đại cường độ (Hình 7a). Dự báo hạn 2 ngày bắt đầu có sự chênh lệch về thời gian và cường độ, một số sự kiện mưa lớn bị bỏ sót hoặc sai lệch (Hình 7b). Đối với hạn 3 ngày, sai lệch tăng đáng kể, với xu hướng phóng đại cường độ mưa và dự báo giả mưa cao hơn, làm giảm độ tin cậy (Hình 7c). Mặc dù độ chính xác giảm dần theo thời gian, mô hình vẫn xác định được xu hướng chung của lượng mưa, góp phần hỗ trợ nhận định về tác động tiềm tàng của mưa cực đoan. Hình 7d dẫn

ra phân bố các hình thể thời tiết (C1, C2, C3, C4) tương ứng với các ngày trong mùa Xuân, được phân loại và dự báo với hạn 1 - 3 ngày cho thấy tương đồng với xu hướng đỉnh mưa lớn quan trắc. Các đỉnh mưa lớn (>40 mm/ngày) xuất hiện không đồng đều nhưng có xu hướng gia tăng về tần suất vào các năm sau (đặc biệt giai đoạn 2020 - 2023).



Hình 7: Lượng mưa (mm/ngày) trong mùa Xuân trung bình ĐBBB khu vực từ 2016 - 2023 của quan trắc tại trạm (màu đỏ) và dự báo GFS (màu xanh) theo các hạn dự báo 1 ngày (a), 2 ngày (b) và 3 ngày (c). (d) là sự xuất hiện (có/không) của ngày mưa theo các nhóm được dự đoán bằng cách sử dụng các trường khí quyển từ dự báo GFS với hạn dự báo từ 1 - 3 ngày

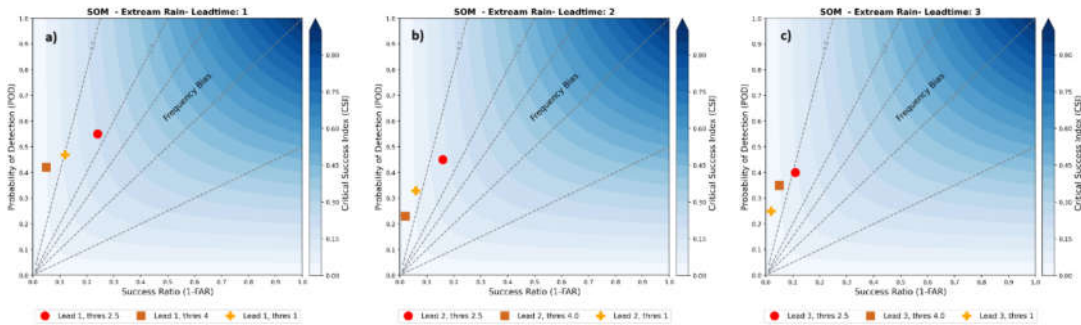
Hình 7d cho thấy phần nào mối liên kết trực tiếp giữa khả năng nhận diện hình thể mưa cực đoan. Để đánh giá cụ thể hơn, việc đánh giá kỹ năng dự báo của SOM sẽ tiến hành xác định các thành phần của chỉ số đánh giá như sau: (1) Trường hợp quan

Nghiên cứu

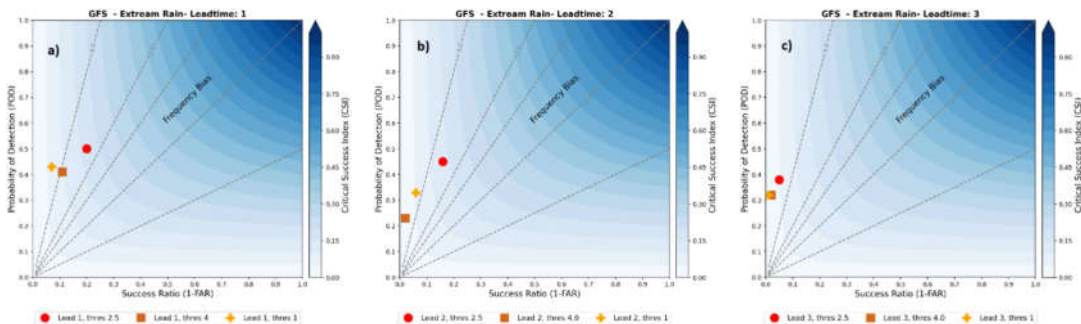
trắc “có” và dự báo “có” mưa cực đoan phải thỏa mãn điều kiện: (i) Mưa quan trắc xảy ra diện rộng với một nửa số trạm; (ii) Dự báo rơi vào nút SOM có hình thể trong quá khứ là 4 cụm; (2) Trường hợp quan trắc “không” và dự báo “không”: (i) Quan trắc không xảy ra diện rộng trên một nửa số trạm; (ii) Dự báo có nút SOM không thuộc 4 hình thể thuộc 4 cụm. Với sự kết hợp trường hợp “có” và “không” có thể sẽ tính toán các chỉ số đánh giá. Điều này cũng được thực hiện để tính A, B, C, D trực tiếp từ lượng mưa của GFS tương ứng với các hình thể đã xác định này.

Kết quả cho thấy dự báo dựa trên SOM, chỉ số POD phổ biến 0,4 - 0,55, trong một vài trường cao hơn hoặc thấp hơn. Nếu xem xét chỉ số FAR (1-FAR thấp) có thể thấy tỉ lệ dự báo không khá

cao phổ biến 0,7 - 0,98 cho các hạn dự báo. Nguyên nhân có thể là do mặc dù điều kiện khí quyển trên diện rộng rất thuận lợi nhưng không phải lúc nào mưa lớn cũng rơi vào khu vực nghiên cứu. Ngoài ra, mưa biến đổi khá mạnh mẽ, hình thể gây mưa lớn trong nghiên cứu chỉ được phân tích trên cơ sở trung bình hàng ngày có thể cũng làm giảm sự phân tích của SOM. Thử nghiệm với các ngưỡng z khác nhau, ở đây đưa ra 3 ngưỡng z là 4, 2.5, 1, kết quả cho thấy $z = 2.5$. Các chỉ số đánh giá cao hơn, cho thấy sử dụng ngưỡng này sẽ là phù hợp hơn các ngưỡng khác trong bài toán thử nghiệm dự báo mưa cực đoan trong mùa Xuân (tháng 3, 4, và 5) (Hình 8 và 9) bởi vì số ngày mưa diện rộng ít cũng có thể dẫn đến khả năng phân tích khách quan của SOM hạn chế đáng kể.



Hình 8: Giản đồ hiệu suất dự báo xảy ra mưa cực đoan cho vùng nghiên cứu dựa trên SOM theo hạn dự báo 1 (a), 2 (b) và 3 (c) ngày; Chấm tròn màu đỏ cho ngưỡng mdj < 2.5; Hình vùng màu cam cho ngưỡng mdj < 1.0 và dấu cộng cho mdj < 4.0



Hình 9: Giản đồ hiệu suất dự báo xảy ra mưa cực đoan cho vùng nghiên cứu dựa trên GFS theo hạn dự báo 1 (a), 2 (b) và 3 (c) ngày; Chấm tròn màu đỏ cho ngưỡng mdj < 2.5; Hình vùng màu cam cho ngưỡng mdj < 1.0 và dấu cộng cho mdj < 4.0

So sánh giữa các chỉ số của dự báo dựa trên SOM và trực tiếp từ GFS cho thấy các chỉ số POD đều cao hơn với trị số khoảng 0,2 - 0,4. Tuy nhiên dự báo không khá cao, chỉ số 1-FAR dao động từ 0 - 0,2. Nhìn chung kết quả nghiên cứu này thấp hơn so với những nghiên cứu trước đây về ứng dụng SOM trong xây dựng công nghệ dự báo mưa lớn vùng nhiệt đới và ngoại nhiệt đới, cũng như khu vực Thái Lan [12]. Tuy nhiên, có thể thấy rằng khả năng dự báo về sự xuất hiện của lượng mưa cực đoan trên khu vực ĐBBB có thể được cải thiện đáng kể trong một số trường hợp bằng cách sử dụng SOM trong cả 3 hạn dự báo.

4. Kết luận

Nghiên cứu này với mục tiêu ứng dụng phương pháp SOM trong dự báo mưa cực đoan trên khu vực đồng bằng Bắc Bộ đã thu được một số kết luận sau: 1) Trên cơ sở ngưỡng lượng mưa 22,3 mm đã xác định các ngày mưa của đợt mưa diện rộng, các ngày mưa này là cơ sở cho xác định 4 véc tơ đầu vào từ ERA5. Kết quả cho thấy SOM có thể phân tích khách quan được 4 hình thể chính: (i) Rãnh áp thấp đi qua khu vực Bắc Bộ kết hợp với hoàn lưu xoáy thấp lên đến độ cao 1.500 m; (ii) Không khí lạnh hoặc không khí lạnh lệch Đông nén rãnh áp thấp kết hợp với hoàn lưu xoáy thấp lên đến độ cao 1.500 m; (iii) Hội tụ gió từ bề mặt lên đến độ cao 5.000 m; (iv) Không khí lạnh kèm front mạnh ở bề mặt kết hợp với dòng xiết cận nhiệt đới tầng cao (500 mb); 2) Trong quá trình dự báo, cấu trúc SOM là tương tự như quá trình học SOM, số liệu đầu vào là dự báo hạn 1 - 3 ngày của GFS (hiệu chỉnh

dự báo dựa trên SOM). Kết quả đánh giá dự báo “có” và “không” xuất hiện mưa cực đoan cho thấy với ngưỡng mdj < 2.5 cho kết quả dự báo tốt hơn, trong một số trường hợp sử dụng SOM, kết quả dự báo được cải thiện đáng kể.

Lời cảm ơn: Bài báo chân thành cảm ơn sự hỗ trợ của đề tài “*Nghiên cứu, tổng hợp và đánh giá tiềm năng các dạng tài nguyên năng lượng mặt trời, gió, sóng, thủy triều, địa nhiệt, sinh khối, lựa chọn các dạng tài nguyên năng lượng phù hợp có thể phát triển ở Việt Nam*”. Mã số TNMT.2023.05.01 do ThS. Trần Quang Năng làm chủ nhiệm đã hỗ trợ số liệu, phương pháp đánh giá kết quả trong quá trình thực hiện và công bố nghiên cứu này.

TÀI LIỆU THAM KHẢO

- [1]. Nguyễn Văn Hưởng (2012). *Xác định khách quan hình thể thời tiết trong các đợt mưa lớn trên khu vực Miền Trung từ số liệu tái phân tích JRA25*. Luận văn Thạc sĩ trường Đại học Khoa học Tự nhiên, Đại học Quốc gia Hà Nội.
- [2]. Vũ Anh Tuấn và cộng sự (2015). *Nghiên cứu xây dựng hệ thống xác định khách quan các hình thể gây mưa lớn điển hình cho khu vực Việt Nam*. Đề tài cấp Bộ Tài nguyên và Môi trường.
- [3]. Trịnh Hoàng Dương và cộng tác viên (2023). *Nghiên cứu đổi mới công nghệ dự báo các đợt rét đậm, rét hại, mưa lớn trong mùa Đông ở khu vực phía Bắc Việt Nam hạn đến 10 ngày dựa trên quy luật hoạt động của dòng xiết cận nhiệt đới*. Đề tài cấp Bộ Tài nguyên và Môi trường.
- [4]. Tổng cục Khí tượng Thủy văn (2020). *Báo cáo diễn biến thời tiết và khí hậu Việt Nam*. Bộ Tài nguyên và Môi trường.
- [5]. Nguyễn Ngọc Thạch và cộng sự (2019). *Hệ thống thời tiết gây mưa lớn ở Miền Bắc Việt Nam*. Tạp chí Khí tượng Thủy văn, Số 702, tr. 11 - 19.

Nghiên cứu

[6]. Viện Khoa học Khí tượng Thủy văn và Biến đổi khí hậu (2018). *Tác động của hội tụ gió và rãnh áp thấp trong điều kiện khí hậu Việt Nam*. Nxb. Khoa học và Kỹ thuật.

[7]. Borah, N., A. K. Sahai, R. Chattopadhyay, S. Joseph, S. Abhilash, and B. N. Goswami (2013). *A self-organizing map-based ensemble forecast system for extended range prediction of active/break cycles of Indian summer monsoon*. J. Geophys. Res. Atmos., 118, 9022 - 9034.

[8]. Chattopadhyay, R., A. K. Sahai, and B. N. Goswami (2008). *Objective identification of nonlinear convectively coupled phases of monsoon intraseasonal oscillation: Implications for prediction*. J. Atmos. Sci., 65, 1549 - 1569. <https://doi.org/10.1175/2007JAS2474.1>.

[9]. Gutiérrez, J. M., R. Cano, A. S. Cofiño, and C. Sordo (2005). *Analysis and downscaling of multi-model seasonal forecasts in Peru using self-organizing maps*. Tellus, 57A, 435 - 447.

[10]. Kohonen, T., (1982). *A simple paradigm for the self-organized formation of structured feature maps*. In *Competition and Cooperation in Neural Nets*; Springer: Berlin, Heidelberg, p. 248 - 266.

[11]. Kohonen, T., (2001). *Self-organizing maps*. 3rd ed.; Springer: Berlin, Germany.

[12]. Nguyen-Le, D., & Yamada, T. J., (2019). *Using weather pattern recognition to classify and predict summertime heavy rainfall occurrence over the Upper Nan River basin, Northwestern Thailand*. Weather and Forecasting. Doi:10.1175/waf-d-18-0122.1.

[13]. Nguyen-Le, D., Yamada, T. J., & Tran-Anh, D., (2017). *Classification and forecast of heavy rainfall in Northern Kyushu during the Baiu season using weather pattern recognition*. Atmospheric Science Letters, 18(8), 324 - 329. Doi:10.1002/asl.759.

[14]. Nishiyama, K., S. Endo, K. Jinno, C. B. Uvo, J. Olsson, and R. Berndtsson (2007). *Identification of typical synoptic patterns causing heavy rainfall in the rainy season in Japan by a self-organizing map*. Atmos. Res., 83, 185 - 200.

[15]. Ohba, M., S. Kadokura, D. Nohara, and Y. Toyoda (2016). *Rainfall downscaling of weekly ensemble forecasts using self-organizing maps*. Tellus, 68A, 29293.