TẠP CHÍ
KHÍ TƯỢNG THỦY VĂN

Check for updates

Bài báo khoa học

Phân loại đặc điểm mưa và ước tính mưa bằng quan trắc Ra đa phân cực cho khu vực Tây Bắc

Nguyễn Khắc Quân[1]*, Phạm Thị Thanh Ngà[2]

[1] Đài khí tượng thủy văn khu vực Tây Bắc; mr.nguyenkhacquan@gmail.com
[2] Viện Khoa học Khí tượng Thủy văn và Biến đổi khí hậu; pttnga.monre@gmail.com

*Tác giả liên hệ: mr.nguyenkhacquan@gmail.com; Tel.: +84–966031851

Ban Biên tập nhận bài: 5/4/2023; Ngày phản biện xong: 20/5/2023; Ngày đăng bài: 25/5/2023

Tóm tắt: Mưa nhìn chung có thể mang đặc điểm đối lưu và phân tầng, mỗi loại mưa gắn với các quá trình phát triển vi vật lí mây là khác nhau dẫn đến sự khác nhau về cường độ, thời gian và diện mưa. Sử dụng số liệu quan trắc Ra đa phân cực trên khu vực Tây Bắc, nghiên cứu đã tiến hành xây dựng một phương pháp nhằm phân loại đặc điểm mưa dựa trên thuật toán phân loại Máy vectơ hỗ trợ. Kết quả phân loại sau đó được sử dụng nhằm tìm ra các hệ số thực nghiệm trong các công thức ước tính mưa dựa trên phương pháp bình phương tối thiểu. Qua đánh giá cho thấy với phương pháp phân loại được đưa ra chỉ dưới 10% khu vực có dải sáng bị phân loại là mưa đối lưu, phương pháp cũng đã khắc phục được các phân loại sai so với phương pháp tham khảo trong các trường trường hợp mưa được xem xét. Kết quả ước tính mưa cho thấy lượng mưa ước tính từ Ra đa thấp hơn giá trị quan trắc, các công thức với việc phân loại đặc điểm mưa và sử dụng các biến phân cực có kết quả tốt hơn công thức chỉ sử dụng giá trị của Z, trong đó công thức R(Z, ZDR, KDP) có hệ số tương quan cao và sai số nhỏ nhất.

Từ khóa: Phân loại đặc điểm mưa; Máy vectơ hỗ trợ; Ra đa phân cực; Tây Bắc.

1. Mở đầu

Trên khu vực nhiệt đới mưa nhìn chung có thể mang đặc điểm phân tầng và đối lưu. Mưa đối lưu xuất phát từ hệ thống mây tích (*Cumulus*), bao phủ một khu vực nhỏ, cường độ mưa lớn, thời gian mưa ngắn. Mưa phân tầng thường từ các đám mây tầng (*Stratus*), có cấu trúc đồng nhất theo phương ngang, cường độ mưa thấp và kéo dài [1]. Việc phân loại đặc điểm mưa có ý nghĩa trong nghiên cứu hệ thống mưa với cơ chế vật lý khác nhau nhằm cải thiện ước tính mưa từ quan trắc Ra đa và dự báo mưa từ mô hình số trị [2].

Trên thế giới đã có nhiều các phương pháp khác nhau được xây dựng nhằm phân loại đặc điểm mưa của một hệ thống mây từ các quan trắc trên mặt đất và trên không. Các phương pháp đầu tiên sử dụng số quan trắc từ các máy đo mưa bề mặt khi giá trị mưa vượt qua một ngưỡng cố định–phương pháp BET [3–4], tuy nhiên phương pháp này chỉ giúp xác định được khu vực tâm mưa đối lưu. Trước yêu cầu mở rộng khu vực phân loại, trong 30 năm trở lại đây các phương pháp phân loại sử dụng số liệu Ra đa thời tiết đã được nghiên cứu và phát triển. Đối với các Ra đa phân cực đơn các thuật toán được phát triển chủ yếu dựa trên giá trị từ Phản hồi vô tuyến (*Reflectivity – Z*) và các biến dẫn xuất được đưa ra [5–9]. Tiêu biểu trong một nghiên cứu của Stainer và CCS [10] đã đưa ra phương pháp sử dụng bán kính ảnh hưởng xung quanh mỗi tâm đối lưu biến đổi theo kính thước và là một hàm phụ thuộc vào độ lớn của giá trị Z trung bình nền khu vực (phương pháp SHY95). Bằng việc kiểm tra bằng giả thuyết dải sáng xuất hiện trong lớp từ 4–5 km, phương pháp cho kết quả phân loại sai nhỏ hơn 10%, tác giả cũng đã tiến hành việc ước tính lượng mưa dựa trên kết quả phân loại bằng cách sử dụng hai mối quan hệ Z-R riêng cho từng loại mưa là: $Z = 82R^{1.47}$ (mưa đối lưu) và $Z = 143R^{1.5}$ (mưa phân tầng) kết quả đã cho thấy có sự cải thiện đáng kể độ chính xác về ước tính lượng mưa, tuy nhiên hạn chế chính đối với phương pháp SHY95 đó là việc đôi khi phân loại quá cao với những vùng PHVT thành mưa đối lưu. Đối với Ra đa thời tiết phân cực đã có một số các nghiên cứu thực hiện [11–15], các nghiên cứu này tập trung chính vào việc phân tích đặc điểm của các biến quan trắc phân cực qua đó nhận dạng hai loại đặc điểm mưa dựa trên các thuật toán máy học hiện đại, kết quả so sánh với các phương pháp được xây dựng trước đó cho thấy có thể phân loại được hai loại mưa trong hệ thống mây và cũng rất hữu ích trong việc cải thiện độ chính xác của ước tính mưa.

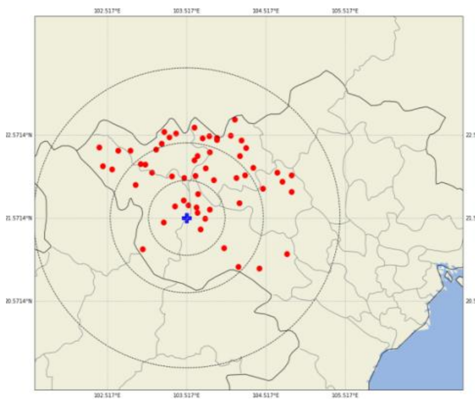
Các nghiên cứu trước đây tại Việt Nam hầu hết chỉ sử dụng số liệu từ các Ra đa phân cực đơn, tập trung chủ yếu vào nhận biết các hiện tương khí tượng dựa trên việc khảo sát, đưa ra các ngưỡng giá trị và xây dựng công thức ước tính mưa dựa trên mối quan hệ giữa Z-R cho từng khu vực nghiên cứu [16–20].

Với sự phát triển của thế hệ các Ra đa thời tiết phân cực thì việc sử dụng số liệu quan trắc phân cực nhằm phân loại đặc điểm mưa đối lưu hay mưa tầng và nâng cao độ chính xác trong ước tính mưa đang được tập trung nghiên cứu. Việc có thêm một lượng lớn số liệu mưa quan trắc được bởi Ra đa nếu sử dụng các thuật toán máy học sẽ góp phần giải quyết được những bài toán phức tạp đặt ra. Do đó nghiên cứu trình bày một phương pháp nhằm hỗ trợ việc phân loại đặc điểm mưa, kết quả phân loại phân loại sau đó được sử dụng nhằm tìm ra các hệ số thực nghiệm trong các công thức ước tính mưa cho khu vực Tây Bắc.

2. Số liệu sử dụng và phương pháp nghiên cứu

2.1. Số liệu sử dụng

Số liệu mưa: Nghiên cứu sử dụng giá trị lượng mưa (mm) quan trắc được từ 55 máy đo mưa tự động thuộc hệ thống Vrain (Hình 1) trong bán kính quan trắc 200 km của Ra đa Pha Đin, độ phân giải thời gian là 10 phút. Số liệu mưa được thu thập từ 68 đợt mưa diễn ra trong khoảng thời gian từ tháng 2 đến tháng 8 trong hai năm 2021 và 2022, trong đó thời gian bắt đầu của một đợt mưa được xác định là thời điểm mà ở đó có ít nhất một nửa số máy đo mưa tự động có báo lượng mưa, thời gian kết thúc khi hơn một nữa số máy đo mưa không còn ghi nhận giá trị lượng mưa.



Hình 1. Vị trí Trạm Ra đa thời tiết Pha Đin (chấm xanh) và 55 máy đo mưa tự động (chấm đỏ).

Số liệu quan trắc phân cực: Nghiên cứu sử dụng số liệu quan trắc từ trạm Ra đa thời tiết Pha Đin (21^{o}34' N, 103^{o}31'E), thu thập trong bán kính 200 km với độ phân giải thời gian là 10 phút và cũng được phân chia theo các đợt mưa để đảm bảo trùng khớp về thời gian với số liệu mưa. Các biến phân cực sử dụng gồm: Phản hồi vô tuyến (Z), Độ lệch phản hồi vô tuyến (*Diferential Reflectivity - ZDR*), Độ lệch pha riêng (*Specific Differential Phase - KDP*), Hệ số tương quan (*Correlation Coefficient - CC*). Sau khi thu thập số liệu trải qua bước thứ nhất là kiểm soát chất lượng nhằm loại bỏ các phản hồi từ nhiễu bề mặt, tiếp theo số liệu sẽ được nội suy trên lưới Đề-các 3 chiều tại 7 mực độ cao (0 tới 9 km) và độ cao 3 km sẽ được lựa chọn sử dụng để tránh ảnh hưởng của dải sáng.

2.2. Phương pháp nghiên cứu

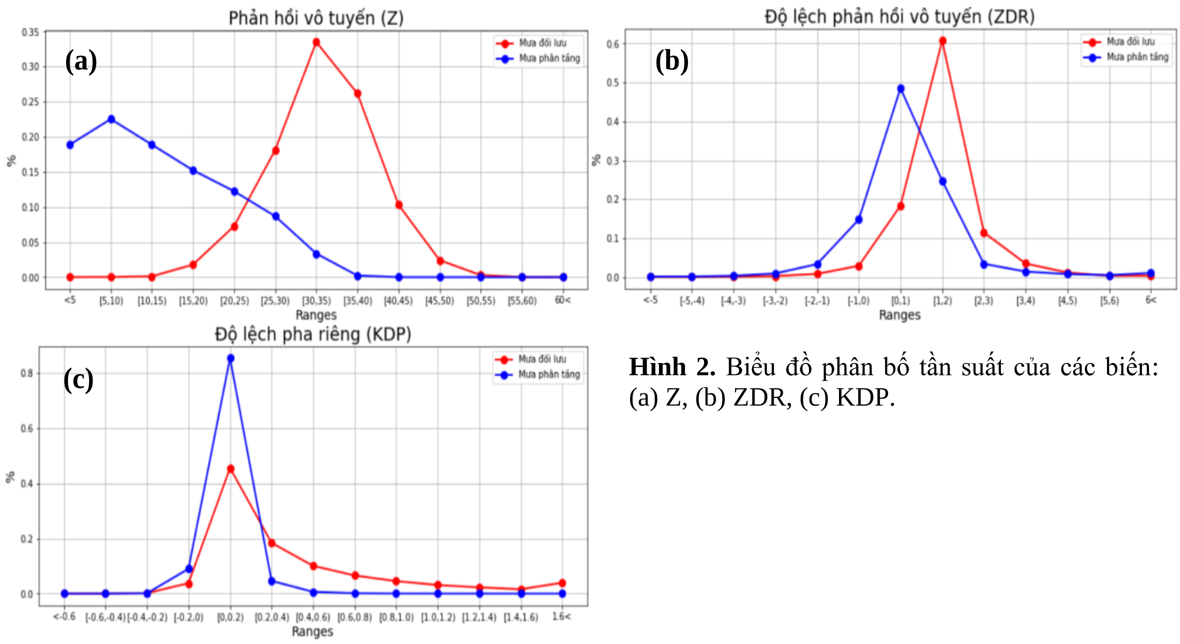
2.2.1. Phương pháp phân loại đặc điểm mưa

Việc sử dụng các thuật toán trí thông minh nhân tạo (AI) đối với số liệu Ra đa thời tiết đã được phát triển mạnh trong hai thập kỉ qua. Dựa trên việc tham khảo kết quả từ các nghiên cứu trước đó, tác giả lựa chọn sử dụng thuật toán Máy vecto hỗ trợ (*Support Vector Machine*) [21–22] nhằm mục đích xây dựng một phương pháp phân loại đặc điểm mưa (phương pháp SVM). Với một bộ các thuộc tính (điểm) thuộc một trong hai lớp cho trước (tập huấn luyện), thuật toán SVM sẽ xây dựng một siêu phẳng (hyperplane) nhằm phân loại các điểm thành hai phần sao cho các điểm cùng một lớp nằm về một phía với siêu phẳng, đồng thời thỏa mãn điều kiện lề phân tách của hai đối tượng cần phân tách đạt giá trị cực đại. Cụ thể bằng cách đưa vào một tập số liệu huấn luyện $\{(X_i, \mathrm{y_i})\}_{\mathrm{i=1}}^{\mathrm{N}}$, mục tiêu sẽ phải tìm Vectơ trọng số tối ưu W và độ lệch b sao cho:

$$\left.\begin{array}{c}\min \frac{<\mathrm{W.W}>}{2} \\ \text{Với điều kiện: } \mathrm{y_i}(\mathrm{WX_i} + \mathrm{b}) \geq 1 \;\; \forall \mathrm{i} = 1,2,\ldots,\mathrm{N},\end{array}\right\} \quad (1)$$

Trong đó $X_i \in R^m$ là vector đầu vào (m là số thuộc tính của hai đối tượng cần phân tách - trong nghiên cứu sử dụng m = 2), N là số mẫu sử dụng để huấn luyện, y_i là đầu ra của thuật toán với giá trị +1 hoặc là -1 (trong nghiên cứu tương ứng với mưa đối lưu hoặc phân tầng). Các điểm với cặp số liệu đặc biệt (X_i, y_i) nếu thỏa mãn điều kiện trong (1) với dấu bằng thì các điểm này được gọi các Vectơ hỗ trợ.

Với mục đích xây dựng và đánh giá phương pháp SVM, nghiên cứu lựa chọn sử dụng phương pháp SHY95 [10] làm phương pháp tham khảo, phương pháp có tính ứng dụng cao bởi việc áp dụng chỉ cần duy nhất số liệu từ Z. Tuy nhiên, để tăng độ chính xác thì thuật toán SVM yêu cầu giá trị của các biến phân cực sử dụng phải có sự khác nhau đáng kể trong hai loại mưa [9, 15], do đó sự biến đổi giá trị tần suất của các biến phân cực tại các khoảng giá trị được xác định và so sánh. Việc khảo sát thực hiện thông qua 18 đợt mưa được lựa chọn trong năm 2021, qua đồ thị tần suất của của các biến phân cực (Hình 2) cho thấy chỉ có giá trị của Z và ZDR có sự khác biệt trong hai loại mưa và do đó có thể được sử dụng trong việc phân loại đặc điểm mưa.



Hình 2. Biểu đồ phân bố tần suất của các biến: (a) Z, (b) ZDR, (c) KDP.

Để sử dụng thuật toán SVM thì W và b trong công thức (1) cần phải được giải thông qua việc sử dụng một tập số liệu huấn luyện. Sau khi xem xét phân bố tần suất nghiên cứu đã lựa chọn được các ngưỡng giá trị để xây dựng tập huấn luyện như sau: với mưa đối lưu giá trị được lựa chọn theo tiêu chí Z > 25dBZ và ZDR > 0dB, đối với mưa phân tầng là Z < 35dBZ và ZDR < 2dB, cuối cùng trong cả hai loại số liệu huấn luyện chỉ những vị trí điểm lưới nào được xác nhận là mưa loại phân tầng hoặc đối lưu bởi phương pháp SHY95 mới được sử dụng để huấn luyện thuật toán SVM.

Tập huấn luyện thuật toán có 236.442 bộ số liệu (226.156 bộ số liệu mưa phân tầng và 10 285 bộ số liệu mưa đối lưu) được sử dụng. Một bộ số liệu được xác định gồm số liệu từ Z và ZDR từ một điểm lưới bất kì thỏa mãn tiêu chí lựa chọn số liệu và một biến E biểu diễn loại mưa (nhãn phân loại) tương ứng đã được xác định trước (với +1/-1 tương ứng là mưa đối lưu và phân tầng). Cụ thể hơn một bộ số liệu huấn luyện là một Vectơ có dạng [Z(i,j); ZDR (i,j); E(i,j);] với i và j lần lượt là vị trí kinh vĩ độ của điểm lưới.

2.2.2. Phương pháp ước tính mưa

Nghiên cứu tiến hành xây dựng các công thức ước tính mưa bằng việc sử dụng kết hợp giữa 3 nguồn số liệu gồm: số liệu mưa quan trắc từ máy do mưa, số liệu từ quan trắc phân cực và kết quả từ phương pháp SVM. Dựa trên việc kết hợp của một hay nhiều biến phân cực với số liệu quan trắc mưa tự động, các dạng công thức được nghiên cứu sử dụng lượt R (Z, ZDR, KDP), R (Z, ZDR), R (Z) và R (KDP) và được viết dưới dạng tổng quát:

$$R = aZ^{b}ZDR^{c}KDP^{d} \tag{2}$$

Trong đó R là lượng mưa ước tính từ Ra đa; Z, ZDR, KDP là các biến phân cực a, b, c, d là các hệ số thực nghiệm.

Tùy vào dạng công thức khác nhau mà các biến Z, ZDR, KDP có thể xuất hiện trong công thức. Việc có sử dụng KDP trong công thức dựa trên cơ sở của các nghiên cứu đã chỉ cho rằng KDP không phải là một phép đo năng lượng sóng điện từ mà là một phép đo dựa trên pha dao động, vì vậy KDP ít bị ảnh hưởng bởi sự suy giảm tín hiệu, nhiễu địa hình vv…do đó việc sử dụng KDP trong các công thức có thể làm tăng sự chính xác trong các công thức ước tính mưa. Để thu được các công thức ước tính mưa từ (2), nghiên cứu sử dụng phương pháp Bình phương tối thiểu [18–20] nhằm xác định các hệ số thực nghiệm thông qua việc sử dụng tập số liệu của 27 đợt mưa được thu thập.

2.2.3. Phương pháp đánh giá

Phương pháp SVM được đánh giá dựa trên phương pháp SHY95 bằng việc xem xét sự khác nhau về phân bố xác định khu vực mưa bởi hai phương pháp, trong đó một khu vực mưa là phân tầng hoặc đối lưu được xác định nhờ vào điểm lưới. Nghiên cứu cũng đánh giá độc lập hai phương pháp thông qua việc tính Tỉ lệ phần trăm phân loại sai để thu được sai số trong từng phương pháp, cho biết tỉ lệ điểm lưới được phương pháp phân loại là mưa đối lưu nhưng lại thể hiện tín hiệu dải sáng so với tổng số điểm lưới có sự xuất hiện của dải sáng trong cấu trúc của Z theo chiều thẳng đứng tại điểm lưới đó [10].

Nhằm đánh giá ước tính mưa một số chỉ số đánh giá thống kê được sử dụng bao gồm:

- Tỉ lệ phát hiện mưa (RD): $RD = \frac{n_{R,G}>0.2}{n_G>0.2}$
- Hệ số tương quan (r): $r = \frac{\sum_{i=1}^{N}(R_i-\bar{R})(G_i-\bar{G})}{\sqrt{\sum_{i=1}^{N}(R_i-\bar{R})^2\sum_{i=1}^{N}(G_i-\bar{G})^2}}$
- Sai số quân phương (RMSE): $RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(R_i - G_i)^2}$
- Sai số hệ thống (ME): $ME = \frac{1}{N}\sum_{i=1}^{N}(R_i - G_i)$

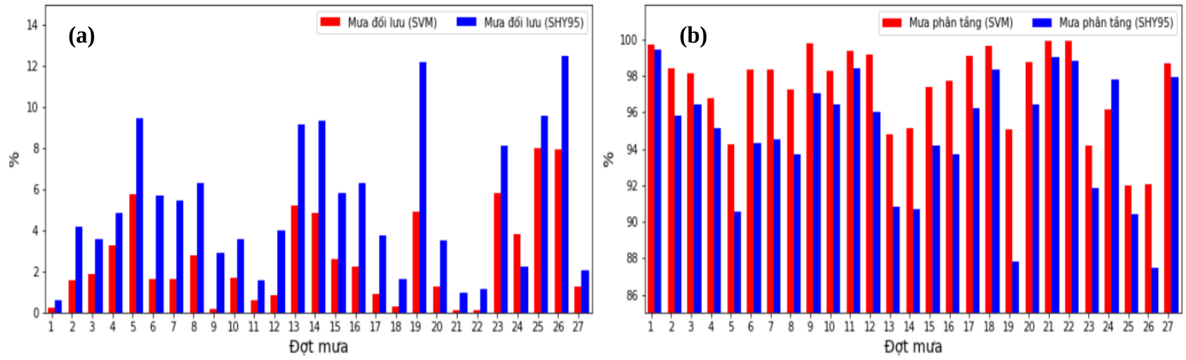
Trong đó $n_{R,G}$ là số cặp giá trị mưa mà ở đó lượng mưa ước tính từ Ra đa và máy đo mưa cùng vượt ngưỡng giá trị 0,2 mm; n_G là số quan trắc từ máy đo mưa vượt quá ngưỡng giá trị 0,2 mm.

R_i và G_i là giá trị ước tính mưa từ Ra đa và giá trị quan trắc mưa từ máy đo mưa (mm); N là dung lượng mẫu. Giá trị R_i được lấy tại độ cao 3 km trung bình trong một miền tròn bán kính 2 km có tâm trùng với vị trí của máy đo mưa.

3. Kết quả và thảo luận

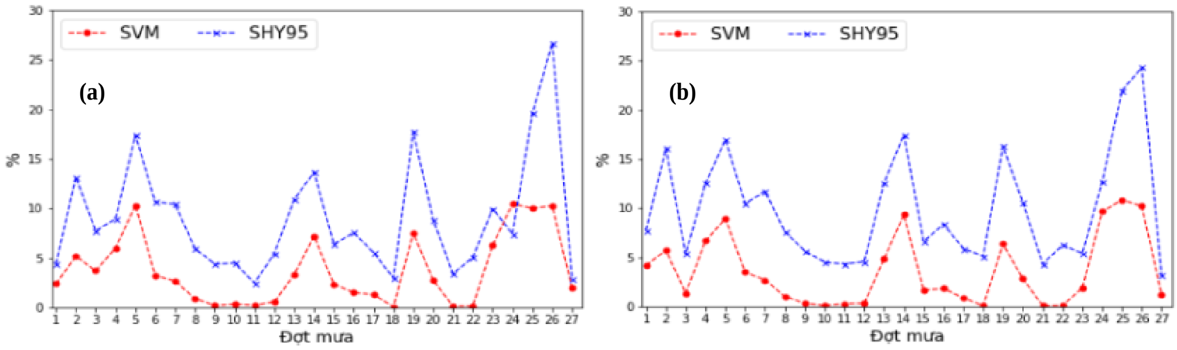
3.1. Kết quả phân loại đặc điểm mưa

Nhằm đánh giá phương pháp được đưa ra trong nghiên cứu, kết quả phân loại đặc điểm mưa từ phương pháp SVM bước đầu được so sánh với phương pháp SHY95 dựa trên việc xem xét tỉ lệ phân bố khu vực mưa và phần trăm phân loại sai của cả hai phương pháp.



Hình 3. Tỉ lệ phân bố của mưa (a) đối lưu và (b) phân tầng đối với hai phương pháp phân loại đặc điểm mưa SVM (màu đỏ) và SHY95(màu xanh) qua các đợt mưa.

Qua tỉ lệ phân bố khu vực điểm lưới được xác định là mưa phân tầng hoặc đối lưu bởi hai phương pháp (Hình 3) có thể thấy cả hai đều đã xác định loại mưa chủ yếu là phân tầng với tỉ lệ trung bình từ 87,1-99,2%, mưa đối lưu chỉ chiếm tỉ lệ từ 0,1-13,7%. Với mưa đối lưu (Hình 3a) phương pháp SVM có xu hướng phân loại khu vực mưa thấp hơn so với phương pháp SHY95, với phương pháp SVM tỉ lệ dao động chỉ từ 0,1-8,0% trong khi đó với SHY95 là từ 0,5-13% khu vực mưa. Đối với mưa phân tầng (Hình 3b) phương pháp SVM lại có xu hướng phân loại cao hơn so với phương pháp tham khảo, tỉ lệ xác định mưa phân tầng với phương pháp SVM dao động từ 92,1-99,9 % trong khi đó SHY95 là từ 87,4-99,4%.



Hình 4. Tỉ lệ phần trăm phân loại sai bởi phương pháp SVM và SHY95 qua từng đợt mưa với mỗi điều kiện xác định độ lớn dải sáng 2dBZ (a) và 5dBZ (b).

Nhằm đánh giá giá độc lập hai phương pháp, tỉ lệ phần trăm phân loại sai sẽ được xem xét (Hình 4). Với cả 2 điều kiện độ lớn nhằm xác định khu vực có sự xuất hiện của dải sáng nhìn chung phần trăm phân loại sai đối với phương pháp SVM thấp hơn so với phương pháp SHY95. Cụ thể phương pháp SVM có tỉ lệ chỉ dao động trong khoảng từ 0-10%, với phương pháp SHY95 tỉ lệ này dao động trong khoảng từ 2-25%. Xét trên toàn bộ tập số liệu đánh giá (Bảng 1), với điều kiện độ lớn dải sáng lần lượt là > 2dBZ và > 5dBZ trung bình phương pháp SHY95 có lần lượt 8,1% và 9,1% khu vực mưa phân tầng thực tế đã bị phân loại sai thành đối lưu. So với phương pháp SHY95 trong cả hai điều kiện thì phân loại sai đối với phương pháp SVM giảm hơn một nửa chỉ còn lần lượt 3% và 2,9%.

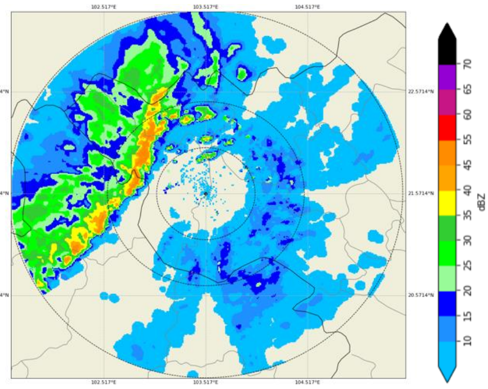
Bảng 1. Trung bình của tỉ lệ phần trăm phân loại sai bởi SHY95 và SVM.

Phương pháp	**Độ lớn dải sáng**	
	> 2 dBZ	**> 5 dBZ**
SHY95	8,1 %	9,1 %
SVM	3,0 %	2,9 %

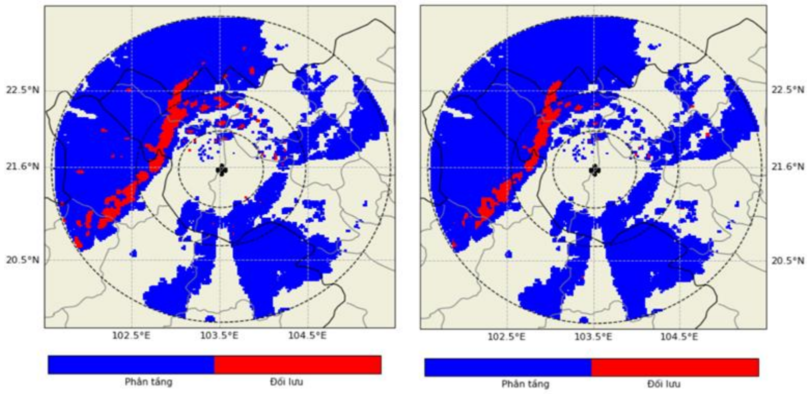
Trong nghiên cứu [8] khi đánh giá hai trường hợp mưa là: Đường tố và Mưa đối lưu trong khu vực mưa phân tầng, dựa trên xem xét số liệu 3 chiều từ Z cho thấy phương pháp SHY95 đôi khi đã phân loại sai những khu vực có mưa phân tầng với cường độ mạnh thành những ổ mưa đối lưu cục bộ làm xuất hiện những khu vực mưa đối lưu bên trong khu vực mưa phân tầng diện rộng, những khu vực mưa đối lưu phân loại sai này có vị trí biến đổi lớn, thời gian tồn tại ngắn khi xem xét liên tục theo thời gian. Để xem xét vấn đề này cũng như đánh giá hiệu quả phương pháp SVM cần thiết phải tiến hành những phân tích độc lập trong những trường hợp mưa thực tế khác nhau, do đó 03 trường hợp mưa bao gồm: Đường tố, Mưa phân tầng diện rộng và Mưa đối lưu trong khu vực phân tầng đã được nghiên cứu lựa chọn xem xét.

a) Trường hợp đường tố

Các đám mây đối lưu đôi khi xếp thành dải kết thành một màn mây gần như liên tục rộng khoảng từ 10-50 km, dài vài trăm km dọc theo đường front lạnh, chuyển động theo hướng ổn định hay gây nên gió giật. Dải này được gọi là đường gió giật hay đường tố. Đường tố hình thành do không khí lạnh di chuyển nhanh đã đẩy không khí nóng ẩm ở phía trước front lên cao, tạo ra các ổ mây đối lưu sâu (thường là mây Cb hình đe) dính liền với nhau thành một dải, trên màn hình hiển thị của Ra đa đường tố thường có hình dạng cánh cung [1]. Hình 5 biểu diễn một trường hợp xảy ra trên khu vực quan trắc bởi Ra đa Pha Đin vào hồi 18:10' (UTC) ngày 5/5/2021, đường tố ban đầu hình thành phía Tây Bắc của Ra đa sau đó mở rộng và dịch chuyển theo hướng Đông Nam.



Hình 5. Sản phẩm MAX của Z từ Ra đa Pha Đin trong trường hợp đường tố lúc 18:10'(UTC) ngày 5/5/2021.



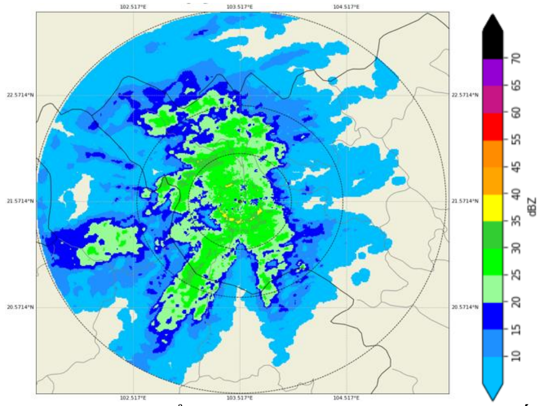
a) b)

Hình 6. Kết quả phân loại cho trường hợp đường tố quan trắc bởi Ra đa Pha Đin lúc 18:10'(UTC) ngày 5/5/202 từ a) phương pháp SHY95, b) phương pháp SVM.

Kết quả phân loại đặc điểm mưa (Hình 6) cho thấy thấy hai phương pháp đã xác định mưa trong khu vực có dạng dải hình cánh cung (gồm tập hợp các ô đối lưu nối liền nhau) là mưa đối lưu. Tuy nhiên với phân loại mưa trong khu vực nằm phía sau đường tố đã có sự khác biệt khi phương pháp SHY95 (Hình 6a) đã mở rộng khu vực phân loại là mưa đối lưu với sự xuất hiện của các ổ mưa đối lưu có kích thước nhỏ nằm riêng lẻ bên trong khu vực mưa phân tầng, trong khi đó phương pháp SVM (Hình 6b) phân loại hoàn toàn khu vực này là mưa phân tầng. Theo [23] sau khi nghiên cứu đã đưa ra kết luận mưa phía sau các ổ đối lưu trong trường hợp đường tố có thể được phân loại là phân tầng bởi chuyển động không khí thẳng đứng trong khu vực này yếu hơn tốc độ rơi của các hạt mưa, điều này thỏa mãn điều kiện khu vực có mưa phân tầng.

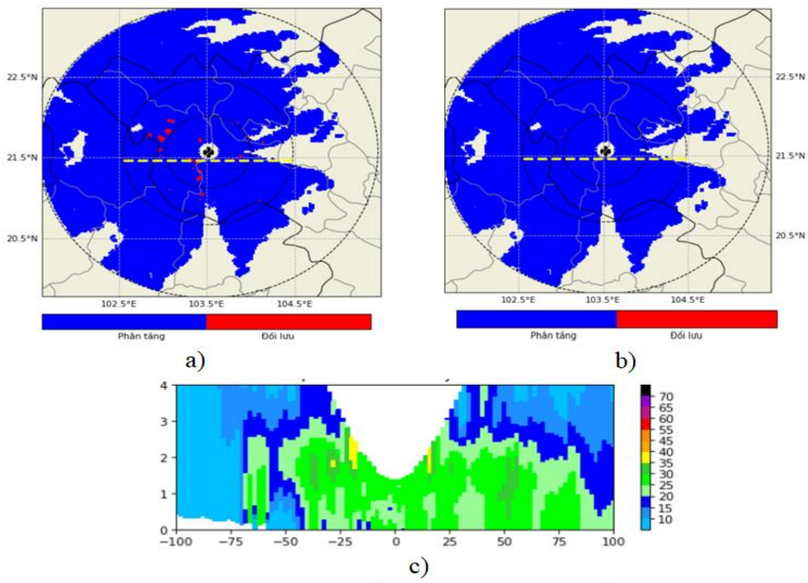
b) Trường hợp mưa phân tầng diện rộng

Trong trường hợp này toàn bộ khu vực quan trắc bởi Ra đa được bao phủ gần như hoàn toàn bởi mưa loại phân tầng, với đặc điểm nhận dạng khi giá trị Z nhỏ cùng với đó là việc có hoặc không có sự xuất hiện của dải sáng đặc trưng. Hình 7 thể hiện một trường hợp mưa diễn ra vào hồi 11:40'(UTC) ngày 8/2/2021, giá trị Z cực đại ≤ 40dBz trên sản phẩm MAX, có thể quan sát thấy được sự xuất hiện của dải sáng hình nhẫn với giá trị Z ~40dBZ, cực đại xung quanh phạm vi bán kính <50km của Ra đa, việc có sự xuất hiện của dải sáng cho thấy đây là một trường hợp mưa phân tầng đặc trưng [2, 24].



Hình 7. Sản phẩm MAX của Z từ Ra đa thời tiết Pha Đin trong trường hợp mưa phân tầng diện rộng lúc 11:40'(UTC) ngày 8/2/2021.

Phương pháp SHY95 (Hình 8a) trong trường hợp này đã phân loại sai khi phân loại mưa đối lưu (các ổ đối lưu nhỏ) tại đúng vị trí của dải sáng, điều này đã tạo ra các ổ nhiễu đối lưu nằm bên trong khu vực mưa phân tầng và làm tăng khu vực mưa đối lưu. Ngược lại, phương pháp SVM (Hình 8b) đã phân loại hoàn toàn mưa phân tầng cho khu vực, với dải sáng xung quanh vị trí của Ra đa cũng được phân loại là mưa phân tầng. Để kiểm tra nghiên cứu sử dụng một lát cắt theo vĩ độ $21^{0}48$'N (Hình 8c) đi qua khu vực được phân loại khác nhau giữa hai phương pháp, qua lát cắt có thể thấy khu vực hai phương pháp phân loại khác nhau thể hiện một cấu trúc PHVT phân tầng với cực đại Z nằm ở ~2km, phần bên trên và dưới cực đại này Z đều giảm. Như vậy việc phân loại khu vực này là mưa phân tầng trong phương pháp SVM là hoàn toàn phù hợp.

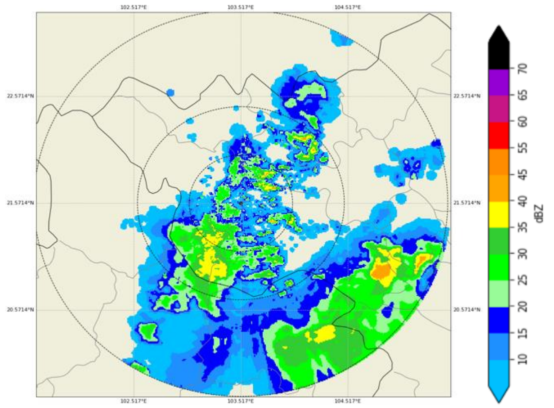


Hình 8. Kết quả phân loại cho trường hợp Mưa phân tầng diện rộng quan trắc từ Ra đa Pha Đin lúc 11:40'(UTC) 8/2/2021 từ: (a) phương pháp SHY95, (b) phương pháp SVM. Đường màu vàng biểu diễn vị trí lát cắt theo vĩ độ $21^{0}48$'N, (c) qua khối quét của Ra đa.

c) Trường hợp mưa đối lưu trong khu vực phân tầng

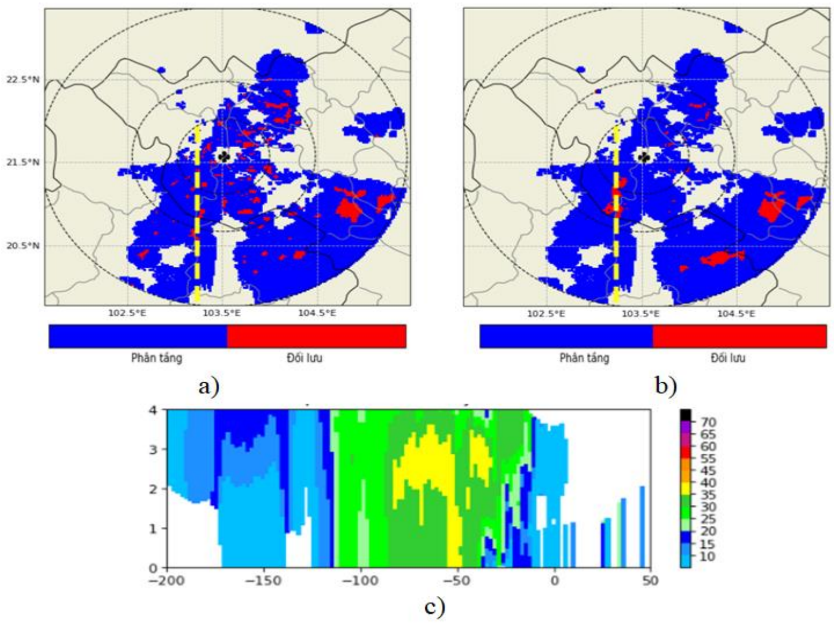
Khác với trường hợp đường tố đã phân tích, đặc trưng của trường hợp mưa loại này đó là khu vực được bao phủ hầu hết bởi mưa phân tầng, tuy nhiên bên trong khu vực mưa phân tầng sẽ có sự xuất hiện của một số ổ mưa đối lưu mới được hình thành hoặc đang trong quá trình tan rã nằm riêng rẽ.

Trên khu vực Tây Bắc qua số liệu thống kê quan trắc Ra đa trong hai năm nghiên cứu cho thấy trường hợp mưa này chiếm phần lớn các trường hợp mưa quan trắc từ Ra đa thời tiết trên khu vực. Hình 9 thể hiện một đợt mưa diễn ra vào lúc 19:20’(UTC) ngày 7/7/2021 từ sản phẩm MAX. Có thể thấy sự xuất hiện của các ổ mưa đối lưu nằm riêng lẻ ở phía Đông Bắc và phía Nam của Ra đa với giá trị PHVT cực đại > 50dBZ bao quanh các trung tâm này là khu vực mưa với giá trị < 35dBZ.



Hình 9. Sản phẩm MAX của Z từ Ra đa Pha Đin trong trường hợp mưa đối lưu trong khu vực phân tầng lúc 19:20’(UTC) ngày 7/7/2021.

Kết quả sử dụng phương pháp phân loại đặc điểm mưa SVM cho trường hợp mưa này được biểu diễn trên Hình 10b qua quan sát có thể thấy phương pháp đã xác định được một số ổ đối lưu lớn nằm ở phía Đông Bắc và phía Nam của Ra đa, phía Đông Bắc cũng suất hiện các ổ đối lưu tuy nhiên có kích thước nhỏ hơn. Các ổ đối lưu này có vị trí trùng với vị trí Z cực đại trên sản phẩm MAX, các khu vực còn lại xung quanh các ổ đối lưu này được phân loại là mưa phân tầng. Trong khi đó phương pháp SHY95 (Hình 10a) phân loại có sự khác biệt khi tạo ra rất nhiều các ổ đối lưu nằm riêng lẻ, có kích thước nhỏ nằm bên trong khu vực được xác định là mưa phân tầng, qua việc kiểm tra trên tập số liệu của cả đợt cho thấy các ổ này tồn tại không có sự liên tục theo thời gian.



Hình 10. Kết quả phân loại cho trường hợp Mưa đối lưu trong khu vực phân tầng quan trắc từ Ra đa Pha Đin lúc 19:20’(UTC) ngày 7/7/ 2021 từ a) phương pháp SHY95, b) phương pháp SVM. Đường màu vàng biểu diễn vị trí lát cắt theo kinh độ $103^{o}23’$E c) qua khối quét của Ra đa.

Sử dụng một lát cắt theo kinh độ $103^{o}23’$E đi qua khu vực có sự phân loại khác nhau giữa hai phương pháp (Hình 10c), có thể thấy khu vực lát cắt đi qua có duy nhất một vị trí có giá trị cực đại của Z dạng cột thẳng đứng thể hiện mưa đối lưu ở vị trí phía Nam Tây Nam cách Ra đa một khoảng 50-100 km, đây là khu vực đối lưu duy nhất các khu vực còn lại đều có giá trị của Z thấp, như vậy phương pháp SVM đã phân loại chính xác, trong khi đó phương pháp SHY95 cũng xác định được vị trí mưa đối lưu nhưng cùng với đó cũng xuất hiện những ổ đối lưu nhỏ khác.

3.2. Kết quả ước tính mưa

Một số các nghiên cứu trước đây đã chỉ ra rằng việc phân loại mưa cũng như sử dụng số liệu phân cực có thể làm tăng độ chính xác của lượng mưa ước tính từ Ra đa thời tiết. Bằng việc sử dụng kết hợp giữa số liệu từ máy đo mưa tự động, quan trắc Ra đa phân cực và kết quả từ phương pháp SVM. Nghiên cứu tiến hành thử nghiệm xây dựng hệ số thực nghiệm trong các công thức ứng với hai loại đặc điểm mưa. Ngoài ra nhằm xem xét hiệu quả phân loại, nghiên cứu cũng sử dụng một công thức ước tính không sử dụng kết quả phân loại SVM (B1) và một công thức được tham khảo của Marshall (B2). Sau khi tính toán hệ số trong các công thức được thể hiện trong Bảng 2.

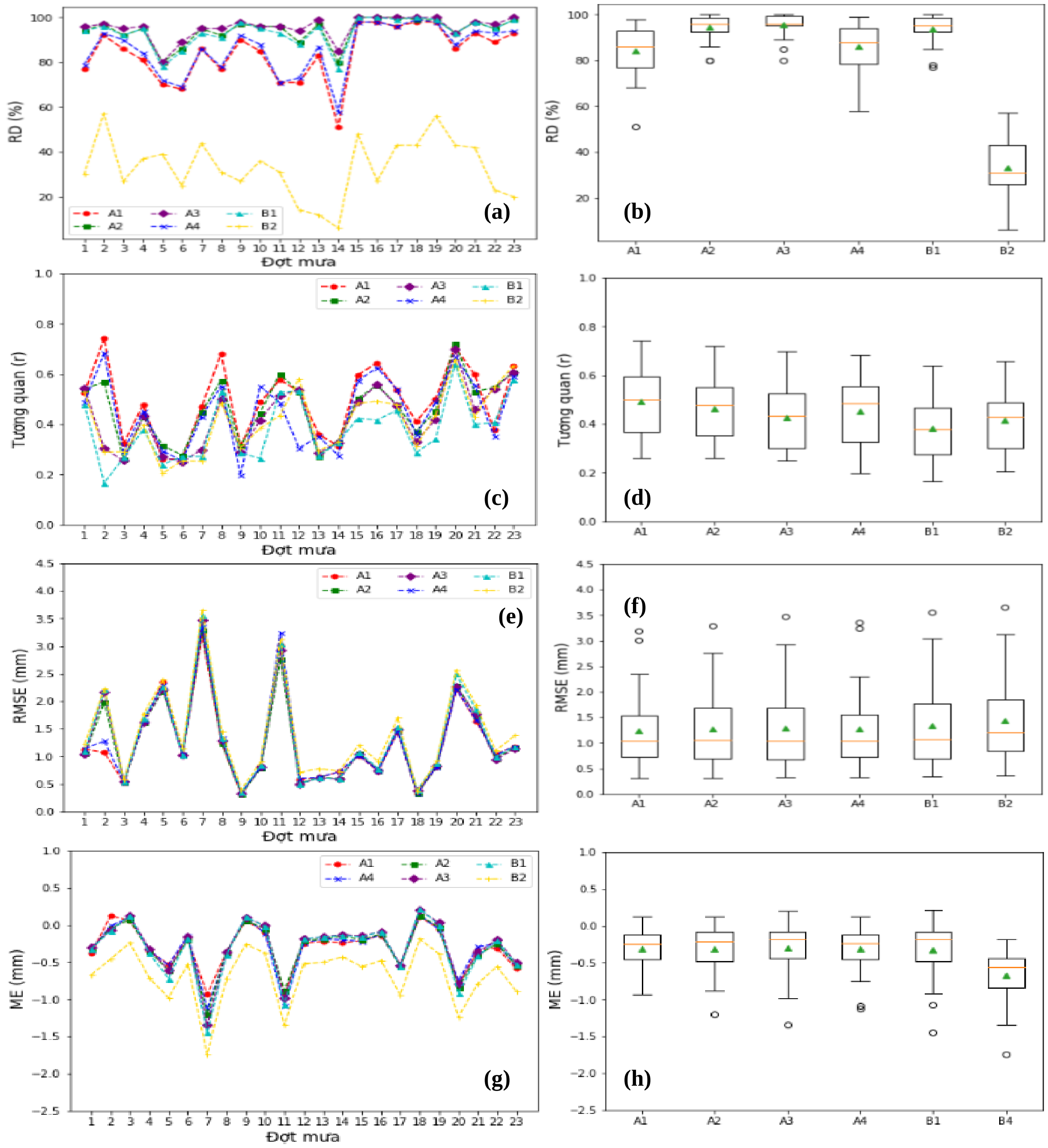
Bảng 2. Hệ số trong các công thức ước tính mưa.

STT	Kí hiệu	Công thức	Đặc điểm mưa	a	b	c	d
1	A1	$R = aZ^bZDR^cKDP^d$	Phân tầng	2,11	0,20	-0,86	0,19
			Đối lưu	64,50	-0,03	-2,25	0,50
2	A2	$R = aZ^bZDR^c$	Phân tầng	1,19	0,23	-0,77	
			Đối lưu	3,03	0,28	-2,10	
3	A3	$R = aZ^b$	Phân tầng	1,15	0,21		
			Đối lưu	3,05	0,19		
4	A4	$R = aKDP^b$	Phân tầng	7,34	0,38		
			Đối lưu	21,86	0,47		
5	B1	$R = aZ^b$	Không phân loại	1,02	0,24		
6	B2	$R = aZ^b$ (Marshall)	Không phân loại	0,0365	0,625		

Với mục đích đánh giá chất lượng của các công thức ước tính mưa, nghiên cứu đã lựa chọn 23 đợt mưa độc lập với các đợt mưa đã được sử dụng để xây dựng các công thức ước tính lượng mưa để đánh giá. Kết quả đánh giá được biểu diễn trong Hình 11.

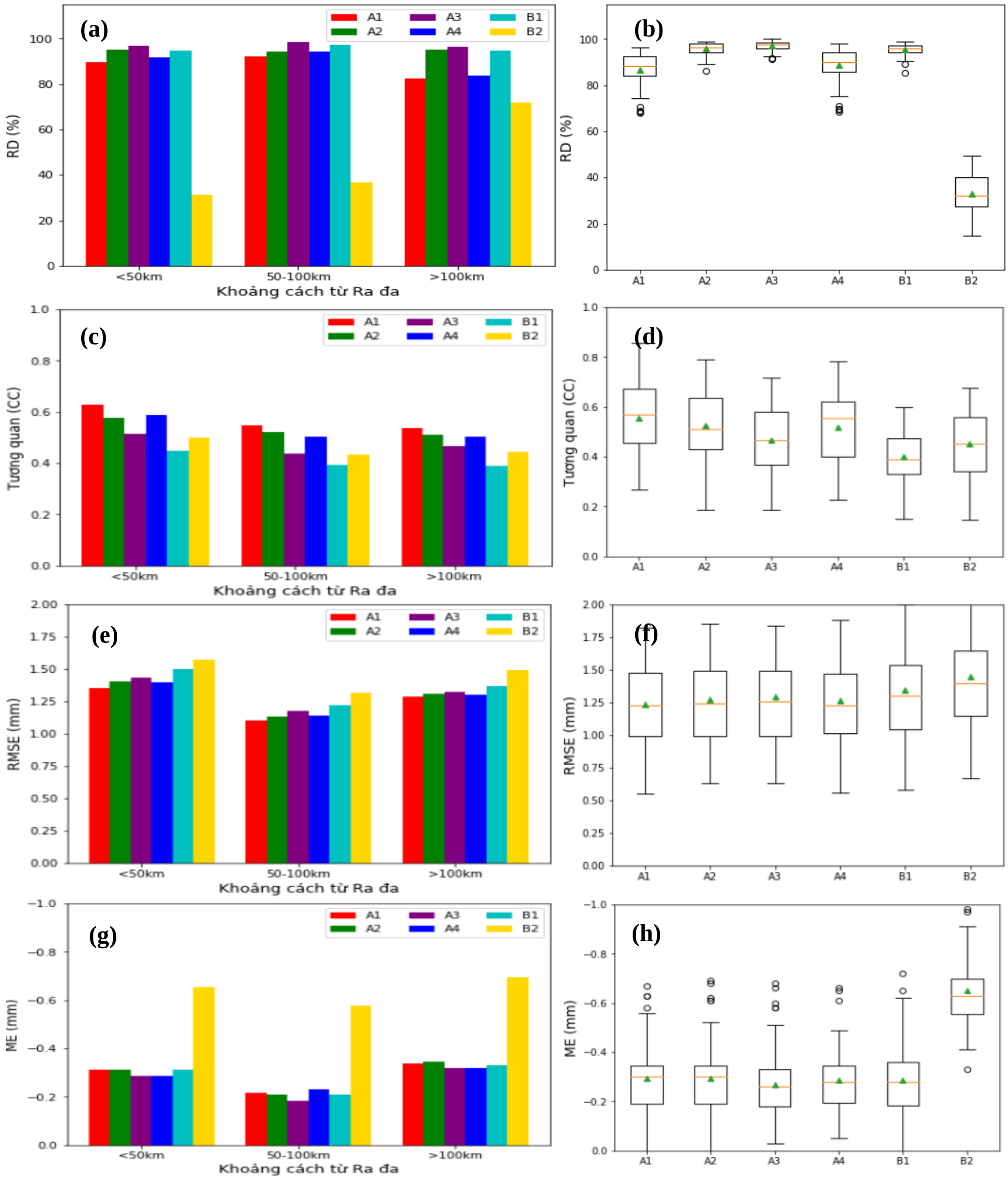
Nhìn chung tất cả các công thức đều cho RD biến đổi qua từng đợt mưa trong khoảng 65-100%, ngoại trừ công thức B2 có RD thấp nhất từ 10-56% (Hình 11a), công thức với Z và ZDR (A2, A3 và B1) có RD cao hơn công thức với KDP (A1 và A4), biểu đồ hộp Hình 11b cũng chỉ ra RD ~95% đối với A2, A3 cùng với đó độ trải giữa (IQR) thấp nhất. Về tương quan giữa lượng mưa ước tính từ Ra đa và máy đo mưa tự động (Hình 11c), công thức với ZDR và KDP (A1, A2, A4) có r cao hơn so với chỉ sử dụng Z (A3, B1, B2) trong đó A1 có r cao nhất trong khoảng từ 0,30-0,74.

Việc sử dụng kết quả phân loại từ phương pháp SVM có tương quan cao hơn so với không phân loại mưa, biểu đồ hộp Hình 11d chỉ ra A1 có r cao nhất và thấp nhất là B1. Hình 11e chỉ ra RMSE giữa các cặp giá trị lượng mưa giữa Ra đa và máy đo mưa trong đó RMSE có sự biến đổi theo từng đợt mưa khác nhau, sai số cao đối với các đợt mưa có cường độ mạnh (đợt 7 và 11) và tương đối nhỏ với các đợt mưa có cường độ nhỏ (đợt 9 và 18), qua biểu đồ hộp Hình 11f có thể thấy A1 có sai số nhỏ nhất ~1,25 mm, B1 và B2 có sai số lớn nhất lần lợt là 1,44 mm và 1,34 mm. Như vậy RMSE cũng có cùng xu thế như đối với r. Hình 11g và 11h biểu diễn phân bố của giá trị ME, có thể thấy hầu hết đều cho kết quả ước tính thấp hơn so với giá trị quan trắc ở tất cả các đợt. Nguyên nhân dẫn tới việc ước tính thấp có thể bởi các quan trắc từ Ra đa bị ảnh hưởng bởi cả sai số ngẫu nhiên lẫn sai số hệ thống gồm: sự suy giảm tín hiệu trong mưa, sự che khuất bởi tín hiệu và việc hiệu chỉnh Ra đa,…



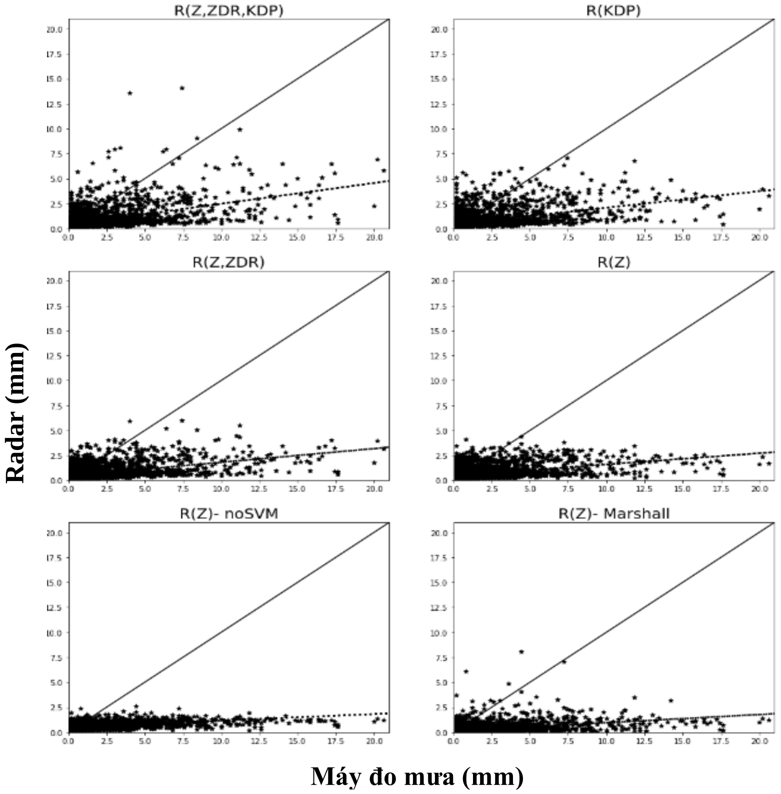
Hình 11. Giá trị RD (a), (b); r (c), (d); RMSE (e), (f); ME (g), (h) giữa lượng mưa ước tính từ Ra đa và quan trắc từ máy đo mưa tự động cho từng đợt mưa.

Chỉ số đánh giá đối với mỗi công thức tại những khoảng cách khác nhau so với vị trí Ra đa cũng được xem xét đánh giá. Hình 12a biểu diễn giá trị RD tại các khoảng cách trong phạm vi hoạt động của Ra đa, ngoại trừ công thức B2 còn lại đều có RD cao trên 85% và có ít sự thay đổi. Biểu đồ hộp Hình 12b cho thấy A2, A3, B1 đạt giá trị cao với RD trung bình ~95% cùng với đó là IQR thấp. Hình 12c chỉ ra giá trị r giữa lượng mưa ước tính từ Ra đa và máy đo mưa, nhìn chung r giảm khi khoảng cách giữa máy đo mưa so với vị trí Ra đa tăng lên, điều này có thể do tín hiệu Ra đa nhận được bị suy giảm theo khoảng cách. Công thức ước tính mưa sử dụng kết hợp việc phân loại đặc điểm và các biến phân cực có r cao hơn so với khi không có sự phân loại. Biểu đồ hộp Hình 12d cho thấy A1 có r cao nhất ~0,66%, trong khi đó B1 có r trung bình thấp nhất ~0,38%. Giá trị ME được thể hiện trong Hình 12g,h. Qua hình 12g có thể thấy tất cả các công thức đều cho giá trị lượng mưa ước tính từ Ra đa thấp hơn so với giá trị quan trắc dao động trong khoảng từ 20-40%, thậm chí B2 cho độ lệch tới hơn 70%. Biểu đồ hộp hình 12h cho thấy không có sự chênh lệnh của giá trị ME giữa các công thức sử dụng biến phân cực, phân loại và không phân loại mưa.



Hình 12. Giá trị RD (a), (b); r (c), (d); RMSE (e), (f); ME (g), (h) giữa lượng mưa ước tính từ Ra đa và quan trắc từ máy đo mưa tự động tại từng khoảng cách.

Biểu đồ phân tán giữa các cặp giá trị lượng mưa ước tính từ Ra đa (R) và lượng mưa quan trắc từ máy đo mưa (G) với mỗi công thức ước tính mưa được thể hiện trong Hình 13. Trong hình các đường nét đứt biểu diễn đường hồi quy tuyến tính dưới dạng R = aG+b của các cặp R-G qua tất cả các đợt mưa, có thể thấy đường hồi quy của tất cả các công thức mưa đều nằm bên dưới đường một - một cho thấy tất cả đều cho kết quả ước tính lượng mưa thấp hơn giá trị quan trắc từ máy đo mưa tự động. Công thức kết hợp giữa việc phân loại đặc điểm mưa và sử dụng các biến phân cực có kết quả tốt hơn so với việc không phân loại và chỉ sử dụng Z (gần với đường một - một nhất). Trong đó công thức R (Z, ZDR, KDP) cho kết quả tốt nhất, với lượng mưa có giá trị ≤ 5 mm cho kết quả tương đối gần sát với giá trị quan trắc, ngược lại với lượng mưa trên > 5 mm thường cho kết quả thấp hơn giá trị quan trắc.



Hình 13. Biểu đồ phân tán giữa lượng mưa ước tính từ Ra đa và quan trắc từ máy đo mưa đối với từng công thức mưa.

4. Kết luận

Nghiên cứu đã tập trung phân loại mưa bằng quan trắc radar phân cực cho các hệ thống mưa trên khu vực tây bắc bằng các áp dụng thuật toán học máy SVM. Kết quả cho thấy phương pháp SVM có xu hướng phân loại thấp khu vực mưa đối lưu và phân loại cao đối với khu vực mưa phân tầng hơn so với phương pháp tham khảo SHY95. Bằng việc sử dụng chỉ tiêu dải cho thấy phương pháp SVM đã hạn chế được sai số, khi chỉ dưới 10% khu vực có dải sáng bị xác định sai là mưa đối lưu. Với 03 trường hợp mưa được xem xét, phương pháp SVM đã khắc phục được một số các khu vực đã bị phân loại sai bởi phương pháp SHY95 như: mưa phía sau đường tố, vị trí dải sáng xung quanh Ra đa bị xác định là mưa đối lưu.

Đối với ước tính lượng mưa qua đánh giá cho thấy lượng mưa ước tính từ Ra đa có xu hướng thấp hơn giá trị quan trắc từ máy đo mưa. Công thức sử dụng các biến phân cực có giá trị tương quan cao và sai số nhỏ hơn chỉ sử dụng giá trị của Z. Các công thức có sự phân loại đặc điểm mưa có độ chính xác cao hơn so với việc không phân loại. Công thức R(Z,ZDR,KDP) ứng với hai loại đặc điểm mưa có r cao nhất và sai số RMSE thấp.

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

Lời cảm ơn: Bài báo hoàn thành nhờ vào kết quả của luận văn thạc sĩ: "Phân loại đặc điểm mưa và ước tính mưa bằng quan trắc Ra đa phân cực cho khu vực Tây Bắc".

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. Điền, N.H.; Đa, T.V. Khí tượng Ra đa, Nhà xuất bản Trường Đại Học Khoa học Tự nhiên, Đại học Quốc gia Hà Nội, 2009, tr. 208.

2. Houze, R.A. Cloud Dynamics. Academic Press: Cambridge, MA, USA, First Eds.: 1993, pp. 573.
3. Austin, P.M.; Houze Jr, R.A. Analysis of the structure of precipitation patterns in New England. *J. Appl. Meteorol*. **1972**, *11*, 926–935.
4. Houze Jr, R.A. A Climatological study of vertical transports by cumulus-scale convection. *J. Atmos. Sci.* **1973**, *30*, 1112–1123.
5. Houze Jr, R.A. Structure and dynamics of a tropical squall-line system. *Mon. Wea. Rev*. **1977**, *105*, 1540–1567.
6. Houze Jr, R.A.; Smull, B.F.; Dodge, P. Mesoscale organization of springtime rainstorms in Oklahoma. *Mon. Wea. Rev*. **1990**, *118*, 613–654.
7. Adler, R.F.; Negri, A.J. A satellite infrared technique to estimate tropical convective and stratiform rainfall. *J. Appl. Meteorol*. **1988**, *27*, 30–51.
8. Biggerstaff, M.I.; Listemaa, S.A. An improved scheme for convective/stratiform echo classification using radar reflectivity. *J. Appl. Meteorol.* **2000**, *39*, 2129–2150.
9. Yang, Y.; Chen, X.; Qi, Y. Classification of convective/stratiform echoes in radar reflectivity observations using a fuzzy logic algorithm. *J. Geophys. Res. Atmos.* **2013**, *118*, 1896–1905.
10. Steiner, M.; Houze Jr, R.A.; Yuter, S.E. Climatological characterization of three-dimensional storm structure from operational radar and rain gauge data. *J. Appl. Meteorol.* **1995**, *34*, 1978–2007.
11. Ikeda, K.; Brandes, E.A. Freezing level determinations with polarimetric radar: Retrieval model and application. *J. Appl. Meteorol.* **2003**, *43*, 1541–1553.
12. Rico-Ramirez, M.A.; Cluckie, I.D.; Han, D. Correction of the bright band using dual-polarisation rada. *Atmos. Sci. Lett.* **2005**, *6*, 40–46.
13. Bringi, V.N.; Williams, C.R.; Thurai, M.; May, P.T. Using dualpolarized radar and dual-frequency profiler for DSD characterization: a case study from Darwin, Australial. *J. Atmos. Oceanic Technol.* **2009**, *26*, 2107–2122.
14. Wang, H.; Shao, N.; Ran, Y. Identification of Precipitation Clouds Based on the Dual-Polarization Doppler Weather Radar Echoes Using Deep-Learning Method. *IEEE Access*. **2018**, *7*, 12822–12831.
15. Wang, Y.; Tang, L.; Chang, P.; Tang, Y. Separation of convective and stratiform precipitation using polarimetric radar data with a support vector machine method. *Atmos. Meas. Tech.* **2021**, *14*, 185–197.
16. Sơn, T.D. Nghiên cứu sử dụng thông tin ra đa thời tiết phục vụ theo dõi, cảnh báo mưa, dông và bão. Báo cáo tổng kết đề tài nghiên cứu khoa học cấp Bộ, Bộ Tài nguyên và Môi Trường, 2007.
17. Thắng, N.V. và cs. Nghiên cứu khai thác các định dạng số liệu, tổ hợp và xây dựng phần mềm xác định vị trí tâm mắt bão, hướng và tốc độ di chuyển của tâm bão cho mạng lưới ra đa thời tiết ở Việt Nam. Báo cáo tổng kết đề tài nghiên cứu khoa học cấp Bộ, Bộ Tài nguyên và Môi Trường, 2012.
18. Toán, H.M. Xây dựng công thức tính lượng mưa từ số liệu Ra đa Đốp-Le cho khu vực Trung Trung Bộ. Luận văn thạc sĩ, Trường Đại Học Khoa học Tự nhiên, Đại học Quốc gia Hà Nội, 2009.
19. Điền, N.H. Công thức thực nghiệm tính toán cường độ mưa từ độ phản hồi vô tuyến quan trắc bởi Ra đa cho khu vực Trung Trung Bộ. *Tạp chí khoa học: Khoa học Tự nhiên và Công nghệ* **2010**, *26(38)*, 317–321.
20. Điền, N.H. Công thức thực nghiệm tính toán cường độ mưa từ độ phản hồi radar cho khu vực Đông Nam Bộ. *Tạp chí Khoa học: Khoa học Tự nhiên và Công nghệ* **2015**, *31(3S)*, 66-70.
21. Burges, C.J.C. A tutorial on Support Vector Machines for pattern recognition. *Data Min. Knowl. Discovery*. **1988**, *2*, 955–974.

22. Haykin, S.O. Neural networks and learning machines, Pearson Higher Education, Pearson Education, Inc., Upper Saddle River, New Jersey, 2011.
23. Houghton, H.G. On precipitation mechanisms and their artificial modification. *J. Appl. Meteor*. **1968**, *7*, 851–859.
24. Houze Jr, R.A. Stratiform precipitation in regions of convection: A meteorological paradox. *Bull. Amer. Meteor. Soc.* **1997**, *78*, 2179–2196.

Study on the classification of rain characteristics and rainfall estimation by dual-polarization radar observations for the Northwest region

Nguyen Khac Quan[1]*, Pham Thi Thanh Nga[2]

[1] Northwestern Regional Hydro - Meteorological Center; mr.nguyenkhacquan@gmail.com
[2] Vietnam Institute of Meteorology, Hydrology and Climate change; pttnga.monre@gmail.com

Abstract: Rain can generally have convective and stratiform characteristics, associating with the different processes of cloud microphysics development which leads to differences in intensity, time and area of rain. Using dual-polarization radar observations in the Northwest region, the study applied the Support Vector Machine classification algorithm for classifying rain characteristics. The classification results are then used to find the empirical coefficients in the rainfall estimation formulas based on the Least Squares Method. The evaluation shows that with the proposed method, only less than 10% of the areas with bright bands are wrongly identified as convective rain, the method has also overcome some misclassifications compare to the reference method in some specific rain cases. Rainfall estimation results show that the estimated rainfall is lower than the observed values, the formulas with the classification of rain characteristics and using dual-polarization variables have better results than the formulas using only the of Z, in which the formula R (Z, ZDR, KDP) provides the highest correlation and the smallest error.

Keywords: Classification of rain characteristics; Support Vector Machine; Northwest region.