Nghiên cứu

NGHIÊN CỨU DỰ BÁO DÔNG CHO SÂN BAY ĐÀ NẴNG BẰNG PHƯƠNG PHÁP THỐNG KÊ VẬT LÝ

Trần Chấn Nam, Nguyễn Công Dũng
Trường Đại học Tài nguyên và Môi trường Hà Nội

Tóm tắt

Dông là hiện tượng thiên tai phổ biến xảy ra ở hầu khắp các khu vực trên cả nước tại bất cứ thời gian nào trong năm và gây ảnh hưởng lớn đến các lĩnh vực của nền kinh tế, xã hội như hàng không, hàng hải, giao thông, du lịch,...và đặc biệt là ảnh hưởng trực tiếp đến đời sống của người dân. Vì vậy việc nghiên cứu dự báo dông phục vụ cho công tác bay tại các sân bay nói chung, cảng hàng không quốc tế Đà Nẵng nói riêng là rất cần thiết. Nghiên cứu này sử dụng số liệu quan trắc tại trạm, số liệu thám không vô tuyến, số liệu mô hình và số liệu tái phân tích để tính toán và phân tích cụ thể, chi tiết về hiện tượng dông cho khu vực sân bay Đà Nẵng và lân cận. Kết quả cho thấy, các phương trình dự báo dông thời hạn đến 24 giờ cho khu vực sân bay Đà Nẵng có kết quả thử nghiệm với bộ số liệu phụ thuộc và độc lập đáp ứng được yêu cầu của bài toán dự báo dông.

Từ khóa: Dự báo dông; Phương pháp thống kê vật lý; Đà Nẵng

Abstract

Research to forecast thunderstorm for Da Nang airport using physical-statistical method

Thunderstorms are common natural disasters that occur in almost all regions of Vietnam at any time of the year. They have great impacts on economic and social sectors such as aviation, maritime, traffic, tourism especially, directly affect people's lives. Therefore, it is very necessary to study the thunderstorms forecast for aviation at airports in general and Da Nang international airports in particular. Using monitoring data at the station, radio detective data, model data and re-analysis data, the research studied, calculated and analyzed details of thunderstorms for Da Nang airport area and adjacent areas. The results show that the thunder forecast equations up to 24 hours for Da Nang airport area using independent and dependent data sets meet the requirements of the thunder forecasting problem.

Keywords: Forecast thunderstorm; Physical-statistical method; Da Nang

1. Mở đầu

Dông là hiện tượng thiên tai phổ biến xảy ra ở hầu khắp các khu vực trên cả nước tại bất cứ thời gian nào trong năm và gây ảnh hưởng lớn đến các lĩnh vực của nền kinh tế, xã hội như hàng không, hàng hải, giao thông, du lịch,... Những thiệt hại của dông gây ra chủ yếu bởi gió giật mạnh, mưa rào (có khi mưa đá), sự phóng điện khí quyển, sự thay đổi khí áp đột ngột,...trong cơn dông.

Ở Việt Nam cùng với sự tiến bộ của khoa học kĩ thuật vấn đề dự báo dông đã có nhiều tiến bộ rõ rệt trong những năm trở lại đây. Các sản phẩm dự báo dông từ các mô hình số trị, số liệu ra đa, số liệu quan trắc tự động, số liệu phân tích synop,...đã góp phần nâng cao chất lượng dự báo dông.

Nguyễn Viết Lành và cs (2000) [1] đã thực hiện đề tài *"Xây dựng phương pháp dự báo dông nhiệt nửa đầu mùa hè ở Hà*

Nội", với hai mục đích chính là: tìm hiểu quy luật thống kê của một số đặc trưng về dông nhiệt và những yếu tố khí tượng có liên quan; xây dựng phương pháp dự báo dông nhiệt cho khu vực Hà Nội trong các tháng đầu mùa hạ. Kết quả đề tài đã tính được: một số đặc trưng của chế độ dông và dông nhiệt trong các tháng 5, 6 và 7 ở khu vực Hà Nội như số ngày dông, số giờ dông, số đợt dông và tần suất xuất hiện dông vào các giờ trong ngày; xác định được mối tương quan giữa sự xuất hiện dông nhiệt ở Hà Nội với lượng mây tích tại Hoà Bình cũng như lượng mây tích tại Hà Nội vào các kỳ quan trắc 7, 10 và 13 giờ; tính được các chỉ số độ bất ổn định của khí quyển cho khu vực Hà Nội trong các tháng nói trên rồi đưa ra được những kết luận có đủ cơ sở khoa học về khả năng áp dụng các chỉ số đó; xây dựng được hai phương trình dự báo khả năng xuất hiện dông nhiệt với thời hạn dự báo từ 6 đến 12 giờ cho khu vực Hà Nội trong các tháng 5, 6 và 7 bằng phương pháp hồi quy nhiều biến và bằng hàm phân lớp với độ chính xác chung và độ tin cậy cao cho cả chuỗi số liệu phụ thuộc và chuỗi số liệu độc lập.

Hoàng Phúc Lâm và cs (2019) [2], đã tiến hành nghiên cứu *"Xây dựng mô hình thống kê hiệu chỉnh kết quả dự báo mưa cho khu vực Bắc Bộ sử dụng xoáy thế"*. Trong nghiên cứu này, tác giả xây dựng mô hình hồi quy từng bước dự báo lượng mưa cho khu vực Bắc Bộ sử dụng yếu tố dự báo là tổng lượng mưa quan trắc 24 giờ trung bình khu vực và nhân tố dự báo là số liệu trung bình ngày các yếu tố xoáy thế, tốc độ thẳng đứng, độ phân kỳ trên các mực đẳng nhiệt độ thế và mực đẳng áp. Kết quả đánh giá độc lập cho thấy lượng mưa dự báo có tương quan tốt (0,6 - 0,8, cá biệt có điểm lưới đạt giá trị tương quan trên 0,9), sai số RMSE còn khá nhỏ (4 - 6mm/ngày) nhưng thường vẫn cho lượng mưa thấp hơn so với thực tế trong các ngày mưa lớn; trong khi, thực tế mưa nhỏ thì mưa tính toán từ phương trình cho giá trị lớn hơn so với thực tế.

Trên thế giới các nhà nghiên cứu cũng đã có những thành công trong việc nghiên cứu dự báo dông. Schmeits và cs (2005) [3] đã tiến hành nghiên cứu *"Dự báo xác suất xảy ra dông/dông nghiêm trọng ở Hà Lan bằng việc sử dụng phương pháp thống kê trên cơ sở đầu ra mô hình"*. Nhóm tác giả xây dựng phương trình hồi quy logistic để dự báo xác suất xuất hiện dông/dông nghiêm trọng trong các tháng mùa nóng (từ giữa tháng 4 đến giữa tháng 10). Các phương trình được xây dựng từ việc sử dụng thống kê đầu ra mô hình (MOS). Nguồn số liệu sử dụng để xây dựng mô hình gồm: dữ liệu sét từ hệ thống định vị phóng điện (SARFIR); 15 chỉ số dông được tính toán từ mô hình HIRLAM và trích xuất (sau xử lý) từ số liệu đầu ra mô hình của Trung tâm Dự báo thời tiết hạn vừa châu Âu (ECMWF). Sau quá trình tuyển chọn nhân tố, tác giả đã lựa chọn được 5 nhân tố dự báo; đồng thời đưa ra kết luận, trong phạm vi nghiên cứu, nhân tố dự báo quan trọng nhất là căn bậc hai của tổng lượng mưa đối lưu 6 giờ một lấy từ ECMWF và nhân tố dự báo quan trọng nhất trong hệ thống dự báo dông nghiêm trọng là chỉ số Boyden tính toán từ mô hình HYRLAM; thành công của nhân tố căn bậc hai của tổng lượng mưa đối lưu 6 giờ một lấy từ ECMWF cho thấy có mối quan hệ chặt chẽ giữa lượng mưa đối lưu dự báo theo mô hình ECMWF và sự xuất hiện của dông trên khu vực Hà Lan. Kết quả đánh giá tổng thể cho thấy hệ thống dự báo tốt và được vận hành tại Viện khí tượng hoàng gia Hà Lan.

Guqian Pang và cs (2019) [4], đã thực hiện nghiên cứu *"Một mô hình hồi quy logistic nhị phân để dự báo đối lưu*

mạnh bằng số liệu mô hình số". Trong nghiên cứu, các tác giả sử dụng số liệu quan trắc khí tượng và số liệu từ mô hình số (GRAPES, ECMWF) từ tháng 3 đến tháng 4 năm 2014, để xây dựng phương trình dự báo. Phương pháp xây dựng mô hình hồi quy logistic nhị phân được sử dụng để xây dựng phương trình dự báo, thời hạn dự báo là 12 giờ, khu vực dự báo là tỉnh Quảng Đông, Trung Quốc; các chỉ số được sử dụng để đánh giá kết quả là chỉ số K, TT (Total total), MDPI (microburst-day potential index), IQ (tích phân Q, Q là xác suất không có hiện tượng đối lưu mạnh xảy ra). Kết quả cho thấy các phương trình có kỹ năng dự báo tương đối tốt, nhưng đều có xu hướng giảm dự báo sót và tăng dự báo khống; đồng thời nhóm nghiên cứu cũng đề cập đến triển vọng áp dụng phương trình dự báo trong nghiên cứu vào thực tế khi chất lượng dự báo của các mô hình số ngày càng tăng lên.

Xuất phát từ thực tế trên, bài báo này sử dụng phương pháp dự báo thống kê đã và đang được sử dụng rất rộng rãi trên thế giới cũng như tại Việt Nam. Bài báo sử dụng phương pháp hồi quy nhị phân logistic để xây dựng phương trình dự báo dông khu vực sân bay Đà Nẵng. Phần tiếp theo của bài báo sẽ mô tả số liệu và phương pháp nghiên cứu ở Mục 2. Mục 3 là kết quả. Cuối cùng phần kết luận ở Mục 4.

2. Số liệu và phương pháp nghiên cứu

2.1. Số liệu

Số liệu được sử dụng trong nghiên cứu trong gồm có số liệu thực đo trong 5 năm 2014 - 2018 tại các trạm quan trắc. Các yếu tố sử dụng trong bài báo là: Dông (trong đó: số đợt dông năm 2014 là 39 đợt; số đợt dông năm 2015 là 26 đợt; số đợt dông năm 2016 là 25 đợt; số đợt dông năm 2017 là 33 đợt; số đợt dông năm 2018 là 25 đợt), nhiệt độ, điểm sương, khí áp, gió và độ ẩm.

Số liệu cao không trong 5 năm 2014 - 2018 lấy từ trang web http://weather.uwyo.edu/ tại hai kì quan trắc 00z và 12z (tương ứng với kì quan trắc lúc 7 giờ và 19 giờ theo giờ Việt Nam), bao gồm các yếu tố khí tượng và chỉ số bất ổn định của khí quyển.

Số liệu từ các mô hình động lực trong 6 năm từ 2013 - 2018 tại 3 kì dự báo chính (00z, 06z, 12z) tại các mực: bề mặt, 925, 850, 700, 500mb. Các yếu tố sử dụng trong nghiên cứu là: Nhiệt độ, điểm sương, các chỉ số bất ổn định của khí quyển, độ cao địa thế vị, gió và độ ẩm. Trên cơ sở số liệu thu thập và các phương pháp thống kê vật lý sẽ xác định phương trình dự báo dông thời hạn tới 24 giờ cho khu vực sân bay Đà Nẵng.

2.2. Phương pháp nghiên cứu

Bài báo sử dụng phương pháp hồi quy nhị phân logistic để xây dựng phương trình dự báo dông khu vực sân bay Đà Nẵng.

1.2.1. Xác định yếu tố dự báo và nhân tố dự báo

Yếu tố dự báo là các biến khí quyển mà giá trị của chúng cần được xác định có thể là các biến liên tục hoặc rời rạc, các biến bị chặn hoặc không bị chặn, các biến phân hạng, biến nhị phân. Nhân tố dự báo là tập các biến mà giá trị của chúng được cho trước như là thông tin ban đầu để xác định giá trị của yếu tố dự báo.

Trong bài toán dự báo dông thì biến dự báo là biến nhị phân nhận một trong hai giá trị (thông thường là 0 và 1, mặc dù sự lựa chọn không làm ảnh hưởng đến kết quả).

Các phương trình dự báo dựa trên cơ sở thống kê cho ta biết thông tin về xác suất xuất hiện của một yếu tố dự báo nào đó. Dự báo xác suất là dự báo trong đó

yếu tố dự báo là một xác suất chứ không phải là giá trị của biến khí tượng. Thông thường, phương trình dự báo xác suất được xây dựng trên cơ sở phép hồi quy bằng việc biến đổi yếu tố dự báo thành biến nhị phân nhận giá trị 0 và 1 như các nhân tố dự báo đã nói ở trên. Khi đó phép hồi quy sẽ được thực hiện. Nghĩa là 0 và 1 được xem như là xác suất của yếu tố dự báo không xuất hiện và có xuất hiện.

Trên cơ sở đó, bài báo lựa chọn bộ nhân tố dự báo như sau:

Nhân tố dự báo

Ngày nay, nhờ sự phát triển mạnh mẽ của khoa học công nghệ cùng với năng lực tính toán ngày càng cao cho phép chúng ta đưa ra số lượng nhân tố dự báo rộng rãi hơn, và nhiều lựa chọn hơn cho việc xây dựng phương trình dự báo dông. Vì vậy, từ nguồn số liệu khai thác đã được đưa ra như trên việc lựa chọn nhân tố được xác định như sau:

Đối với dự báo 24h các giá trị quan trắc tại trạm được lấy lùi lại một ngày tại các obs 01h, 07h, 13h và 19h giờ Việt Nam.

Đối với số liệu dự báo được trích xuất ở thời đoạn 07h, 13h và 19h giờ Việt Nam đối với ngày dự báo.

1.2.2. Mô hình hồi quy nhị phân logistic

Hồi quy nhị phân logistic sử dụng biến phụ thuộc dạng nhị phân để ước lượng xác suất một sự kiện sẽ xảy ra với những thông tin của biến độc lập mà ta có được [6].

Đối với bài toán dự báo xác suất xuất hiện dông trong nghiên cứu này, sự kiện có dông được mã hóa nhận giá trị 1, hiện tượng không có dông được mã hóa nhận giá trị 0. Khi biến phụ thuộc ở dạng nhị phân thì nó không thể được nghiên cứu với dạng hồi quy thông thường vì nó sẽ xâm phạm các giả định, dễ thấy nhất là khi biến phụ thuộc chỉ có hai biểu hiện thì rất không phù hợp khi giả định rằng phần dư có phân phối chuẩn mà thay vào đó nó sẽ có phân phối nhị thức, điều này sẽ làm mất hiệu lực thống kế của các kiểm định trong phép hồi quy thông thường. Một khó khăn khác khi sử dụng hồi quy tuyến tính thông thường là giá trị dự đoán được của biến phụ thuộc không thể được diễn dịch như xác suất (giá trị ước lượng của biến phụ thuộc trong hồi quy Binary logistic rơi vào khoảng (0;1))

Với hồi quy nhị phân logistic, thông tin chúng ta cần thu thập về biến phụ thuộc là một sự kiện nào đó có xảy ra hay không, biến phụ thuộc Y lúc này có giá trị 0 (không xảy ra dông) và 1 (có dông), và thông tin về các biến độc lập X. Từ biến phụ thuộc nhị phân này, một thủ tục được sử dụng để dự đoán xác suất sự kiện xảy ra theo quy tắc nếu xác suất được dự đoán lớn hơn 0,5 thì kết quả dự đoán sẽ cho là "có" xảy ra sự kiện, ngược lại thì kết quả dự đoán sẽ là "không" xảy ra sự kiện.

Dựa vào những thông tin thu thập được từ biến phụ thuộc và các biến độc lập có thể xây dựng được phương trình hồi quy nhị phân logistic như sau:

$$log_e[\frac{P_i}{1-P_i}]=B_o+ B_1X_1+ B_2X_2+...+ B_kX_k$$

Trong đó: P_i là xác xuất xảy ra sự kiện; B_o, B_o, B_1, ..., B_k là hệ số hồi quy; X_o, X_1, X_2, ..., X_k là biến độc lập

Từ phương trình hồi quy bên trên có thể xác định phương trình mô hình hàm dự báo như sau:

$$P_i=E(Y=1/X)=\frac{e^z}{1+e^z}$$

Trong đó: $z=B_o+B_1X_1+...+ B_kX_k$,

P_i= E(Y=1/X)=P(Y=1) gọi là xác suất để sự kiện xảy ra (Y=1) khi biến độc lập X có giá trị cụ thể X_i.

a) Độ phù hợp của mô hình hồi quy

Hồi quy nhị phân logistic đòi hỏi phải đánh giá độ phù hợp của mô hình. Đo lượng độ phù hợp tổng quát của mô hình nhị phân logistic được dựa trên chỉ số -2LL (viết tắt của -2 log likelihood), thước đo này có ý nghĩa giống như SSE (sum of squares of error) nghĩa là giá trị của nó càng nhỏ thì càng tốt. Quy tắc đánh giá -2LL ngược lại với quy tắc dựa trên hệ số xác định mô hình R^2, nghĩa là giá trị -2LL càng nhỏ càng thể hiện độ phù hợp cao. Giá trị nhỏ nhất của -2LL là 0 (trường hợp không có sai số) khi đó mô hình là hoàn hảo.

b) Kiểm định ý nghĩa thống kê của các hệ số

Hồi quy nhị phân logistic cùng đòi hỏi kiểm định giả thuyết hệ số hồi quy khác không. Nếu hệ số hồi quy B_o, B_1, ..., B_k đều bằng 0 thì tỷ lệ chênh lệch giữa các xác xuất sẽ bằng 1, tức xác suất để sự kiện xảy ra hay không xảy ra là như nhau, lúc đó mô hình không còn khả dụng trong việc dự báo.

Trong hồi quy nhị phân logistic, đại lượng Wald Chi Square được sử dụng để kiểm định ý nghĩa thống kê của hệ số hồi quy tổng thể. Cách thức sử dụng mức ý nghĩa Sig. cho kiểm định Wald cùng theo quy tắc thông thường. Wald Chi Square được tính bằng cách lấy ước lượng của hệ số hồi quy của biến độc lập trong mô hình (hệ số hồi quy mẫu) nhị phân logistic chia cho sai số chuẩn của hệ số hồi quy này, sau đó bình phương lên theo công thức sau:

$$\text{Wald Chi Square}=\left(\frac{B}{s.e(B)}\right)^2$$

c) Kiểm định mức độ phù hợp tổng quát

Ở hồi quy nhị phân logistic, tổ hợp liên hệ tuyến tính của toàn bộ các hệ số trong mô hình ngoại trừ hằng số cũng được kiểm định xem có thực sự có ý nghĩa trong việc giải thích cho biến phụ thuộc không. Với hồi quy tuyến tính bội ta dùng thống kê F để kiểm định giả thuyết H_o: $\rho_1=\rho_2=...=\rho_k=0$, còn với hồi quy nhị phân logistic dùng kiểm định Chi- bình phương.

d) Các phương pháp đưa biến độc lập vào mô hình hồi quy nhị phân logistic

Với phương pháp hồi quy từng bước, số thống kê được sử dụng cho các biến được đưa vào và dời ra căn cứ trên số thống kê likelihood-ratio (tỷ lệ thích hợp) hay số thống kê Wald. Cũng có thể chọn một trong các phương pháp thay thế sau:

- Enter: Đưa vào bắt buộc, các biến trong khối biến độc lập được đưa vào trong một bước;

- Forward: Conditional là phương pháp đưa vào dần theo điều kiện. Nó kiểm tra việc loại biến căn cứ trên xác suất của số thống kê Likelihood-ratio dựa trên ước lượng thông số có điều kiện.

- Forward: LR là phương pháp đưa vào dần kiểm tra việc loại biến căn cứ trên xác suất của số thống kê Likelihood-ratio dựa trên ước lượng khả năng xảy ra tối đa (maximum-likelihood estimates)

- Forward: Wald là phương pháp đưa vào dần kiểm tra việc loại biến căn cứ trên xác suất của số thống kê Wald.

- Backwald: Conditional là phương pháp loại trừ dần theo điều kiện. Nó kiểm tra việc loại biến căn cứ trên xác suất của số thống kê Likelihood-ratio dựa trên ước lượng của những thông số có điều kiện.

- Backwald: LR là phương pháp loại trừ dần kiểm tra loại biến căn cứ trên xác suất của số thống kê Likelihood-ratio dựa trên ước lượng khả năng xảy ra tối đa.

- Backwald: Wald là phương pháp loại trừ dần kiểm tra loại biến căn cứ trên xác suất của số thống kê Wald.

1.2.3. Đánh giá chất lượng dự báo

Đối với các hiện tượng được dự báo theo 2 pha «có/ không» như dông người ta thường sử dụng đánh giá thống kê theo loại (categorical statistics) để biết tần suất xảy ra của hiện tượng dự báo.

Các điểm số đánh giá được dựa vào bảng ngẫu nhiên (Bảng 1).

Bảng 1. Bảng phân loại tần suất cho biến dự báo dạng nhị phân

Có		Quan trắc	
		Không	
Dự báo	Có	H	F
	Không	M	CN

Hay có thể diễn giải cụ thể thư sau:

Hits (H) = dự báo có + quan trắc có

Misses (M) = dự báo không + quan trắc có

False alarms (F) = dự báo có + quan trắc không

Correct negatives (CN) = dự báo không + quan trắc không

Dựa trên bảng tần suất này và giả sử có N bản ghi số liệu được nghiên cứu (H+M+F+CN=N), một số các chỉ số đánh giá được thiết lập để đánh giá kỹ năng dự báo của các mô hình dự báo cho yếu tố dự báo nhị phân bao gồm các chỉ số:

a) Chỉ số PC/FC (Percent Correct/ Fraction Correct)

b) Chỉ số FBI (BS hay FBI- Bias score)

c) Chỉ số POD (Probability of Detection)

d) Chỉ số FAR (False Alarm Ratio)

e) Chỉ số MR/POFD (Miss Rate/ Probability of False Detection)

f) Chỉ số TS/CSI (Threat Score/ Critical Suscess Index)

3. Kết quả

3.1. Xây dựng phương trình dự báo

Kết quả xây dựng phương trình dự báo dông cho Đà Nẵng tại các tháng 5, 6, 7 và 8 được trình bày trong Bảng 2.

Bảng 2. Các phương trình dự báo dông tại trạm Đà Nẵng ***từ tháng 5 đến 8***

Tháng	Phương trình dự báo
5	$P_i=\frac{e^{-10.324-3.627*13BP12h+0.769*13VGRD700+0.288*B7HGT18h925}}{1+e^{-10.324-3.627*13BP12h+0.769*13VGRD700+0.288*B7HGT18h925}}$
6	$P_i=\frac{e^{-9.482-1.216*19FF925+0.161*7RH500+0.399*B19K24h}}{1+e^{-9.482-1.216*19FF925+0.161*7RH500+0.399*B19K24h}}$
7	$P_i=\frac{e^{-25.77+0.271*19RH700+0.346*B7UGRD18h850-0.049*B7SWEAT12h}}{1+e^{-25.77+0.271*19RH700+0.346*B7UGRD18h850-0.049*B7SWEAT12h}}$
8	$P_i=\frac{e^{-58.179+0.084*7HGT925-0.184*7BU18h-0.375*B7VGRD24h700}}{1+e^{-58.179+0.084*7HGT925-0.184*7BU18h-0.375*B7VGRD24h700}}$

Từ Bảng 2 ta thấy các yếu tố chủ yếu được lựa chọn làm biến dự báo trong mô hình bao gồm cả yếu tố quan trắc, yếu tố dự báo và cả các chỉ số độ bất ổn định của khí quyển. Cụ thể như sau:

Đối với phương trình dự báo trong tháng 5, các nhân tố được lựa chọn làm biến dự báo bao gồm:

TT	Ký hiệu	Nhân tố
1	13BP12h	Biến áp 12h tại obs 13 giờ
2	13VGRD700	Gió kinh hướng mực 700mb tại obs 13 giờ
3	B7HGT18h925	Biến thiên độ cao địa thế vị 18h mực 925mb obs 7 giờ

Đối với phương trình dự báo trong tháng 6, các nhân tố được lựa chọn làm biến dự báo bao gồm:

TT	Ký hiệu	Nhân tố
1	19FF925	Tốc độ gió mực 925mb lúc 19 giờ
2	7RH500	**Độ ẩm riêng mực 500 mb tại obs 7 giờ**
3	B19K24h	Biến thiên chỉ số K 24h tại obs 19 giờ

Đối với phương trình dự báo trong tháng 7, các nhân tố được lựa chọn làm biến dự báo bao gồm:

TT	Ký hiệu	Nhân tố
1	19RH700	**Độ ẩm riêng mực 700 mb tại obs 19 giờ**
2	B7UGRD18h850	Biến thiên gió vĩ hướng 18h mực 850mb obs 7 giờ
3	B7SWEAT12h	Biến thiên SWEAT 12h tại obs 7 giờ

Đối với phương trình dự báo trong tháng 8, các nhân tố được lựa chọn làm biến dự báo bao gồm:

TT	Ký hiệu	Nhân tố
1	7HGT925	Độ cao địa thế vị mực 925 tại obs 7 giờ
2	7BU18h	Biến thiên độ ẩm 18h tại obs 7 giờ
3	B7VGRD24h700	Biến thiên gió kinh hướng 24h mực 700mb obs 7 giờ

Mỗi nhân tố được lựa chọn chỉ xuất hiện một lần trong các phương trình. Nhóm các yếu tố được sử dụng trong mô hình bao gồm: tốc độ gió, độ ẩm, độ cao địa thế vị, khí áp và hai chỉ số K, SWEAT; cũng có thể kết luận rằng những nhân tố này đóng vai trò quan trọng đối với sự xuất hiện dông trên khu vực Đà Nẵng.

Từ thống kê ở trên cũng có thể thấy rằng số liệu dự báo từ mô hình GFS được sử dụng với tần suất tương đối lớn trong các mô hình chứng tỏ rằng tương quan giữa số liệu dự báo từ mô hình GFS với hiện tượng dông trên khu vực Đà Nẵng là tương đối tốt. Đồng thời cũng chứng tỏ khả năng dự báo của mô hình dự báo GFS đối với hiện tượng quy mô nhỏ như dông tại Việt Nam.

3.2. Thử nghiệm dự báo

Từ những phương trình đã xây dựng được bài báo tiến hành đánh giá khả năng dự báo của các phương trình này thông qua các chỉ tiêu đánh giá. Chuỗi số liệu phụ thuộc dùng để đánh giá lấy từ năm 2014 đến năm 2017, chuỗi số liệu độc lập được sử dụng để tính toán lấy trong năm 2018.

Kết quả đánh giá chất lượng của các phương trình được trình bày trong Bảng 3.

Bảng 3. Kết quả dự báo dông trên chuỗi số liệu phụ thuộc (từ năm 2014 đến năm 2017)

Trạm	Tháng	Tần suất				Tổng	Chỉ tiêu đánh giá						
		H	F	M	CN		PC	% dự báo	FBI	POD	FAR	MR	TS
Đà Nẵng	5	7	7	13	59	86	0,77	76,74	0,7	0,35	0,5	0,11	0,26
	6	4	4	7	78	93	0,88	88,17	0,73	0,36	0,5	0,05	0,27
	7	16	5	8	40	69	0,81	81,16	0,88	0,67	0,24	0,11	0,55
	8	19	17	13	49	98	0,69	69,39	1,13	0,59	0,47	0,26	0,39

Từ kết quả đánh giá các phương trình dự báo dông đối với bộ số liệu phụ thuộc, có thể thấy độ chính xác PC của các phương trình là tương đối cao, dao động trong khoảng từ 0,69 đến 0,8 (trong đó, PC thấp nhất trong phương trình dự báo dông tháng 8 và cao nhất trong phương trình dự báo dông tháng 6), chứng tỏ rằng tỷ lệ trùng khớp giữa kết quả của phương trình đã xây dựng với quan trắc hiện tượng thực tế trong cả hai pha có và không có hiện tượng là tương đối lớn.

Chỉ số FBI của các phương trình dao động trong khoảng từ 0,7 đến 1,13. Giá trị FBI của các phương trình dự báo trong các tháng 5, 6, 7 nhỏ hơn 1 chứng tỏ trong các tháng này các phương trình cho kết quả dự báo sót nhiều và dự báo sót nhiều nhất là phương trình dự báo dông trong tháng 5. Ngược lại, trong tháng 8, FBI của phương trình dự báo cho giá trị dương, chứng tỏ rằng phương trình cho kết quả dự báo khống; tuy nhiên sai khống không đáng kể (do FBI không khác giá trị 1 quá nhiều).

Xác xuất phát hiện hiện tượng POD của các phương trình trong tháng 7 và tháng 8 tương đối cao hơn so với các phương trình trong tháng 5 và tháng 6 thể hiện ở giá trị POD của các phương trình dự báo trong tháng 7, 8 cao hơn so với các phương trình dự báo trong tháng 5, 6. Tuy nhiên, như đã nói ở phần định nghĩa, POD chỉ nhạy với những hiện tượng không dự báo được chứ không nhạy với những phát hiện sai do đó cần xem xét thêm các chỉ số như FAR và MR để đánh giá chất lượng phương trình.

Tỷ lệ dự báo phát hiện sai FAR của các phương trình tương đối nhỏ, giá trị FAR dao động từ 0,24 đến 0,5 và nhỏ nhất trong phương trình dự báo dông tháng 7. Điều này chứng tỏ, đối với chuỗi số liệu nghiên cứu, các mô hình dự báo tháng 5, 6, 8 cho kết quả số lần dự báo đúng trong pha có xảy ra hiện tượng (Hits) và dự báo khống (False alarms) là tương đương nhau, còn phương trình dự báo trong tháng 7 cho kết quả dự báo khống ít hơn.

Giá trị chỉ số MR của tất cả các phương trình đều rất nhỏ (rất gần giá trị 0, là giá trị mà ở đó phương trình được xem là hoàn hảo), chứng tỏ các phương trình cho kết quả dự báo rất tốt ở pha không xuất hiện hiện tượng; đặc biệt là vào tháng 6, phương trình cho kết quả dự báo rất gần với giá trị 0 (MR của trong trình dự báo trong tháng 6 là 0,05).

Giá trị chỉ số TS của các phương trình dự báo tháng 7, 8 tương đối cao hơn so với các tháng 5, 6. Điều này chứng tỏ, đối với pha "có xảy ra hiện tượng" thì phương trình dự báo tháng 7, 8 tỏ ra tốt hơn các phương trình trong các tháng 5, 6. Tuy nhiên, cũng cần nhấn mạnh rằng đối với chỉ số này thì những trường hợp "không xuất hiện hiện tượng" đã bị bỏ

qua và nó chỉ xem xét đến tỷ số giữa số lần mô hình cho kết quả hiện tượng có xuất hiện và số lần quan trắc được hiện tượng có xuất hiện.

Khi so sánh các chỉ số tính toán từ các phương trình dự báo đối với chuỗi số liệu phụ thuộc, có thể thấy, tất cả các phương trình dự báo dông trong tất cả các tháng được xét đến đều có thể dự báo hiện tượng dông với mức độ chính xác nhất định. Để đánh giá một cách khách quan hơn nhóm tác giả tiến hành đánh giá các phương trình trên với bộ số liệu độc lập năm 2018 để có thể đưa đến những kết luận khách quan hơn. Kết quả đánh giá các mô hình dự báo được trình bày trong Bảng 4.

Bảng 4. Kết quả dự báo dông trên chuỗi số liệu độc lập (năm 2018)

Tháng	Tần suất				Tổng	Chỉ tiêu đánh giá						
	H	F	M	CN		PC	% dự báo	FBI	POD	FAR	MR	TS
5	1	3	3	23	30	0,80	80,00	1,00	0,25	0,75	0,12	0,14
6	1	1	9	19	30	0,67	66,67	0,20	0,10	0,50	0,05	0,09
7	1	6	4	18	29	0,66	65,51	1,40	0,20	0,86	0,25	0,09
8	2	2	4	21	29	0,79	79,31	0,67	0,33	0,50	0,09	0,25

Từ kết quả tính toán các chỉ số dựa trên chuỗi số liệu độc lập ở trên có thể thấy sự phân hóa rõ rệt hơn chất lượng các phương trình dự báo dông trong các tháng.

Đối với phương trình dự báo dông trong tháng 5, cho kết quả PC khá cao so với các phương trình còn lại và giá trị FBI bằng 1 (với ý nghĩa là vùng quan trắc bằng với vùng dự báo). Tuy nhiên khi xem xét các chỉ số khác thì có vẻ như các giá trị PC và FBI cao có xu hướng là do các dự báo không chính xác mang lại chẳng hạn như khi xét đến chỉ số FAR có thể thấy tỷ lệ dự báo khống của phương trình này là tương đối cao, chỉ số MR cũng lớn hơn so với các phương trình dự báo tháng 6, 8; điểm số thành công TS và xác xuất phát hiện hiện tượng cũng không cao vượt trội so với các phương trình khác; chỉ số FBI đạt giá trị lý tưởng là do số lần dự báo khống và dự báo sót bằng nhau, trong khi dự báo đúng trong pha "có xuất hiện hiện tượng" chỉ là 1 trường hợp (lưu ý trong bảng tần suất, H=1, F=M=3), trường hợp này xảy ra có thể là do chuỗi số liệu được lựa chọn chưa đủ dài để cho kết quả chính xác, hơn nữa FBI chỉ cho biết mức độ phù hợp giữa mô hình và quan trắc về tần số xuất hiện nhưng không phản ánh độ chính xác của mô hình.

Đối với phương trình dự báo dông trong tháng 6, độ chính xác PC chỉ cao hơn so với phương trình dự báo trong tháng 7 một chút, các chỉ số TS, POD đều nhỏ hơn so với các phương trình còn lại; FBI nhỏ chứng tỏ phương trình này chủ yếu đưa ra các dự báo sót; cũng có thể thấy chỉ số FAR cho kết quả tương đối nhỏ so với các phương trình dự báo tháng 5, 7 chứng tỏ phương trình này cho số dự báo khống ít hơn các phương trình còn lại. Chỉ số MR của phương trình này cũng nhỏ nhất trong số các phương trình dự báo và gần giá trị lý tưởng nhất (MR=0,05), chứng tỏ phương trình dự báo tương đối tốt đối với pha "không xuất hiện hiện tượng". Cũng cần xem xét kết quả khi đánh giá phương trình này với bộ số liệu phụ thuộc (từ 2014 - 2017), ta thấy các chỉ số của phương trình này tương đối ổn định, có thể lưu ý để lựa chọn giữ lại hay loại bỏ khi so sánh với các phương trình dự báo khác.

Phương trình dự báo dông trong tháng 7 có PC thấp nhất và FAR, MR, FBI cao nhất so với các phương trình còn lại. Giá trị POD nằm ở mức trung bình, trong khi giá trị TS lại khá thấp. Nhìn chung phương trình này cho kết quả dự báo thiên về dự báo khống.

Phương trình dự báo dông trong tháng 8 nhìn chung cho kết quả dự báo tương đối tốt khi thử nghiệm với chuỗi số liệu độc lập. Các chỉ số PC, POD, TS nhìn chung cao hơn so với các phương trình dự báo khác. Chỉ số FBI nhỏ hơn 1 và các chỉ số FAR, MR nhỏ hơn các chỉ số khác chứng tỏ phương trình này thiên về hướng dự báo sót và ít đưa ra các dự báo khống.

Từ việc tính toán thử nghiệm các phương trình dự báo dựa trên bộ số liệu phụ thuộc và độc lập có thể đi đến kết luận: Phương trình dự báo trong tháng 6 và tháng 8 là phù hợp để dự báo dông cho khu vực Đà Nẵng và ta loại bỏ hai phương trình dự báo trong tháng 5 và tháng 7 do hai phương trình này đưa ra nhiều dự báo khống, không chính xác có thể dẫn đến nhận định sai lầm trong quá trình dự báo. Phương trình dự báo trong tháng 6 tuy cho chất lượng không cao đối với bộ số liệu độc lập tuy nhiên khi xét về độ ổn định các chỉ số và mức độ thành công khi đánh giá với bộ số liệu phụ thuộc vẫn có thể sử dụng phương trình này như một phương án tham khảo trong dự báo.

3.3. Kết quả lựa chọn phương trình dự báo dông cho khu vực sân bay Đà Nẵng và lân cận

Từ những kết quả ở trên bài báo lựa chọn được hai phương trình phù hợp để dự báo dông cho khu vực sân bay Đà Nẵng và lân cận gồm:

a) Tháng 6:

$$P_i=\frac{e^{-9.482-1.216*19FF925+0.161*7RH500+0.399*B19K24h}}{1+e^{-9.482-1.216*19FF925+0.161*7RH500+0.399*B19K24h}}$$

Trong đó: 19FF925 là tốc độ gió mực 925mb lúc 19 giờ; 7RH500 là độ ẩm tương đối mực 500mb tại obs 7 giờ; B19K24h là biến thiên chỉ số K 24h tại obs 19 giờ.

- Ý nghĩa của phương trình dự báo mưa dông tháng 6 tại khu vực sân bay Đà Nẵng: Để hiểu được ý nghĩa của phương trình trên ta tiến hành thay thế một số liệu thực tế từ một ngày cụ thể, VD: ngày 25 tháng 6 năm 2018, là ngày xảy ra dông tại khu vực Đà Nẵng các giá trị 19FF925, 7RH500, B19K24h ngày 24 tháng 6 lần lượt là 4.33m/s, 85%, 3.2 thay vào phương trình ta tính được P_{25}=0,55 (giá trị này >0,5). Từ đó, ta có thể đưa ra dự báo ngày 25 tháng 6 năm 2018, trên khu vực sân bay Đà Nẵng có dông và độ chính xác của dự báo này là khoảng 88% (dựa vào chỉ số PC (độ chính xác) đã tính toán trên bộ số liệu phụ thuộc ở trên).

b) Tháng 8:

$$P_i=\frac{e^{-58.179+0.084*7HGT925-0.184*7BU18h-0.375*B7VGRD24h700}}{1+e^{-58.179+0.084*7HGT925-0.184*7BU18h-0.375*B7VGRD24h700}}$$

Trong đó: 7HGT925 là độ cao địa thế vị mực 925mb tại obs 7 giờ; 7BU18h là biến thiên độ ẩm 18h tại obs 7 giờ; B7VGRD24h700 là biến thiên gió kinh hướng 24 giờ mực 700mb tại obs 7 giờ

- Ý nghĩa của phương trình dự báo mưa dông tháng 8 tại khu vực sân Bay Đà Nẵng: Để hiểu được ý nghĩa của phương trình trên ta tiến hành thay thế một số liệu

thực tế từ một ngày cụ thể, VD: ngày 01 tháng 8 năm 2018, là ngày xảy ra dông tại khu vực Đà Nẵng các giá trị 7HGT925, 7BU18h, B7VGRD24h700 ngày 01 tháng 8 lần lượt là 746,838; 30; -3,48 thay vào phương trình ta tính được P_{01}=0,58 (giá trị này >0,5). Từ đó, ta có thể đưa ra dự báo ngày 01 tháng 8 năm 2018, trên khu vực sân bay Đà Nẵng có dông và độ chính xác của dự báo này là khoảng 69% (dựa vào chỉ số PC (độ chính xác) đã tính toán trên bộ số liệu phụ thuộc ở trên).

Như vậy từ tập 414 nhân tố ban đầu bài báo đã tính toán, phân tích, chọn lọc ra bộ 6 nhân tố có tương quan cao với yếu tố dự báo từ đó xây dựng 02 phương trình dự báo dông cho khu vực Đà Nẵng. Các kết quả thử nghiệm, đánh giá dựa trên bộ số liệu độc lập và phụ thuộc ở phần trên cho thấy khả năng dự báo của phương trình là đáng kể, có thể đưa vào hoạt động dự báo thực tiễn như một phương án tham khảo để so sánh, đối chứng với các phương án dự báo khác.

4. Kết luận

Từ những nghiên cứu, tính toán và phân tích cụ thể, chi tiết về hiện tượng dông cho khu vực sân bay Đà Nẵng và lân cận, bài báo đã đạt được một số kết quả như sau:

Đã xây dựng được các phương trình dự báo dông thời hạn đến 24 giờ cho khu vực sân bay Đà Nẵng. Kết quả thử nghiệm với bộ số liệu phụ thuộc và độc lập cho thấy các phương trình xây dựng đáp ứng được yêu cầu của bài toán dự báo dông. Từ tập 414 nhân tố ban đầu đã tính toán, phân tích, chọn lọc ra bộ 6 nhân tố có tương quan cao với yếu tố dự báo từ đó xây dựng 02 phương trình dự báo dông cho khu vực Đà Nẵng.

Khi so sánh các chỉ số tính toán từ các phương trình dự báo đối với chuỗi số liệu phụ thuộc, có thể thấy, tất cả các phương trình dự báo dông trong tất cả các tháng được xét đến đều có thể dự báo hiện tượng dông với mức độ chính xác nhất định.

Khi so sánh các chỉ số tính toán từ các phương trình dự báo đối với chuỗi số liệu độc lập có thể thấy sự phân hóa rõ rệt hơn chất lượng các phương trình dự báo dông trong các tháng.

Từ việc tính toán thử nghiệm các phương trình dự báo dựa trên bộ số liệu phụ thuộc và độc lập có thể đi đến kết luận: Phương trình dự báo trong tháng 6 và tháng 8 là phù hợp để dự báo dông cho khu vực Đà Nẵng và ta loại bỏ hai phương trình dự báo trong tháng 5 và tháng 7 do hai phương trình này đưa ra nhiều dự báo khống, không chính xác có thể dẫn đến nhận định sai lầm trong quá trình dự báo.

TÀI LIỆU THAM KHẢO

[1]. Nguyễn Viêt Lành và cộng sự (2000). *Xây dựng phương pháp dự báo dông nhiệt nửa đầu mùa hè ở Hà Nội*. Đề tài nghiên cứu khoa học cấp Tổng cục, Trường Cán bộ Khí tượng Thủy văn Hà Nội, Tổng cục Khí tượng Thủy văn.

[2]. Hoàng Phúc Lâm, Nguyễn Đức Nam, Đỗ Thị Thanh Thủy, Hoàng Đức Cường (2019). *Xây dựng mô hình thống kê hiệu chỉnh kết quả dự báo mưa cho khu vực Bắc Bộ sử dụng xoáy thế.* Tạp chí Khí tượng Thủy văn, 697, 26 - 32.

[3]. Schmeits, Maurice J. et al (2005). *Probabilistic Forecasting of (Severe) Thunderstorms in the Netherlands Using Model Output Statistics*. Weather and Forecasting, Vol. 20, No2., pp. 134 - 148.

[4]. Guqian Pang et al (2019). *A Binary Logistic Regression Model for Severe Convective Weather with Numerical Model Data*. Hindawi, Advances in Meteorology, Volume 2019, Article ID 6127281, 15 pages, https://doi.org/10.1155/2019/6127281

BBT nhận bài: 11/9/2020; Phản biện xong: 16/9/2020; Chấp nhận đăng: 28/9/2020