

THỐNG KÊ ĐÁNH GIÁ DIỄN BIẾN DÔNG TẠI MỘC HÓA, MỸ THO VÀ SỬ DỤNG HÀM PHÂN LỚP ĐỂ DỰ BÁO, CẢNH BÁO DÔNG

Lê Đình Quyết¹, Lê Ngọc Quyền¹, Trần Thị Thu Uyên²

Tóm tắt: Nghiên cứu thống kê đánh giá diễn biến dông có ý nghĩa quan trọng làm cơ sở cho công tác dự báo, cảnh báo chỉ ra thời gian dông xảy ra trong ngày. Nghiên cứu này tập trung phân tích đánh giá dựa trên dữ liệu hai trạm Mộc Hóa, Long An và Mỹ Tho, Tiền Giang với chuỗi số liệu từ 2007-2018. Hàm phân lớp được áp dụng để đánh giá phục vụ cho dự báo. Kết quả nghiên cứu này chỉ ra dông tập trung nhiều nhất vào thời điểm từ 13 đến 19 giờ, tần suất xảy ra khung giờ này tại Mộc Hóa, Long An là 57,55% Mỹ Tho, Tiền Giang là 56,19%, thời gian xảy ra trong năm thường tập trung nhiều vào tháng 5, tháng 6 và tháng 10. Việc áp dụng các chỉ số CAPE, SI, LI, KI như là công cụ hỗ trợ đánh giá khả năng dông có thể xảy ra. Như vậy, những kết quả này có ý nghĩa nhất định trong việc xác định dông khu vực Nam Bộ đặc biệt là khu vực 2 trạm nghiên cứu.

Từ khóa: Dông, Nam Bộ, hàm phân lớp (phân biệt), dự báo cảnh báo dông.

Ban Biên tập nhận bài: 8/4/2020

Ngày phản biện xong: 22/5/2020

Ngày đăng bài: 25/5/2020

1. Đặt vấn đề

Khu vực Nam Bộ hàng năm dông, sét thường xảy ra trong mùa mưa từ tháng 5 đến tháng 11, xuất hiện nhiều nhất vào thời kỳ chuyển mùa, mùa dông ở đây thường kéo dài rất lâu với khoảng 120 đến 140 ngày trong một năm. Dông là một hiện tượng thời tiết nguy hiểm xảy ra do mây đối lưu phát triển mạnh, là một nhiễu động có tính chất địa phương, thường kèm theo hiện tượng phóng điện gây nên sấm và chớp. Hiện tượng dông thường xảy ra bất ngờ, bắt đầu và chấm dứt nhanh nhưng để lại hậu quả rất nặng nề. Dông là hiện tượng thời tiết quy mô nhỏ nên việc dự báo, cảnh báo là vấn đề khó khăn, phức tạp. Bên cạnh những phương pháp truyền thống như việc sử dụng phân tích bản đồ synop, ảnh mây vệ tinh, radar thời tiết để dự báo, cảnh báo, hàm phân lớp sẽ là một cách tiếp cận mới. Hàm phân lớp được sử dụng nhiều cách tiếp cận phương pháp phân lớp nhằm cực tiểu hóa vùng

diện tích giao thoa giữa hai pha có dông và không dông. Nghiên cứu đặc điểm dông ở vùng đồng bằng Bắc Bộ [1] cũng sử dụng các phương pháp thống kê để đánh giá diễn biến, đặc điểm dông, đồng thời đưa ra những phương trình dự báo trên cơ sở thống kê số liệu khí tượng [2-3]. Một nghiên cứu khác của Tajbakhsh và cộng sự (2012) [4] đã nỗ lực thực hiện nghiên cứu các ngưỡng tối ưu xác định trước của một số chỉ số bất ổn định được sử dụng để khảo sát trước các cơn dông ở Iran kết quả của nghiên cứu này cho thấy rằng ngưỡng 30°C cho KI, -2°C cho LI và -3°C cho SI sẽ không xảy ra dông. Bên cạnh đó, [5] cũng đã nghiên cứu dự báo dông cho khu vực Việt Nam bằng chỉ số bất ổn định tính theo kết quả của mô hình ETA. Trong nghiên cứu này chúng tôi sử dụng số liệu trong khoảng thời gian 10 năm từ năm 2007 đến năm 2017 của trạm thám không vô tuyến Tân Sơn Hòa, số liệu dông được thống kê 11 năm từ năm 2007 đến năm

¹Đài Khí tượng thủy văn khu vực Nam Bộ

²Trường Đại học Khoa học tự nhiên, Đại học quốc gia thành phố Hồ Chí Minh

Email: quyet.le74@gmail.com

2018 của trạm khí tượng Mộc Hóa, Mỹ Tho. Mục tiêu nghiên cứu này nhằm đánh giá được đặc điểm đồng trên hai tỉnh Long An và Tiền Giang, xác định thời điểm nào đồng thường xuyên xuất hiện nhất, sau đó tiến hành sử dụng hàm phân lớp, trên cơ sở số liệu bất ổn định khí quyển với số liệu quan trắc đồng, để làm cơ sở dự báo, cảnh báo hiện tượng đồng.

2. Số liệu và phương pháp nghiên cứu

2.1 Phương pháp nghiên cứu

2.1.1 Giới thiệu hàm phân lớp

Hàm phân lớp được tóm tắt như sau: Giả sử có hai pha thời tiết là (phi 0) và (phi 1) cùng với x_1, \dots, x_n là các nhân tố dự báo được chia thành hai lớp tương ứng với hai pha thời tiết đó. Dựa vào chuỗi số liệu ban đầu và mục đích của dự báo hiện tượng, với mục đích là dự báo có đồng hay không có đồng nên xác định 2 lớp là có xảy ra đồng và không xảy ra đồng. Từ tập số liệu dự trữ, ta biết mỗi vector nhân tố dự báo $X = \{x_1, \dots, x_n\}$ thuộc lớp nào. Từ đó cần tìm cách giải để khi có bất kỳ một vector nhân tố dự báo nào nằm trong bộ số liệu lưu trữ thì ta cũng chỉ ra được thời tiết thuộc pha (phi 0) hay (phi 1). Từng vector nhân tố ảnh hưởng là 1 điểm trong không gian m chiều. Những điểm ứng với vector nhân tố dự báo thuộc 1 lớp sẽ tạo nên một miền nào đó trong không gian này. Để giải bài toán này ta phải thừa nhận tính phân bố chuẩn của các yếu tố khí tượng. Phương trình sử dụng là phương trình phân lớp đa nhân tố vì sử dụng 4 yếu tố là các chỉ số: SI, LI, KI, CAPE. Trong lý thuyết về đồng, có nhiều yếu tố khí tượng gây nên sự bất ổn định khí quyển. Số liệu thám không vô tuyến đưa ra nhiều chỉ số bất ổn định. Thông thường khi tính toán phải đưa các nhân tố vào tính toán, sau đó lọc ra những nhân tố có quan hệ tốt nhất (tương quan cao nhất) để chọn làm nhân tố dự báo. Nhưng trong nghiên cứu này không tiến hành tuyển chọn mà lấy trực tiếp để tính toán (dựa trên cơ sở của những nghiên cứu khác thấy 4 nhân tố này có hệ số tương quan

cao).

2.1.2 Phương pháp tính toán

Mục đích của phương pháp phân lớp là xác định tổ hợp tuyến tính của các nhân tố sao cho sự phân biệt là cực đại có thể. Bài toán dự báo sử dụng hàm phân lớp có thể có hai bước như sau:

Bước 1: Xác định không gian phân biệt từ tập nhân tố ban đầu X_1, X_2, \dots, X_m thực hiện phép biến đổi tọa độ để nhận được các biến mới sao cho biến này thỏa mãn điều kiện có sự phân chia (phân biệt) giữa các pha là lớn nhất. Có sự phân tán (độ tản mạn) trong mỗi pha là nhỏ nhất.

Bước 2: Tính xác suất có điều kiện của từng pha với điều kiện cho trước là các biến mới.

Trong trường hợp không thực hiện bước thứ nhất vẫn có thể tiến hành với các nhân tố dự báo ban đầu, tuy nhiên chất lượng dự báo có thể kém hơn.

2.1.3 Xác định không gian phân biệt

Để đơn giản trong trình bày, cách gọi được quy ước như sau: Nhóm tương đương với pha thời tiết; số nhóm tức là số pha cần xem xét. Giả thiết có G nhóm đầy đủ xung khắc của một yếu tố dự báo.

Giả sử ng là số quan trắc của nhóm g trong tập mẫu phụ thuộc. Khi đó tổng số quan trắc (kích thước mẫu phụ thuộc) là: $N = \sum_{g=1}^G n_g$ (1)

Với m nhân tố dự báo X_1, X_2, \dots, X_m Có 4 phép lấy tổng. Để xử lý trường hợp nếu giá trị phương trình rơi vào vùng giao thoa giữa 2 không gian nghiệm thì ta phải quay lại xét số liệu đưa vào. Với bộ số liệu đưa vào ta đi tính xác suất xảy ra ở từ pha (có hoặc không), cụ thể ở đây là tính xác suất có điều kiện của sự kiện xảy ra vào một thời điểm cụ thể, nếu pha nào có xác suất cao hơn thì dự báo sẽ đưa vào pha đó. Phương trình sử dụng là phương trình phân lớp đa nhân tố vì sử dụng tới 4 yếu tố là các chỉ số: SI, LI, KI, CAPE, các chỉ số thám không vô tuyến này được đưa trực tiếp vào tính toán, không qua bước lọc để tuyển chọn nhân tố có mối tương quan tốt nhất.

2.1.4 Cách tính khoảng giờ xảy ra dông nhiều nhất

Từ báo biểu số liệu của trạm Mỹ Tho, Mộc Hóa nghiên cứu thống kê được số lượng cơn

dông tương ứng theo từng khoảng giờ. Từ đó tính được tỷ lệ phân bố của dông theo từng khoảng giờ trong ngày từ công thức sau:

$$\text{Tỷ lệ \% khoảng giờ } x = \frac{\text{Tổng số dông khoảng giờ } x \text{ đó trong cả chuỗi} \times 100}{\text{Tổng số dông toàn bộ chuỗi}} \quad (2)$$

Tiếp theo nghiên cứu tiến hành tính tỷ lệ phần trăm tần suất đạt cực đại về cơn dông của từng

khoảng giờ trong toàn bộ chuỗi số liệu theo công thức sau:

$$\text{Tỷ lệ \% khung giờ } x = \frac{\text{Số năm khung giờ } x \text{ đạt cực đại cơn dông} \times 100}{\text{Tổng số năm}} \quad (3)$$

2.1.5 Cách tính tháng có dông nhiều nhất

Để tìm ra tháng có tần suất xuất hiện nhiều ngày dông nhất, nghiên cứu tiến hành tính ra tỷ

lệ phần trăm tần suất đạt cực đại về ngày dông của từng tháng theo công thức sau:

$$\text{Tỷ lệ \% tháng } x = \frac{\text{Số năm tháng } x \text{ đạt cực đại ngày dông} \times 100}{\text{Tổng số năm}} \quad (4)$$

Tiếp theo, số liệu về dông sẽ được tính trung bình theo từng tháng trên toàn bộ chuỗi số liệu để đánh giá số lượng dông trung bình trong một tháng của các năm. Diễn biến dông theo từng tháng trong toàn bộ chuỗi số liệu sẽ được biểu diễn bằng 12 đồ thị thể hiện được số lượng các cơn dông quan trắc được trong mỗi tháng. Từ đó rút ra được diễn biến dông theo 2 mùa là mùa mưa (từ tháng 5 đến tháng 10) và mùa khô (từ tháng 11 đến tháng 4 năm sau).

2.1.6 Các chỉ số dự báo dông

Chỉ số SI (Showalter Index); Chỉ số nâng LI (Lifted Index); Chỉ số KI (K Index); $KI = (T850 - T500) + Td850 - (T700 - Td700)$; Năng lượng đối lưu tiềm năng (CAPE).

2.2 Cơ sở dữ liệu

Số liệu về dông: Số liệu ngày và giờ xảy ra dông được thống kê từ năm 2007 đến năm 2018 tại các trạm Mỹ Tho, Mộc Hóa. Đối với nghiên cứu ngày, một ngày được chia thành 4 khoảng thời gian để xem xét sự xuất hiện dông: Sáng: Từ 7h đến 13h; Trưa chiều: Từ 13h đến 19h; Tối: Từ 19h đến 00h; Rạng sáng: Từ 00h đến 7h ngày hôm sau. Căn cứ vào khoảng giờ xảy ra dông để ghi nhận hiện tượng dông theo từng khoảng thời gian trong ngày. Nghiên cứu sẽ tiến hành đánh giá thống kê ngày và giờ xuất hiện của hiện tượng dông trong các chuỗi số liệu các năm. Sử dụng đồ thị để tìm ra diễn biến dông, tháng xảy ra dông nhiều nhất, khoảng thời gian xảy ra dông nhiều nhất trong một ngày.

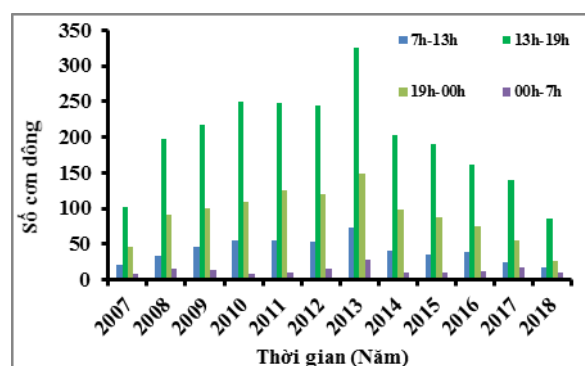
3. Kết quả nghiên cứu và thảo luận

3.1. Diễn biến dông theo thời gian

3.1.1. Kết quả tại trạm Mộc Hóa

a) Đánh giá dông theo khung giờ

Từ dữ liệu của trạm Mộc Hóa, nghiên cứu thống kê được số lượng cơn dông tương ứng theo từng khung giờ từ năm 2007 đến năm 2018 (Hình 1). Theo kết quả phân tích khung giờ từ 13-19h dông tập trung xuất hiện nhiều nhất. Căn cứ theo đánh giá số liệu dông trong từng khung giờ, thì các cơn dông đa số thường hình thành và hoạt động từ 13h tới 19h hằng ngày với 2363 cơn dông quan trắc được; tiếp theo là từ 19h tới 7h có 1087 quan trắc được trong toàn chuỗi số liệu; tiếp theo vào buổi sáng từ 7h tới 13h với 494; cuối cùng dông thường tập trung khá ít từ 00h-7h với 162 cơn dông quan trắc được trong khoảng giờ này.



Hình 1. Đồ thị biểu diễn sự phân bố dông theo từng năm từ năm 2007 đến năm 2018 ứng với từng khoảng giờ thu được từ trạm Mộc Hóa

Kết hợp với số liệu thống kê, tính được tỉ lệ phần trăm của từng khung giờ hằng ngày trong toàn bộ chuỗi số liệu trong khoảng thời gian từ năm 2007 đến 2018 có kết quả như sau:

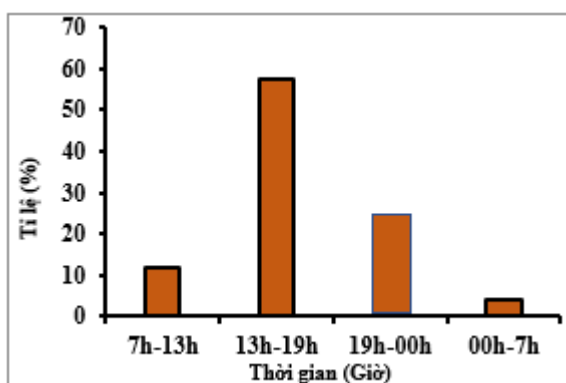
$$\% \text{ khung } 7\text{h}-13\text{h} = \frac{494 \cdot 100}{4106} = 12,03\% \quad (5)$$

$$\% \text{ khung } 13\text{h}-19\text{h} = \frac{2363 \cdot 100}{4106} = 57,55\% \quad (6)$$

$$\% \text{ khung } 19\text{h}-7\text{h} = \frac{1087 \cdot 100}{4106} = 26,47\% \quad (7)$$

$$\% \text{ khung } 00\text{h}-7\text{h} = \frac{162 \cdot 100}{4106} = 3,95\% \quad (8)$$

Tỉ lệ phần trăm từng khung giờ được biểu diễn trên Hình 2.



Hình 2. Đồ thị biểu diễn tỉ lệ phân bố đông trong ngày ứng với từng khoảng giờ từ năm 2007 đến năm 2018 thu được từ trạm Mộc Hóa

b) Đánh giá số ngày đông theo tháng

Từ số liệu đông của trạm Mộc Hóa, nghiên cứu thống kê được số lượng cơn đông tương ứng theo từng khung giờ, số ngày đông trong 12 tháng của 12 năm được biểu thị qua (Bảng 1). Tại trạm đo có thể quan sát tới 3 hoặc hơn 3 cơn đông trong 1 ngày vì thế số cơn đông trong 1 tháng cao nhất có thể vượt 80 cơn/tháng.

Tần suất xuất hiện của tháng có nhiều ngày đông nhất được tính theo công thức (9) và biểu diễn qua đồ thị (Hình 3). Hình 3 cho thấy, tháng 5 có số ngày có đông nhiều nhất trong cả chuỗi số liệu, có đến 5 lần tháng 5 đạt số ngày đông nhiều nhất trong năm chiếm 41,67% trong 12 năm; tiếp đến là tháng 10 với mỗi tháng có 4 lần có số ngày đông nhiều nhất trong năm chiếm 33,33% trong 12 năm; bên cạnh đó là tháng 6 với mỗi tháng có 2 lần có số ngày đông nhiều nhất trong năm chiếm 16,67% trong 12 năm và cuối cùng là tháng 9 với mỗi tháng có 1 lần có số ngày đông nhiều nhất trong năm chiếm 8,33% trong 12 năm. Trung bình số ngày đông của mỗi tháng trong cả chuỗi số liệu từ năm 2007 đến năm 2018 được thống kê trên bảng 1.

Bảng 1. Số ngày đông trong 12 tháng từ năm 2007 đến năm 2018 thu được từ trạm Mộc Hóa

Thời gian	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
1	0	10	11	12	13	12	12	5	5	3	3	3
2	0	23	23	23	22	23	22	4	4	4	4	1
3	0	29	28	30	33	29	28	5	5	7	5	2
4	10	30	37	36	38	45	67	34	20	14	15	7
5	23	34	38	43	46	42	66	46	45	47	42	26
6	26	39	47	45	45	46	47	41	34	33	31	17
7	25	36	38	44	40	43	42	42	39	28	25	18
8	23	28	42	45	46	45	47	42	44	38	33	17
9	28	32	40	46	46	41	75	42	33	31	28	24
10	33	43	39	38	47	46	83	44	49	44	33	23
11	9	23	24	42	44	42	38	32	30	26	13	0
12	2	12	11	18	21	19	48	14	16	12	6	0

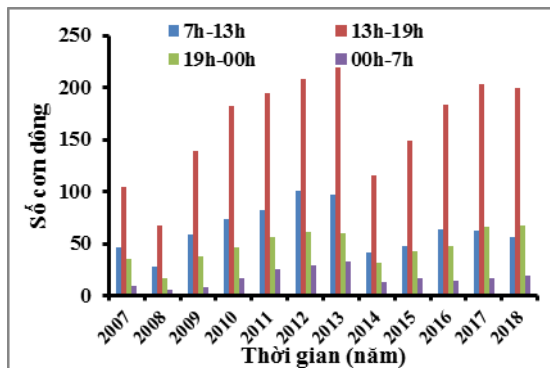
Trong một năm đông xảy ra nhiều nhất vào mùa mưa từ tháng 5 tới tháng 10 với trung bình 1 tháng có hơn 38 cơn đông. Tháng 5 có số cơn đông trung bình lên đến 41,5 cơn đông trong tháng. Tháng 6 đến tháng 9 số lượng cơn đông giảm so với tháng 5 nhưng vẫn duy trì ở mức cao, trung bình 37 đến 38 cơn 1 tháng. Những

tháng còn lại trong năm như tháng 1, 2, 3, 11, 12 vẫn có đông diễn ra nhưng do các tháng này ở vào giai đoạn mùa khô của Nam Bộ nên số lượng ngày có đông không nhiều, trung bình số ngày đông thấp hơn đáng kể so với mùa mưa với chỉ trung bình ít hơn 8 cơn đông trong 1 tháng.

3.1.2 Kết quả tại trạm Mỹ Tho

a) Đánh giá dòng theo khung giờ

Từ số liệu dòng của trạm Mỹ Tho, nghiên cứu thông kê được số lượng cơn dông tương ứng theo từng khoảng giờ từ năm 2007 đến năm 2018 (Hình 4). Từ số liệu thống kê, phân bố của dòng trong toàn bộ chuỗi thời gian theo từng khung giờ từ năm 2007 đến năm 2018 được biểu diễn bằng đồ thị (Hình 5).



Hình 4. Đồ thị biểu diễn sự phân bố dòng theo từng năm ứng với từng khoảng giờ thu được từ trạm Mỹ Tho

Căn cứ theo số liệu thống kê cho thấy rằng đa số các cơn dông thường hình thành và hoạt động từ 13h tới 19h hằng ngày với 1964 cơn dông. Tiếp theo là từ 7h đến 13h với 758 cơn dông. Tiếp đến là từ 19h đến 00h với 567 cơn dông. Và cuối cùng dông ít xuất hiện vào khoảng 00h đến 7h chỉ với 206 cơn dông quan trắc được trong toàn chuỗi số liệu. Sử dụng các công thức (5-8) kết hợp với số liệu thống kê, tỷ lệ phần trăm của từng khung giờ trong toàn bộ chuỗi số liệu từ năm 2007 đến năm 2018 có kết quả như sau:

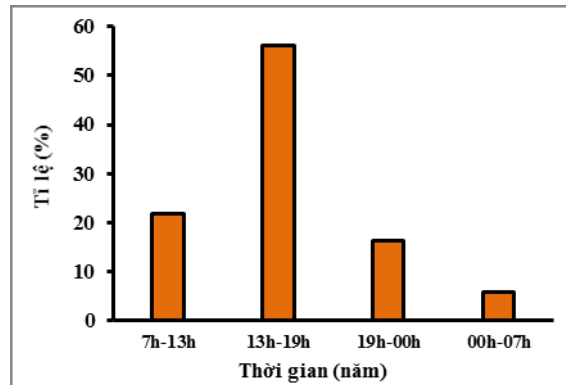
$$\% \text{ khung } 7\text{h}-13\text{h} = \frac{758 \cdot 100}{3495} = 21,69\%$$

$$\% \text{ khung } 13\text{h}-19\text{h} = \frac{1964 \cdot 100}{3495} = 56,19\%$$

$$\% \text{ khung } 19\text{h}-00\text{h} = \frac{567 \cdot 100}{3495} = 16,22\%$$

$$\% \text{ khung } 00\text{h}-7\text{h} = \frac{206 \cdot 100}{3495} = 5,89\%$$

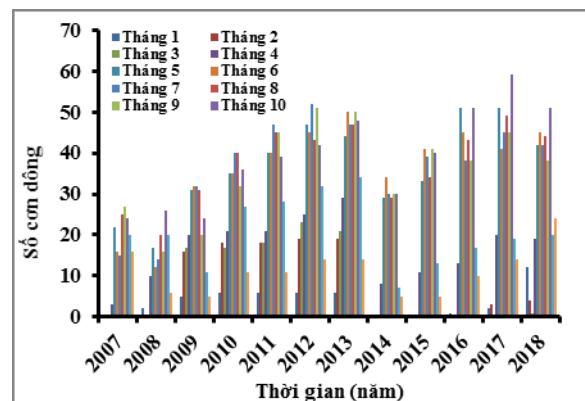
Biểu diễn kết quả sự phân bố dòng trong ngày từ năm 2007 đến năm 2018 theo từng khoảng giờ thu được từ trạm Mỹ Tho được thể hiện trên hình 5.



Hình 5. Đồ thị biểu diễn sự phân bố dòng trong ngày từ năm 2007 đến năm 2018 theo từng khoảng giờ thu được từ trạm Mỹ Tho

b) Đánh giá số ngày dông theo tháng

Từ kết quả số liệu dòng của trạm Mỹ Tho, nghiên cứu thông kê được số lượng cơn dông trong toàn chuỗi số liệu theo 12 tháng từ năm 2007 đến năm 2018, từ số liệu thống kê, phân bố của dòng trong toàn bộ chuỗi thời gian được biểu diễn bằng đồ thị (Hình 6).



Hình 6. Đồ thị biểu diễn số ngày dông trong 12 tháng từ năm 2007 đến 2018 tại trạm Mỹ Tho

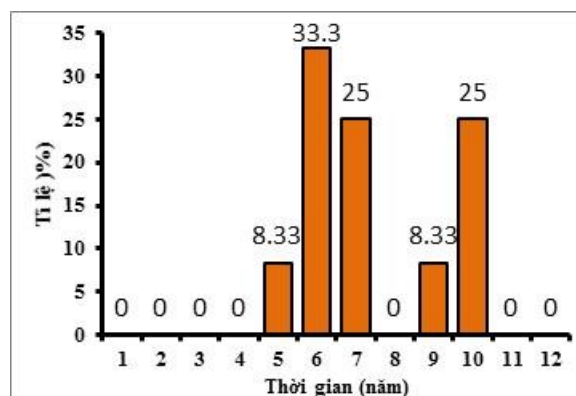
Sử dụng công thức (4) tính tỷ lệ % tần suất đạt cực đại về ngày dông từng tháng trong mỗi năm của chuỗi số liệu, kết quả biểu diễn qua đồ thị (Hình 7) ta thấy tháng 6 có số ngày dông nhiều nhất trong cả chuỗi số liệu, có 4 lần tháng 6 đạt số ngày dông nhiều nhất trong năm, chiếm 33,3% của 12 năm. Tiếp theo tháng 7 và tháng 10 có 2 lần số ngày dông đạt cực đại chiếm 25% của 12 năm. Cuối cùng là tháng 5 và tháng 9 có 1 lần số ngày dông đạt cực đại, chiếm 8,33% của 12 năm. Trung bình số ngày dông của mỗi tháng

trong cả chuỗi số liệu từ năm 2007 đến năm 2018 thu được từ trạm Mỹ Tho được thống kê. Như vậy, trong một năm dông xảy ra nhiều nhất vào mùa mưa từ tháng 5 đến tháng 10. Tháng 8 và tháng 10 có số cơn dông xảy ra nhiều nhất với trung bình lên đến 39,2 cơn dông. Các tháng 5, 6, 7, 9 có số cơn dông ít hơn so với tháng 8 và tháng 10 nhưng vẫn ở mức độ cao với trung bình khoảng 36.5 cơn dông. Tháng 10 có trung bình cơn dông cao nhất trong toàn bộ chuỗi số liệu. Với trung bình lên tới 39,2 cơn dông. Tăng cao so với tháng 7 đến tháng 9 vì đây là giai đoạn chuyển mùa. Những tháng còn lại trong năm như tháng 1, 2, 3, 11, 12 vẫn có dông diễn ra nhưng do các tháng này ở vào giai đoạn mùa khô của Nam Bộ nên số lượng ngày có dông không nhiều, trung bình số ngày dông thấp hơn đáng kể

so với mùa mưa với chỉ trung bình khoảng 11,4 cơn dông trong 1 tháng.

3.2 Dự báo dông bằng hàm phân lớp

3.2.1 Chỉ số thám không



Hình 7. Đồ thị biểu diễn tần suất xuất hiện của tháng có nhiều dông từ năm 2007 đến năm 2018 thu được từ trạm Mỹ Tho

Bảng 2. Kết quả dự báo dông theo các chỉ số

Chỉ số	SI			LI			
	SI<-4	-4≤SI≤4	SI≥4	-6≤LI≤-2	-2≤LI≤0	1≤LI≤6	LI≥6
Xác suất dự báo đúng (%)	50,0	45,45	51,72	54,91	48,85	60,67	66,67

Chỉ số	KI			CAPE			
	KI<25	25<KI<35	KI>35	CAPE<1000	1000<CAPE<1500	1500<CAPE<2000	CAPE>2000
Xác suất dự báo đúng (%)	53,49	57,81	47,79	49,86	51,85	59,77	69,12

Từ số liệu của các chỉ số thám không vô tuyến từ trạm Tân Sơn Hòa thống kê được các giá trị của chỉ số CAPE, SI, KI, LI và đưa vào chương trình dự báo dông bằng hàm phân lớp, từ đó cho kết quả xác suất dự báo đúng, các kết quả được trình bày ở Bảng 2.

3.3 Mối tương quan giữa các chỉ số thám không vô tuyến với dông

Kết quả xác suất cao nhất xảy ra dông đối với số liệu các chỉ số thám không vô tuyến được trình bày như sau: (1) Chỉ số SI (*Showalter Index*): Xác suất xảy ra dông thấp nhất là 45,45% khi SI nhận giá trị từ -4 đến 4. Xác suất xảy ra dông cao nhất là 51,72% khi SI nhận giá trị lớn hơn 4; (2) Chỉ số LI (*Lifted Index*): Xác suất xảy ra dông thấp nhất là 48,85% khi LI nhận giá trị từ -2 đến 0. Xác suất xảy ra dông cao nhất là 66,67% khi LI nhận giá trị lớn hơn hoặc bằng 6;

(3) Chỉ số KI (*K-Index*): Xác suất xảy ra dông thấp nhất là 47,79% khi KI nhận giá trị lớn hơn 35. Xác suất xảy ra dông cao nhất là 57,81% khi KI nhận giá trị từ 25 đến 35; (4) Chỉ số CAPE: Xác suất xảy ra dông thấp nhất là 49,86% khi CAPE nhận giá trị nhỏ hơn 1000. Xác suất xảy ra dông cao nhất là 69,12% khi CAPE nhận giá trị lớn hơn 2000.

3.4 Đánh giá kết quả dự báo dông

Trong nghiên cứu này chúng tôi sử dụng hai chỉ số đánh giá gồm chỉ số xác suất phát hiện (*Probability of Detection-POD*) và chỉ số tỷ phần dự báo phát hiện sai (*False Alarms Ratio-FAR*). Các công thức tính như sau:

$$POD = \frac{H}{H+M} \quad (9)$$

$$FAR = \frac{F}{H+F} \quad (10)$$

Trong đó Hits (H) là có dự báo và có quan trắc; Misses (M) là không dự báo và có quan trắc; False alarm (F) là có dự báo và không quan trắc.

POD dao động từ 0 đến 1 và giá trị tối ưu $POD = 1$. Với chỉ số $FAR = 0$ là giá trị tối ưu. Chúng tôi tiến hành dự báo thử nghiệm, chọn tháng 5 năm 2019, sau đó tiến hành đánh giá trên cơ sở số liệu tháng 5 có 22 ngày có dông, với 32 lần xuất hiện dông, 9 ngày không dông. Các kết quả thống kê được thể hiện trong bảng 3.

Bảng 3. Kết quả đánh giá khả năng dự báo dông bằng hàm phân lớp theo các chỉ số

		CAPE	LI	KI	SI
H	bd	15	15	15	15
	qt	10	6	5	5
M	bd	7	7	7	7
	qt	5	4	2	4
F	bd	15	15	15	15
	qt	5	9	8	9
POD (%)		0,67	0,63	0,68	0,64
FAR (%)		0,44	0,53	0,53	0,54

Giá trị POD đều trên 60% tức là sử dụng các chỉ số thám không vô tuyến để đưa vào hàm phân biệt dự báo dông thì xác suất phát hiện ở mức khá cao, nhưng tỉ phần dự báo phát hiện sai cũng vẫn dao động ở mức trên 50%, riêng chỉ số CAPE có thấp hơn so với 3 chỉ số còn lại.

4. Kết luận

Nghiên cứu này chỉ ra thời gian dông xảy ra trong ngày tập trung nhiều nhất vào thời điểm từ 13 đến 19 giờ, tần suất xảy ra khung giờ này tại Mộc Hóa, Long An là 57,55% Mỹ Tho, Tiền Giang là 56,19%, thời gian xảy ra trong năm thường tập trung nhiều vào tháng 5, tháng 6 và tháng 10. Hàm phân lớp có thể là một phương pháp dự báo, cảnh báo dông phù hợp, bởi trong dự báo, nhiều trường hợp xảy ra rất rõ hai pha có hoặc không dông, tuy nhiên cũng luôn tồn tại

vùng giao thoa giữa có và không có dông. Nếu giá trị phương trình rơi vào vùng giao thoa giữa 2 không gian nghiệm thì ta phải quay lại xét số liệu đưa vào. Với bộ số liệu đưa vào sẽ tính xác suất xảy ra ở từ pha (có hoặc không) cụ thể ở đây là tính xác suất có điều kiện của sự kiện xảy ra vào một thời điểm cụ thể, nếu pha nào có xác suất cao hơn thì dự báo sẽ đưa vào pha đó. Cần xác định được phép biến đổi sao cho cực đại hóa được tỷ số giữa độ phân tán giữa các nhóm và độ phân tán trong từng nhóm. Xác định được các biến mới. Mục đích của phương pháp phân lớp là xác định tổ hợp tuyến tính của các nhân tố sao cho sự phân biệt là cực đại có thể. Kết quả của nghiên cứu trước hết đánh giá chi tiết diễn biến dông xảy ra tại hai tỉnh Tiền Giang và Long An theo chuỗi số liệu từ năm 2007 đến 2018, dựa vào chuỗi số liệu xác định được thời gian dông tập trung xuất hiện nhiều nhất.

Một số đánh giá cụ thể: 1) Vào thời kì mùa mưa, thời gian dông thường xảy ra là 13h đến 19h, trong ngày, tiếp theo là khoảng thời gian từ 19h đến 00h và thời gian dông xảy ra ít nhất là 00h đến 7h sáng hôm sau; 2) Các tháng có nhiều dông nhất gồm tháng 5, tháng 6 và tháng 10; 3) Sử dụng các chỉ số CAPE, SI, LI, KI đưa chương trình Fortran tính toán, dự báo, cảnh báo. Dùng các chỉ số thám không vô tuyến và chương trình tính toán xác suất dự báo đúng của các chỉ số bằng hàm phân lớp để có thể dự báo dông. Những khoảng giá trị sau đây sẽ có khả năng xảy ra dông cao: (1) Giá trị chỉ số SI: Từ 4 trở lên sẽ cho khả năng xảy ra dông là 51,52%; (2) Giá trị chỉ số LI: Từ 6 trở lên sẽ cho khả năng xảy ra dông là 66,67%; (3) Giá trị chỉ số KI: Thuộc khoảng từ 25 đến 35 sẽ cho khả năng xảy ra dông là 57,81%; (4) Giá trị chỉ số CAPE: Từ $1500 < CAPE < 2000$ là 59,77%; trên 2000 trở lên sẽ cho khả năng xảy ra dông là 69,12%.

Lời cảm ơn: Nhóm tác giả xin chân thành cảm ơn các đơn vị của Đài Khí tượng Thủy văn khu vực Nam Bộ, đề tài “Nghiên cứu xây dựng bộ công cụ dự báo, cảnh báo sớm mưa, lũ, dông khu vực đồng bằng sông Cửu Long” đã tạo điều kiện để nhóm tác giả tiếp cận, tìm hiểu các hệ thống dự báo, cảnh báo thời tiết, cung cấp số liệu cao không - mặt đất để nghiên cứu.

Tài Liệu Tham Khảo

1. Nguyễn Viết Lành (2002), *Nghiên cứu dự báo dông nhiệt vùng Đồng bằng Bắc Bộ trong các tháng nửa đầu mùa hè*. Luận án Tiến sĩ Khí tượng, Viện Khí tượng - Thủy văn, Hà Nội.
2. Trần Tân Tiến, Nguyễn Đăng Quế (2001), *Xử lý số liệu khí tượng và dự báo thời tiết bằng phương pháp thống kê vật lý*. Nhà xuất bản Đại học quốc gia Hà Nội.
3. Phan Văn Tân (2003), *Các phương pháp thống kê trong khí hậu*. Nhà xuất bản Đại học quốc gia Hà Nội, 260 trang.
4. Tajbakhsh, S., Ghafarian, P., Sahraian, F., (2012), *Instability indices and forecasting thunderstorms: the case of 30 April 2009*. Natural hazards and Earth System Sciences, 12(2), 403-413.
5. Trần Tân Tiến, Nguyễn Khánh Linh (2004), *Dự báo dông cho khu vực Việt Nam bằng chỉ số bất ổn định tính theo kết quả của mô hình ETA*. Tạp chí Khoa Học, ĐHQG Hà Nội, 3PT, 90-100.

Statistics for evaluation of thunderstorm development in Moc Hoa, My Tho and using classification function for thunderstorm forecast and warning

Le Dinh Quyet¹, Le Ngoc Quyen¹, Tran Thi Thu Uyen²

¹Southern Regional Hydro-Meteorological Center

²University of Science Ho Chi Minh City

Abstract: Statistical studies assessing thunderstorm activity are important as a basis for forecasting and warning activities indicating thunderstorm time during the day. This study focuses on analysis and evaluation based on data of two stations Moc Hoa, Long An province and My Tho, Tien Giang province with data period from 2007-2018. The classification function is applied to evaluate for the forecast. The results of this study indicated that thunderstorms were most concentrated from 13 p.m to 19 p.m, the frequency of this time frame in Moc Hoa, Long An was 57.55% My Tho, Tien Giang was 56.19%. During the year, thunderstorms occurs mainly in May, June and October. The application of CAPE, SI, LI, KI indicators as a supportive tool for assessing possible thunderstorms. Thus, these results have a certain significance in identifying thunderstorms in the Southern region, especially in the area of 2 research stations.

Keywords: Thunderstorm, Southern region of Vietnam, Classification function, Thunderstorm forecast and warning.