



Original Article

Seasonal Predictions of the Number of Tropical Cyclones in the Vietnam East Sea Using Statistical Models

Dinh Ba Duy¹, Ngo Duc Thanh², Tran Quang Duc³, Phan Van Tan^{3,*}

¹*Vietnam-Russia Tropical Center, 63 Nguyen Van Huyen, Nghia Do, Cau Giay, Hanoi, Vietnam*

²*University of Science and Technology of Hanoi, Vietnam Academy of Science and Technology, 18 Hoang Quoc Viet, Cau Giay, Hanoi, Vietnam*

³*VNU University of Science, 334 Nguyen Trai, Thanh Xuan, Hanoi, Vietnam*

Received 04 April 2019

Revised 19 April 2019; Accepted 19 May 2019

Abstract: In this study, the equations for estimating the number of tropical cyclones (TCs) at a 6-month lead-time in the Vietnam East Sea (VES) have been developed and tested. Three multivariate linear regression models in which regression coefficients were determined by different methods, including 1) method of least squares (MLR), 2) minimum absolute deviation method (LAD), 3) minimax method (LMV). The artificial neural network model (ANN) and some combinations of the above regression models were also used. The VES was divided into the northern region above 15°N (VES_N15) and the southern one below that latitude (VES_S15). The number of TCs was calculated from the data of the Japan Regional Specialized Meteorological Center (RMSC) for the period 1981-2017. Principal components of the 14 climate indicators were selected as predictors. Results for the training period showed that the ANN model performed best in all 12 times of forecasts, following by the ANN-MLR combination. The poorest result was obtained with the LMV model. Results for the independent dataset showed that the number of adequate forecasts based on the MSSS scores decreased sharply compared to the training period and the models generated generally similar errors. The MLR model tended to give out the best results. Better-forecast results were obtained in the VES_N15 region followed by the VES and then the VES_S15 regions.

Keywords: Tropical cyclone, Seasonal prediction, Vietnam East Sea (VES).

* Corresponding author.

E-mail address: phanvantant@hus.edu.vn

<https://doi.org/10.25073/2588-1094/vnuees.4379>



Dự báo hạn mùa số lượng xoáy thuận nhiệt đới trên Biển Đông bằng các mô hình thống kê

Đình Bá Duy¹, Ngô Đức Thành², Trần Quang Đức³, Phan Văn Tân^{3,*}

¹Trung tâm Nhiệt đới Việt–Nga, 63 Nguyễn Văn Huyền, Nghĩa Đô, Cầu Giấy, Hà Nội, Việt Nam

²Trường Đại học Khoa học và Công nghệ Hà Nội, Viện Hàn lâm Khoa học và Công nghệ Việt Nam, 18 Hoàng Quốc Việt, Cầu Giấy, Hà Nội, Việt Nam

³Trường Đại học Khoa học Tự nhiên, Đại học Quốc gia Hà Nội, 334 Nguyễn Trãi, Thanh Xuân, Hà Nội, Việt Nam

Nhận ngày 04 tháng 4 năm 2019

Chỉnh sửa ngày 19 tháng 4 năm 2019; Chấp nhận đăng ngày 19 tháng 5 năm 2019

Tóm tắt: Trong nghiên cứu này các phương trình dự báo số lượng xoáy thuận nhiệt đới (XTNĐ) hạn 6 tháng trên khu vực Biển Đông (VES) đã được xây dựng và thử nghiệm. Ba mô hình hồi quy tuyến tính đa biến trong đó các hệ số hồi quy được xác định bằng các phương pháp khác nhau, gồm 1) bình phương tối thiểu (MLR), 2) độ lệch tuyệt đối nhỏ nhất (LAD), 3) minmax (LMV), và mô hình mạng thần kinh nhân tạo (ANN) cộng với một số tổ hợp các mô hình trên với nhau được sử dụng. Khu vực Biển Đông được chia thành 2 vùng phía bắc (VES_N15) và phía nam (VES_S15) vĩ tuyến 15 độ. Yếu tố dự báo được tính từ tập số liệu XTND tại Trung tâm Khí tượng chuyên vùng của Nhật Bản giai đoạn 1981-2017. Các thành phần chính của 14 chỉ số khí hậu được lựa chọn làm nhân tố dự báo. Kết quả đánh giá sai số dự báo trên bộ số liệu phụ thuộc cho thấy mô hình ANN cho kết quả dự báo tốt nhất ở cả 12 thời điểm dự báo, tiếp đó là các kết quả tổ hợp với mô hình ANN, MLR và kém nhất ở mô hình LMV. Kết quả thử nghiệm dự báo trên bộ số liệu độc lập cho thấy số lượng các sơ đồ dự báo “đạt” theo ngưỡng chỉ số MSSS giảm so với thời kỳ dự báo phụ thuộc và nhìn chung sai số của các mô hình có sự tương đồng nhau. Mô hình MLR có xu hướng cho kết quả dự báo tốt nhất. Kết quả dự báo tốt hơn cũng nhận được trên vùng VES_N15, tiếp theo là vùng VES và thấp nhất trên vùng VES_S15.

Keywords: Tropical cyclone, Seasonal prediction, Vietnam East Sea (VES).

1. Mở đầu

Cho đến nay đã có khá nhiều công trình nghiên cứu dự báo hạn mùa bão, áp thấp nhiệt

đới (gọi chung là xoáy thuận nhiệt đới - XTND) đã được thực hiện. Hầu hết trong số đó tập trung vào hai ổ bão lớn nhất thế giới là ổ bão Tây bắc Đại Tây dương và ổ bão Tây bắc Thái Bình

*Tác giả liên hệ.

Địa chỉ email: phanvantan@hus.edu.vn

<https://doi.org/10.25073/2588-1094/vnuees.4379>

đương trong đó có khu vực Biển Đông. Về phương pháp dự báo có thể chia thành ba nhóm chính: 1) Nhóm phương pháp thống kê truyền thống hay thống kê kinh điển (TKTT); 2) Nhóm phương pháp thống kê - động lực và 3) Nhóm phương pháp mô hình động lực.

Dự báo bằng phương pháp TKTT là dựa trên mối quan hệ thống kê giữa yếu tố dự báo và tập các nhân tố dự báo được xây dựng từ một tập số liệu phụ thuộc trong quá khứ, sau đó áp dụng mối quan hệ này cho tương lai. Các mô hình TKTT được sử dụng rộng rãi trong bài toán dự báo hoạt động của XTNĐ [1-6]. Trong các công trình này, yếu tố dự báo là số lượng và số ngày hoạt động của XTNĐ trên các vùng khác nhau, chủ yếu ở Đại Tây Dương và Tây Thái Bình Dương. Các nhân tố dự báo có thể là những nhân tố thuộc nhóm ENSO, như dị thường nhiệt độ bề mặt biển (SSTA) trên các vùng Niño1+2, 3, 4, 3.4, hoặc các đặc trưng hoàn lưu khí quyển và đại dương qui mô lớn như chỉ số dao động tựa hai năm tầng bình lưu (QBO), chỉ số dao động nam (SOI), v.v. Xuất phát từ cơ sở này hoạt động của XTNĐ trên khu vực Tây Bắc Thái Bình Dương cũng đã được một số tác giả quan tâm [7-9]. Bên cạnh đó, phương pháp hồi quy từng bước nhiều biến cũng đã được sử dụng để lọc nhân tố dự báo nhằm dự báo hạn mùa sự hoạt động của XTNĐ [10]. Nhìn chung, các tác giả không chỉ dự báo số lượng XTNĐ mà còn dự báo cả số lượng các cơn bão mạnh cũng như chỉ số “hoạt động bão chuẩn hoá” - NTA (Normalized Typhoon Activity). Các nhân tố dự báo được lựa chọn theo nguyên tắc làm cực tiểu hoá sai số dự báo chứ không đơn thuần dựa trên sự phù hợp với số liệu quá khứ. Kết quả của nghiên cứu cho thấy xác suất dự báo dựa trên tổ hợp các mô hình thống kê là hợp lý so với thực tế. Ngoài ra, còn có thể kể đến các công trình sử dụng mô hình thống kê để dự báo tần suất hoạt động hoặc vùng hoạt động của bão được trình bày tại [12-16] v.v.

Đối với Việt Nam, các công trình dự báo XTNĐ sớm nhất thuộc lớp dự báo thời tiết chủ yếu nằm trong khoảng những thập kỷ 1970-1980 và thường tập trung vào các mô hình thống kê và một số thử nghiệm bằng mô hình số trị, đặc biệt sau khi các mô hình số bắt đầu được ứng dụng,

cả các mô hình chính áp hai chiều và mô hình ba chiều đầy đủ. Trong số các công trình nghiên cứu về XTNĐ ở Việt Nam, ảnh hưởng của ENSO đến hoạt động của XTNĐ trên Biển Đông là chủ đề được nhiều tác giả quan tâm [17]. Xu thế hoạt động của bão cũng được đề cập đến [18-23]. Dự báo mùa sự hoạt động của bão trên khu vực Tây Bắc Thái Bình Dương và Biển Đông đã được Nguyễn Văn Tuyên (2008) trình bày khá chi tiết [24, 25]. Trên cơ sở công cụ thống kê với bộ nhân tố dự báo là các chỉ số khí hậu tác giả đã khảo sát khả năng dự báo hạn mùa số lượng bão cho khu vực Biển Đông. Có thể nói đây là một trong những công trình đáng chú ý nhất về dự báo mùa số lượng bão bằng phương pháp thống kê. Việc nghiên cứu dự báo hạn mùa số lượng XTNĐ cũng đã được đề cập ở chừng mực nhất định thông qua việc các tác giả đã chỉ ra mối quan hệ giữa sự tăng nhiệt độ mặt nước biển trung bình mùa bão với hoạt động của XTNĐ trên Biển Đông [26].

Bài báo này sẽ trình bày một số kết quả dự báo số lượng XTNĐ hoạt động trên khu vực Biển Đông bằng các mô hình thống kê khi sử dụng các chỉ số khí hậu làm nhân tố dự báo. Phương pháp và số liệu được sử dụng trong bài báo được trình bày trong mục 2. Mục 3 là những kết quả nghiên cứu chính. Một vài kết luận được đưa ra trong mục 4.

2. Số liệu và phương pháp nghiên cứu

2.1. Nguồn số liệu

Số lượng XTNĐ trên khu vực Biển Đông (VES) được khai thác từ bộ số liệu quỹ đạo bão của Trung tâm Khí tượng chuyên vùng Nhật Bản (RSMC) tại [27]. Toàn bộ khu vực VES được chia thành hai vùng là phía bắc và phía nam vĩ tuyến 15 độ, tương ứng ký hiệu là VES_N15 và VES_S15. Trên mỗi một vùng, XTNĐ được tổng hợp cho từng tháng làm yếu tố dự báo, theo đó yếu tố dự báo của tháng hiện tại là tổng số XTNĐ đạt cường độ từ áp thấp nhiệt đới trở lên (tức tốc độ gió mạnh nhất đạt từ cấp 6 trở lên) của 6 tháng tiếp theo. Chẳng hạn, VES_N15_01 là tổng số XTNĐ từ tháng 2 đến tháng 7, VES_N15_02 là

tổng số XTNĐ từ tháng 3 đến tháng 8 của vùng Biển Đông nằm ở phía bắc vĩ tuyến 15 độ, v.v. Do việc phân vùng hoạt động nên một cơn bão nếu xuất hiện ở cả hai vùng phía bắc và phía nam vĩ tuyến 15 độ sẽ được tính là hai cơn.

Số liệu các chỉ số khí hậu (ClimIDX) được khai thác tại website của Cơ quan quản lý biển và khí quyển Hoa Kỳ [28]. Trong số các chỉ số ClimIDX này, một số có độ dài chuỗi từ khoảng

1979-1980 đến nay, một số không được cập nhật thường xuyên. Với mục đích sử dụng các chỉ số như là những nhân tố dự báo nghiệp vụ, trong nghiên cứu này 14 chỉ số được lựa chọn như thể hiện trong Bảng 1. Cả hai tập số liệu yếu tố dự báo và 14 nhân tố dự báo được chia làm hai giai đoạn: 1) tập số liệu phụ thuộc từ 1981-2010 (30 năm) và 2) tập số liệu độc lập từ 2010-2017 (8 năm).

Bảng 1. Danh sách 14 chỉ số khí hậu được lựa chọn làm nhân tố dự báo

TT	Chỉ số	Mô tả
1	ONI	Chỉ số Niño đại dương (Oceanic Niño Index)
2	SOI	Chỉ số dao động Nam (Southern Oscillation Index)
3	Nino3.4	Nhiệt độ bề mặt biển (SST) phía Đông trung tâm Thái Bình Dương nhiệt đới (East Central Tropical Pacific SST)
4	MEI	Chỉ số ENSO đa biến (Multivariate ENSO Index)
5	RINDO_SLPA	Dị thường khí áp mực biển vùng xích đạo gần khu vực Indonesia (Equatorial SOI Indonesia SLP (Standardized Anomalies)
6	REQSOI	Dị thường dao động Nam vùng xích đạo (Equatorial SOI (Standardized Anomalies))
7	BEST	Chỉ số ENSO theo chuỗi thời gian (Bivariate ENSO Timeseries)
8	PDO	Dao động thập kỷ Thái Bình Dương (Pacific Decadal Oscillation)
9	DMI	Chỉ số dao động lưỡng cực (Dipole Mode Index)
10	TNI	Chỉ số khuynh hướng Niño hay chỉ số Niño-chuyển dịch (Trans-Niño Index)
11	PNA	Chỉ số Bắc Mỹ Thái Bình Dương (Pacific North American Index)
12	WHWP	Vực nóng Bán cầu Tây (Western Hemisphere Warm Pool)
14	WP	Chỉ số Tây Thái Bình Dương (Western Pacific Index)
14	QBO	Dao động tựa 2 năm (Quasi-Biennial Oscillation)

2.2. Phương pháp nghiên cứu

Mô hình thống kê truyền thống: Ba mô hình TKTT được sử dụng trong nghiên cứu này đều là các mô hình hồi quy tuyến tính đa biến:

$$y = a_0 + \sum_{j=1}^m a_j x_j \quad (1)$$

Trong đó y là yếu tố dự báo, x_j với $j = 1, 2, \dots, m$ là các nhân tố dự báo, các hệ số hồi quy $a_0, a_1, a_2, \dots, a_m$ được xác định theo ba phương pháp khác nhau:

Phương pháp bình phương tối thiểu (MLR):

$$\sum_{i=1}^n (y_i - a_0 - \sum_{j=1}^m a_j x_{ij})^2 \rightarrow \min \quad (2)$$

Phương pháp độ lệch tuyệt đối nhỏ nhất (LAD):

$$\sum_{i=1}^n |y_i - a_0 - \sum_{j=1}^m a_j x_{ij}| \rightarrow \min \quad (3)$$

Phương pháp minimax (LMV):

$$\min(\max |y_i - a_0 - \sum_{j=1}^m a_j x_{ij}|, i = 1, 2, \dots, n) \quad (4)$$

Trong các công thức (2)-(4), y_i, x_{ij} tương ứng là giá trị quan trắc thứ i ($i=1, 2, \dots, n$) của yếu tố dự báo y và các nhân tố dự báo x_j ($j=1, 2, \dots, m$).

Mô hình mạng thần kinh nhân tạo (ANN): Bên cạnh 3 mô hình TKTT kể trên, nghiên cứu này còn thử nghiệm với mô hình mạng thần kinh nhân tạo (ANN). ANN là sự mô phỏng cấu trúc và hoạt động của bộ não người, được hình thành từ những đơn vị riêng lẻ gọi là những tế bào thần kinh nhân tạo (hay neuron nhân tạo). Nghiên cứu này sử dụng loại ANN phổ biến là mạng feed-forward với một lớp ẩn, các neuron thuộc lớp ẩn nhận hàm kích hoạt dạng sigma với hệ số $a=1$, các neuron thuộc lớp đầu ra không sử dụng hàm kích hoạt. Tổng các sai số E trên tất cả các neuron đầu ra được xác định qua biểu thức:

$$E(\vec{w}) = \frac{1}{2} \sum_{d \in D} \sum_{k \in \text{outputs}} (t_{kd} - o_{kd})^2 \quad (5)$$

Ở đây D là tập số liệu luyện, t_{kd} là giá trị đích (giá trị mong muốn), o_{kd} là giá trị đầu ra của ví dụ luyện thứ d . Mục đích của nghiên cứu là cực tiểu hóa E trong không gian \vec{w} . Quá trình cực tiểu hóa này (còn được hiểu là quá trình học của ANN) được tiến hành theo thuật toán lan truyền ngược [30]. Kết quả cuối cùng sẽ là tập các trọng số w và áp dụng mạng ANN thu được vào dự báo.

Để giảm bớt số lượng nhân tố dự báo có mặt trong các phương trình (1) nhưng vẫn bảo toàn được lượng thông tin cần thiết, kỹ thuật phân tích thành phần chính (PCA) được áp dụng cho tập số liệu ClimIDX. Ký hiệu $\mathbf{X}(n, m)$ là ma trận số liệu phụ thuộc của các ClimIDX, trong đó n là độ dài chuỗi thời gian, m là số chỉ số ClimIDX ($m=14$). Có thể biểu diễn $\mathbf{X}(n, m)$ dưới dạng:

$$\mathbf{X}(n, m) \approx \text{EOF}(n, k) \cdot \text{PC}(k, m), k \ll m \quad (6)$$

Trong đó $\text{EOF}(n, k)$ là k hàm trực giao thực nghiệm, $\text{PC}(k, m)$ là k thành phần chính được giữ lại. Khi đó thay $m=14$ chỉ số \mathbf{X} ban đầu làm nhân tố dự báo là $k \ll m$ hàm trực giao thực nghiệm EOF được sử dụng. Ưu điểm của việc sử dụng các EOF làm nhân tố dự báo là chúng không tương quan với nhau. Khi tiến hành dự báo trên chuỗi số liệu độc lập, ứng với mỗi vector nhân tố dự báo ban đầu $\mathbf{X}(i, m)$, xem $\text{PC}(k, m)$ là không đổi, có thể nhận được giá trị các nhân tố dự báo $\text{EOF}(i, k)$ như sau:

$$\text{EOF}(i, k) = \mathbf{X}(i, m) \cdot \text{PC}'(m, k) \quad (7)$$

Trong đó $\text{PC}'(m, k)$ là ma trận chuyển vị của $\text{PC}(k, m)$.

Mô hình tổ hợp: Từ 4 mô hình thống kê kể trên (MLR, LAD, LMV và ANN), để tăng thêm các trường hợp khảo sát, nghiên cứu này đã xem xét thêm đối với các tổ hợp giữa mô hình MLR và LAD kí hiệu là E12; tương tự như vậy là E123 (tổ hợp của các mô hình MLR, LAD, LMV), E124 (tổ hợp MLR, LAD, ANN), E14 (tổ hợp MLR, ANN) và E1234 (tổ hợp MLR, LAD, LMV, ANN).

2.3. Phương pháp đánh giá chất lượng mô hình dự báo

Để đánh giá chất lượng các mô hình dự báo, nghiên cứu sử dụng các chỉ số sai số trung bình (ME), sai số tuyệt đối trung bình (MAE), sai số bình phương trung bình (RMSE) và hệ số tương quan (Corr) giữa giá trị dự báo và giá trị quan trắc. Với kí hiệu F_i và O_i tương ứng là giá trị dự báo của mô hình và giá trị quan trắc về số lượng XTNĐ của 6 tháng liên kế sau đó (trong đó $i=1, 2, \dots, n$, với n là dung lượng mẫu), các sai số này được tính toán như sau:

Sai số trung bình:

$$ME = \frac{1}{n} \sum_{i=1}^n (F_i - O_i) \quad (8)$$

Sai số tuyệt đối trung bình:

$$MAE = \frac{1}{n} \sum_{i=1}^n |F_i - O_i| \quad (9)$$

Sai số bình phương trung bình:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (F_i - O_i)^2} \quad (10)$$

Hệ số tương quan:

$$\text{Corr} = \frac{\sum_{i=1}^n (F_i - \bar{F})(O_i - \bar{O})}{\sqrt{\sum_{i=1}^n (F_i - \bar{F})^2} \sqrt{\sum_{i=1}^n (O_i - \bar{O})^2}} \quad (11)$$

Ngoài 4 chỉ số trên, để đánh giá kết quả dự báo của mô hình với dự báo khí hậu, nghiên cứu này xem xét thêm điểm kỹ năng MSSS của các phương trình dự báo, MSSS được tính toán dựa trên sai số bình phương trung bình:

$$MSSS = 1 - \frac{\frac{1}{n} \sum_{i=1}^n (F_i - \bar{O}_i)^2}{\frac{1}{n} \sum_{i=1}^n (O_i - \bar{O}_{cl})^2} \quad (12)$$

Ở đây, $\overline{O_{cl}}$ là giá trị trung bình khí hậu, được tính trên chuỗi số liệu giai đoạn 1981-2010 (30 năm). Giá trị MSSS sẽ nằm trong khoảng $(-\infty, 1)$ với giá trị 1 có nghĩa mô hình hoàn hảo trong khi giá trị nhỏ hơn hoặc bằng 0 có nghĩa mô hình không có kỹ năng hay chính xác hơn mô phỏng, dự báo của mô hình không tốt hơn so với dự báo khí hậu. Ở nghiên cứu này xác định tiêu chí “đạt” đối với một mô hình khi chỉ số MSSS >0 tức là mô hình “đạt” khi có kỹ năng dự báo cao hơn dự báo khí hậu.

3. Kết quả và thảo luận

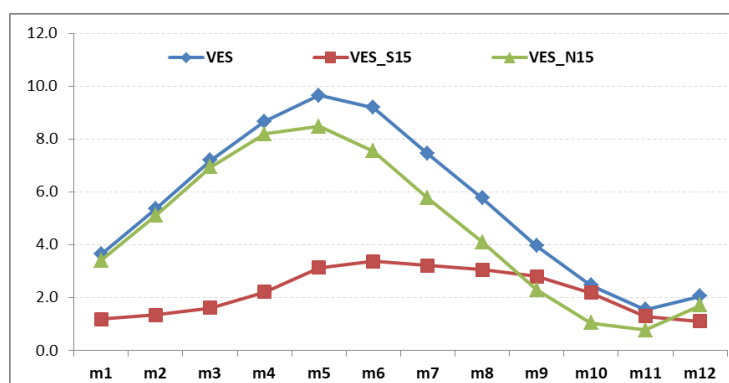
3.1. Một số đặc điểm số lượng xoáy thuận nhiệt đới 6 tháng liên tiếp sau tháng dự báo

Số lượng XTNĐ trung bình tích lũy 6 tháng liên tiếp tính từ sau tháng dự báo (TCs_6mon) trên khu vực VES và VES_N15, VES_S15 được thể hiện tại hình 1 với các giá trị đặc trưng chỉ ra tại bảng 2.

Các kết quả cho thấy phần lớn TCs_6mon thống kê được ở khu vực VES cũng được ghi nhận xuất hiện tại vùng VES_N15 với những diễn biến tương đồng nhau ở các thời điểm các tháng trong năm, đặc biệt là ở thời điểm từ tháng 1 tới tháng 4 hàng năm. Nhìn chung số lượng XTNĐ của 6 tháng liên trước thời điểm dự báo tại 3 vùng có sự khác biệt rõ rệt hơn so số liệu XTNĐ hàng tháng, tương ứng với từng tháng trong năm giá trị này cao nhất trên khu vực VES và thấp nhất tại vùng VES_S15. Mặc dù vậy, đồ

thị TCs_6mon vẫn duy trì sự tương đồng giữa khu vực VES và VES_N15, trong đó tập trung phần lớn vào các tháng 4, 5, 6 và đạt cực đại vào tháng 5 trong khi TCs_6mon tại vùng VES_S15 có xu hướng tập trung ở các tháng 5, 6, 7 và 8, với mức chênh lệch giữa các thời điểm là không lớn và thấp hơn so với 2 khu vực còn VES và VES_N15 (thể hiện chỉ có một phần số ít XTNĐ xuất hiện tại VES dịch chuyển xuống phía Nam vĩ tuyến 15 độ trong các tháng cuối năm (tháng 10, 11 và 12).

Như vậy có thể thấy rằng, việc lựa chọn TCs_6mon như là các nhân tố dự báo thì để dự báo cho mùa bão chính (tháng 6 tới tháng 11 hàng năm) ta quan tâm tới sơ đồ dự báo tại tháng 5 hay để xem xét những XTNĐ “bất thường” xuất hiện vào những thời điểm được cho là không phải mùa bão ta nghiên cứu tới sơ đồ dự báo tháng 11 và để dự báo cho 1 tháng thì ta tiến hành dự báo 2 thời điểm tháng liền trước sau khi đã xử lý hiệu chỉnh với kết quả đã có của tháng trước đó. Theo kinh nghiệm của các nhà dự báo hoạt động bão thì do kết quả dự báo XTNĐ nhìn chung chưa cao nên cùng với giá trị dự báo các giá trị trung bình nhiều năm hay độ lệch chuẩn của tập luyện để tiện cân nhắc khi phát báo hoặc người dùng có thể dễ dàng cân nhắc trong từng ứng dụng cụ thể [24, 25, 30-32]. Tại nghiên cứu này, các giá trị đặc trưng thống kê số lượng XTNĐ tích lũy 6 tháng liên tiếp sau tháng dự báo (TCs_6mon) cho các vùng VES, VES_N15, VES_S15 được tính toán và chỉ ra tại bảng 2.



Hình 1. Số lượng TCs_6mon trung bình tháng giai đoạn 1981-2017 của các vùng VES, VES_S15 và VES_N15.

Bảng 2. Đặc trưng thống kê yếu tố dự báo TCs_6mon trên khu vực VES, VES_N15, VES_S15 theo từng tháng tương ứng là các thời điểm dự báo của tháng 1 tới tháng 12

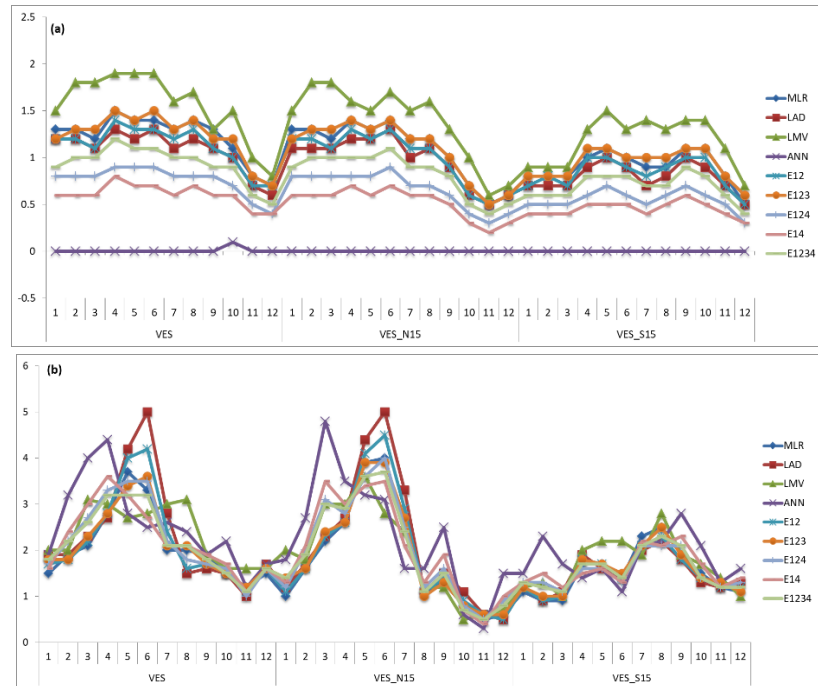
Số liệu	Đặc trưng thống kê	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12
VSE	Cực đại	7.0	10.0	12.0	13.0	16.0	17.0	13.0	11.0	7.0	6.0	4.0	5.0
	Cực tiểu	1.0	2.0	3.0	4.0	6.0	5.0	4.0	2.0	0.0	0.0	0.0	0.0
	Trung bình	3.6	5.4	7.2	8.7	9.6	9.2	7.4	5.8	4.0	2.5	1.6	2.1
	Trung vị	4.0	5.0	7.0	8.0	9.0	9.0	7.0	5.5	4.0	2.0	1.0	2.0
VES_S15	Cực đại	4.0	4.0	5.0	6.0	7.0	6.0	6.0	7.0	6.0	6.0	4.0	4.0
	Cực tiểu	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Trung bình	1.2	1.4	1.6	2.2	3.1	3.4	3.2	3.1	2.8	2.2	1.3	1.1
	Trung vị	1.0	1.0	1.0	2.0	3.0	3.0	3.0	3.0	3.0	2.0	1.0	1.0
VES_N15	Cực đại	7.0	10.0	12.0	12.0	14.0	12.0	9.0	9.0	5.0	4.0	2.0	4.0
	Cực tiểu	1.0	2.0	3.0	4.0	4.0	3.0	2.0	1.0	0.0	0.0	0.0	0.0
	Trung bình	3.4	5.1	6.9	8.2	8.5	7.5	5.8	4.1	2.3	1.1	0.8	1.7
	Trung vị	3.0	5.0	7.0	8.0	8.0	7.0	6.0	4.0	2.0	1.0	1.0	2.0

3.2. Đánh giá chất lượng của các mô hình thống kê dự báo hạn mùa số lượng xoáy thuận nhiệt đới trên Biển Đông

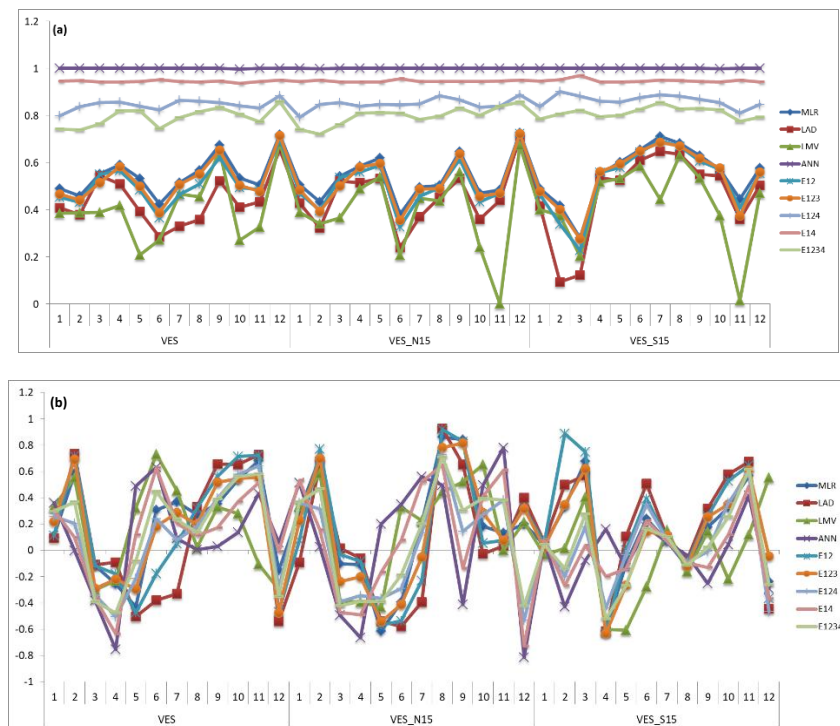
Sử dụng 4 mô hình thống kê (MLR, LAD, LMV, ANN) và 5 mô hình tổ hợp (E12, E123, E124, E14, E1234) như đã trình bày ở trên nghiên cứu đã tiến hành dự báo hạn 6 tháng số lượng TCs_6mon hoạt động trên các vùng VES, VES_S15 và VES_N15 trên tập số liệu phụ thuộc (giai đoạn 1981-2010) và trên bộ số liệu độc lập (giai đoạn 2010-2017). Sai số MAE và hệ số tương quan (Correlation) được dẫn ra tại các hình 2, 3 dưới đây (sai số ME, RMSE được tính toán nhưng không hiển thị ở đây).

Tổng thể chung qua các giá trị sai số có thể thấy rõ trên bộ số liệu phụ thuộc mô hình ANN cho kết quả dự báo tốt nhất ở cả 12 thời điểm dự báo (tương ứng với 12 tháng), thậm chí ME, MAE và RMSE ở mô hình này đạt tới gần mức giá trị lý tưởng (giá trị sai số ME, MAE và RMSE gần bằng 0 và hệ số tương quan gần bằng 1), tiếp đó là các kết quả tổ hợp với mô hình

ANN. Ở nhóm các mô hình thống kê truyền thống, mô hình MLR có kết quả tốt hơn cả, tiếp đến là mô hình tổ hợp E12 (MLR với LAD) và kết quả dự báo kém nhất khi sử dụng mô hình LMV (hình 2, hình 3). Kết quả các chỉ số sai số này có sự thay đổi khi thử nghiệm dự báo trên bộ số liệu độc lập, các giá trị ME, MAE và RMSE ở mô hình ANN và LMV cho thấy mức độ hiệu quả giảm đi nhiều so với thời kỳ luyện trên bộ số liệu phụ thuộc, nhìn chung giá trị sai số của các mô hình còn lại thể hiện tương đối giống nhau, giá trị sai số MAE, RMSE trên khu vực VES và VES_N15 cao ở những thời điểm dự báo cho mùa bão chính (thời điểm dự báo tại tháng 4, 5, 6) trong khi ở khu vực VES_S15 các giá trị này có phần tăng nhẹ ở những tháng cuối năm. Bên cạnh đó cũng cần lưu ý sự xuất hiện các thời điểm tương quan âm (hệ số Corr < 0) của các sơ đồ dự báo có sự thay đổi theo các vùng dự báo, đối với vùng VES các sơ đồ dự báo ở thời điểm từ tháng 3 đến tháng 5 có xu hướng cho kết quả tương quan âm, vùng VES_N15 ở các thời điểm tháng 3 đến tháng 7 và vùng VES_S15 ở các thời điểm tháng 4 đến tháng 9.



Hình 2. Sai số MAE theo từng thời điểm dự báo ở giai đoạn luyện (hình a) và dự báo độc lập (hình b) trên ba vùng VES, VES_S15 và VES_N15.



Hình 3. Hệ số tương quan theo từng thời điểm dự báo trên bộ số liệu phụ thuộc (hình a) và dự báo độc lập (hình b) trên ba vùng VES, VES_N15 và VES_S15.

Để có cái nhìn tổng quát hơn về chất lượng của 9 mô hình dự báo khảo sát tại 12 thời điểm dự báo trong năm, điểm kỹ năng MSSS tương ứng của từng sơ đồ dự báo này được tính toán và dẫn ra tại bảng 3 dưới đây.

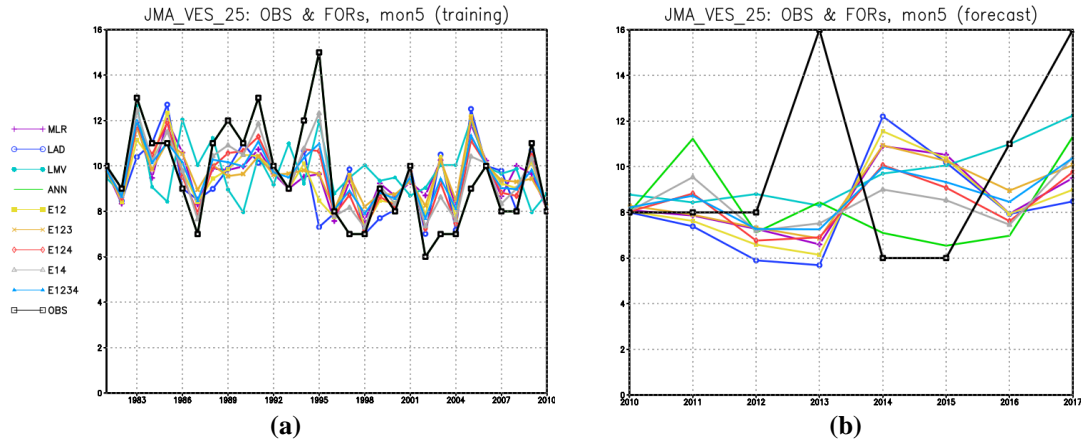
Kết quả tại bảng 3 cho thấy ở giai đoạn 1981-2010 các mô hình MLR, ANN, E12, E123, E124, E14 và E1234 đạt tỷ lệ 100% (12/12 sơ đồ dự báo đạt) có số lượng các sơ đồ dự báo “đạt” tiếp đến là mô hình LAD với tỷ lệ tương ứng 75% (9/12) và thấp nhất ở mức trên dưới 58% (7/12) khi sử dụng mô hình LMV. Ở giai đoạn này khu vực VES_S15 có kết quả dự báo tốt nhất, tiếp đến là vùng VES_N15 và VES. Ở giai đoạn dự báo độc lập (2011-2017) số lượng các sơ đồ dự báo “đạt” giảm mạnh ở cả 9 mô hình khảo sát trên 3 vùng nghiên cứu. Cụ thể ở vùng VES_N15 cho kết quả dự báo tốt nhất với khoảng trên dưới 33% sơ đồ đạt, tiếp đến là vùng VES với tỷ lệ tương ứng khoảng 25% và thấp nhất là vùng VES_S15 với chỉ khoảng 16% sơ đồ dự báo đạt. Cá biệt mô hình ANN thể hiện sự không ổn định khi thay đổi từ tỷ lệ 100% ở giai đoạn phụ thuộc xuống chỉ còn khoảng trên dưới 16% ở dự báo độc lập. Trong số các mô hình thống kê khảo sát, mô hình MLR có xu hướng cho kết quả dự báo tốt nhất, tiếp đến là mô hình E12, E123 và LAD.

Để minh họa thêm về chất lượng của các mô hình dự báo, nghiên cứu đã dẫn ra tại hình 4 đến hình 9 dưới đây các giá trị quan trắc và kết quả

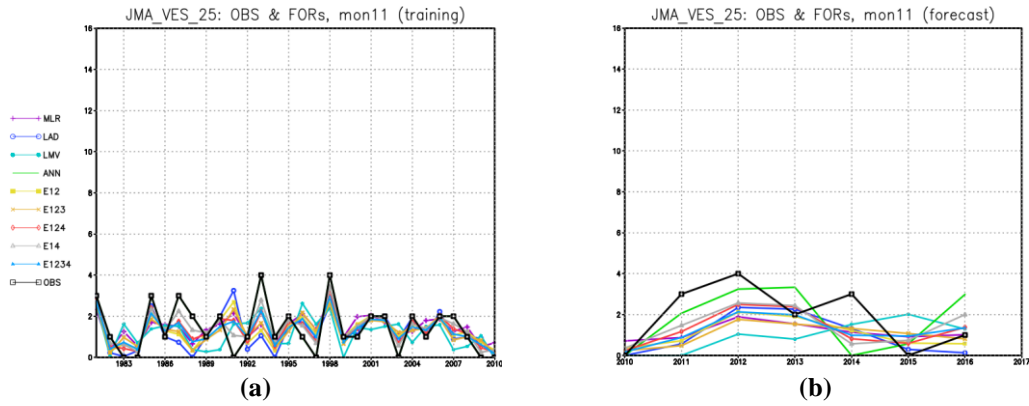
dự báo (trên bộ số liệu phụ thuộc và độc lập) về số lượng TCs_6mon cho mùa XTND chính (tháng 6 tới tháng 11) và mùa “ít” bão (tháng 12 tới tháng 5 năm liền tiếp) trên ba vùng VES, VES_S15 và VES_N15. Các kết quả này (không được thể hiện đầy đủ tại đây) cho thấy ở cả 3 vùng nghiên cứu kết quả dự báo trên bộ số liệu phụ thuộc đã mô phỏng khá tốt so với số liệu quan trắc ở cả 9 mô hình thống kê khảo sát, trong đó tốt nhất có thể nhận thấy ở mô hình ANN. Kết quả dự báo trên bộ số liệu độc lập đã có những diễn biến khác so với trên bộ số liệu phụ thuộc, thể hiện sự khác nhau ở các vùng không gian nghiên cứu và các thời điểm dự báo. Nhìn chung, tồn tại nhiều sơ đồ có kết quả dự báo độc lập khá sát với số liệu quan trắc, đặc biệt ở 2 năm đầu tiên tính từ cuối thời điểm chọn số liệu phụ thuộc (năm 2011 và 2012) và kết quả dự báo cho mùa ít XTND thể hiện tốt hơn. Các kết quả này cũng cho thấy, việc lựa chọn khoảng thời gian cho bộ số liệu làm dự báo độc lập quá xa so với thời điểm cuối cùng của bộ số liệu phụ thuộc có thể ảnh hưởng không tốt đến kết quả dự báo bởi khi đó những quan hệ thống kê ở bộ số liệu phụ thuộc không được mô hình “nắm bắt” được. Tuy vậy các kết quả này là cơ sở cho phép kỳ vọng có thể tiếp tục nghiên cứu ứng dụng các mô hình TKTT để lựa chọn ra các sơ đồ có chất lượng dự báo tốt với độ ổn định cao cho dự báo nghiệp vụ hạn mùa XTND tại VES nói chung và 2 khu vực VES_S15, VES_N15 nói riêng.

Bảng 3. Số lượng các sơ đồ dự báo “đạt” qua đánh giá chỉ số kỹ năng dự báo MSSS cho giai đoạn dự báo phụ thuộc và độc lập

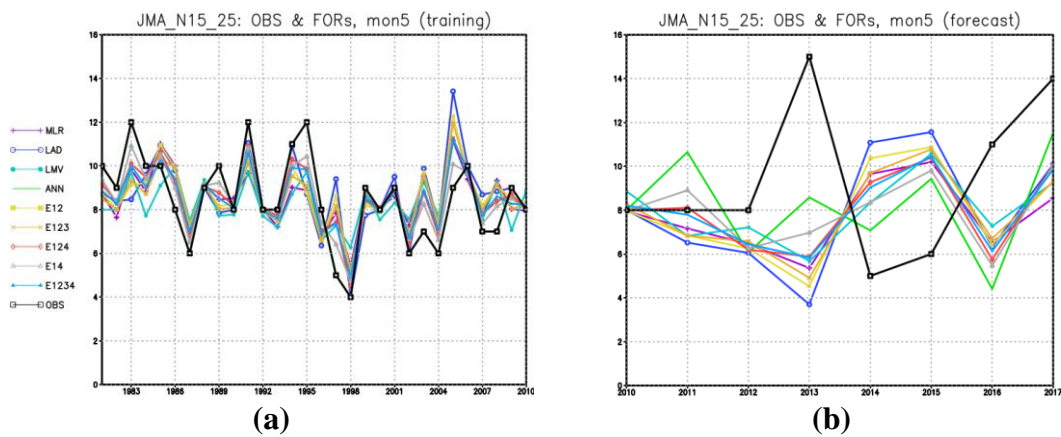
Dự báo	Vùng nghiên cứu	Mô hình thống kê								
		MLR	LAD	LMV	ANN	E12	E123	E124	E14	E1234
Phụ thuộc (1981-2010)	VES	12	9	5	12	12	12	12	12	12
	VES_N15	12	9	6	12	12	12	12	12	12
	VES_S15	12	9	10	12	12	12	12	12	12
Độc lập (2011-2017)	VES	5	3	2	1	4	3	3	3	3
	VES_N15	6	3	4	3	4	4	3	4	5
	VES_S15	2	3	2	0	3	3	0	0	1



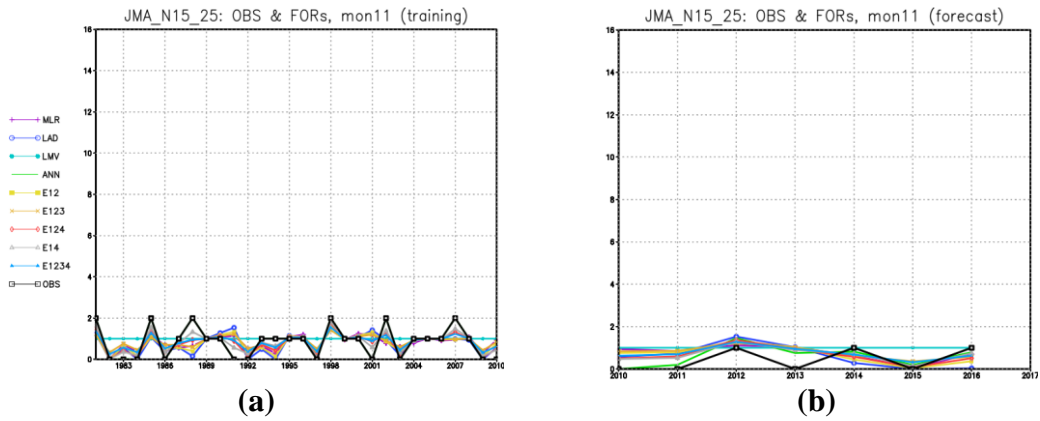
Hình 4. Số lượng XTNĐ mùa bão chính trên vùng VES (tháng 5 đến tháng 11) quan trắc và dự báo trên bộ số liệu phụ thuộc (hình a) và dự báo độc lập (hình b).



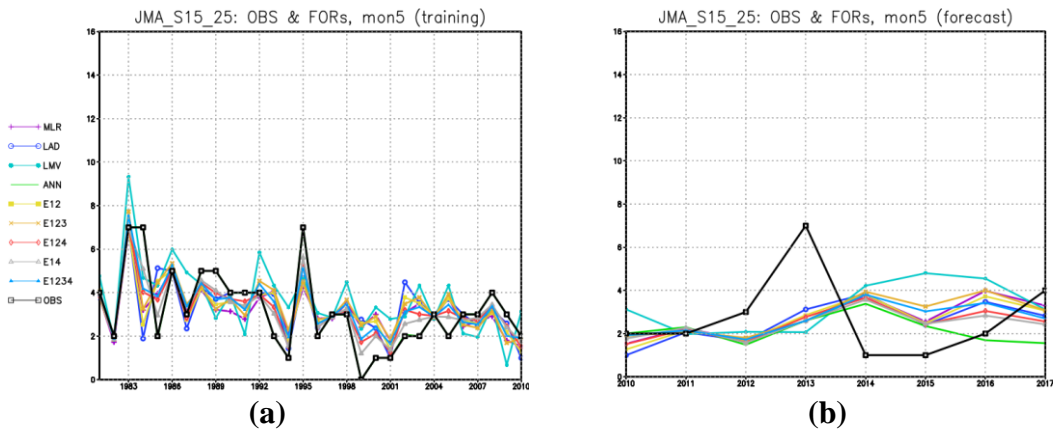
Hình 5. Số lượng XTNĐ mùa “ít” bão trên vùng VES (tháng 12 đến tháng 5 của năm liền sau) quan trắc và dự báo trên bộ số liệu phụ thuộc (hình a) và dự báo độc lập (hình b).



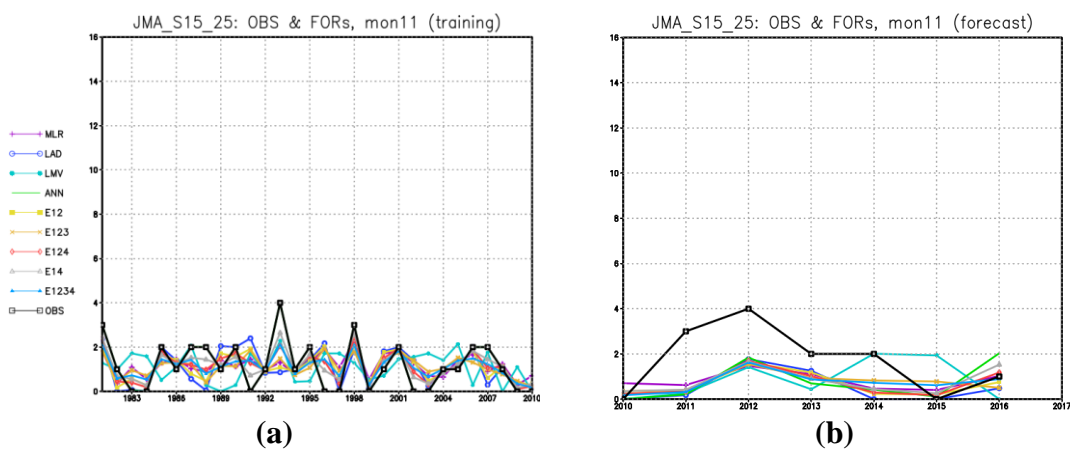
Hình 6. Số lượng XTNĐ mùa bão chính trên vùng VES_N15 (tháng 5 đến tháng 11) quan trắc và dự báo trên bộ số liệu phụ thuộc (hình a) và dự báo độc lập (hình b).



Hình 7. Số lượng XTNĐ mùa “ít” bão trên vùng VES_N15 (tháng 12 đến tháng 5 năm liền sau) quan trắc và dự báo trên bộ số liệu phụ thuộc (hình a) và dự báo độc lập (hình b).



Hình 8. Số lượng XTNĐ mùa bão chính trên vùng VES_S15 (tháng 5 đến tháng 11) quan trắc và dự báo trên bộ số liệu phụ thuộc (hình a) và dự báo độc lập (hình b).



Hình 9. Số lượng XTNĐ mùa “ít” bão trên vùng VES_S15 (tháng 12 đến tháng 5 năm liền sau) quan trắc và dự báo trên bộ số liệu phụ thuộc (hình a) và dự báo độc lập (hình b).

4. Kết luận

Sử dụng 4 mô hình TKTT và 5 mô hình tổ hợp được kết hợp từ các mô hình trên với nhau nghiên cứu này đã tiến hành xây dựng các phương trình dự báo số lượng XTND hạn 6 tháng cho các khu vực VES, VES_N15, VES_S15, những kết quả bước đầu cho phép rút ra một số kết luận sau:

-Trên bộ số liệu phụ thuộc mô hình MLR, ANN, E12, E123, E124, E14 và E1234 cho kết quả dự báo tốt nhất ở cả 12 thời điểm dự báo, thậm chí các chỉ số đánh giá sai số ở mô hình ANN đạt gần mức giá trị lý tưởng, tiếp đó là các tổ hợp với mô hình ANN, kết quả dự báo kém nhất khi sử dụng mô hình LMV.

- Thử nghiệm dự báo trên bộ số liệu độc lập cho thấy số lượng các sơ đồ dự báo “đạt” theo ngưỡng chỉ số MSSS giảm so với thời kỳ dự báo phụ thuộc và nhìn chung sai số của các mô hình có sự tương đồng nhau. Mô hình MLR có xu hướng cho kết quả dự báo tốt nhất, tiếp đến là mô hình E12, E123, LAD... và thấp nhất là mô hình ANN. Vùng VES_N15 có xu hướng cho kết quả dự báo tốt nhất, tiếp đến là vùng VES và thấp nhất là vùng VES_S15.

Lời cảm ơn

Nghiên cứu này được tài trợ bởi Đề tài nghiên cứu khoa học KC.09.15/16-20, thuộc chương trình KC.09, Bộ Khoa học và Công nghệ.

Tài liệu tham khảo

- [1] W. Landsea Christopher, Gerald D. Bell, William M. Gray, Stanley B. Goldenberg, The extremely active 1995 Atlantic hurricane season: Environmental conditions and verification of seasonal forecasts, *Mon. Wea. Rev.* 126 (1998) 1174-1193.
- [2] W. Landsea Christopher, William M. Gray, Paul W. Mielke, Jr, Kenneth J. Berry, Seasonal Forecasting of Atlantic hurricane activity, *Weather* 49 (1994) 273-284.
- [3] M. Gray William, Christopher W. Landsea, Paul W. Mielke, Predicting Atlantic basin seasonal tropical cyclone activity by 1 June, *Weather and Forecasting* 9 (1994) 103-115.
- [4] Neville Nicholls, Chris Landsea, Jon Gill, Recent trends in Australian region tropical cyclone activity, *Meteorol. Atmos. Phys.* 65 (1998) 197-205.
- [5] Elsner, James B., Kam-biu Liu, Bethany Kocher, Spatial Variations in Major U.S., Hurricane Activity: Statistics and a Physical Mechanism, *J. Climate.* 13 (2000) 2293-2305.
- [6] J. C. L. Chan, J. E. Shi and C. M. Lam, Seasonal forecasting of tropical cyclone activity over the Western North Pacific and the South China Sea, *Wea. Forecasting* 13 (1998) 997-1004.
- [7] J. C. L. Chan, Tropical cyclone activity over the Western North Pacific associated with El Niño and La Niña events, *J. Climate.* 13 (2000) 2960-2972.
- [8] Pao-Shin Chu, Xin Zhao, Chang-Hoi Ho, Hyeong-Seog Kim, Mong-Ming Lu, Joo-Hong Kim, Bayesian forecasting of seasonal typhoon activity: A track-pattern oriented categorization approach, *J. Climate.* 23 (2010) 6654-6668
- [9] M. Lu, P.-S. Chu, and Y.-C. Lin, Seasonal prediction of tropical cyclone activity near Taiwan using the Bayesian multivariate regression method, *Wea. Forecasting* 25 (2010) 1780-1795.
- [10] H. J. Kwon, W.-J. Lee, S.-H. Won, E.-J. Cha, Statistical ensemble prediction of the tropical cyclone activity over the Western North Pacific, *Geophys. Res. Lett.* 34 (2007) L24805 1-5. doi:10.1029/2007GL032308
- [11] J. C. L. Chan, Tropical cyclone activity in the Western North Pacific in relation to the stratospheric quasi-biennial oscillation, *Mon. Wea. Rev.* 123 (1995) 2567-2571.
- [12] J. C. L. Chan, Prediction of annual tropical cyclone activity over the Western North Pacific and the South China Sea, *Int'l J. Climatol.* 15 (1995) 1011-1019.
- [13] J. C. L. Chan, J. E. Shi and C. M. Lam, Seasonal forecasting of tropical cyclone activity over the Western North Pacific and the South China Sea, *Wea. Forecasting* 13 (1998) 997-1004.
- [14] J.C.L. Chan, J.E. Shi, K.S. Liu, Improvements in the seasonal forecasting of tropical cyclone activity over the Western North Pacific. *Wea. Forecasting* 16 (2001) 491-498.
- [15] P. J. Klotzbach, Recent developments in statistical prediction of seasonal Atlantic basin tropical cyclone activity, *Journal compilation C* (2007) Blackwell Munksgaard. <https://doi.org/10.1111/j.1600-0870.2007.00239.x>, *Tellus* 59A (4) (2007) 511-518.

- [16] W. Zhang, Y. Zhang, D. Zheng, F. Wang, and L. Xu, Relationship between lightning activity and tropical cyclone intensity over the northwest Pacific, *J. Geophys. Res. Atmos.* 120 (2015) 4072-4089. <https://doi.org/10.1002/2014JD022334>.
- [17] Phan Van Tan, On the tropical cyclone activity in the Northwest Pacific basin and South China sea in relationship with ENSO, *Journal of Science, Vietnam National University, Hanoi* (1) 18 (2002) 51-58. (In English)
- [18] Nguyễn Văn Tuyên, Xu hướng hoạt động của xoáy thuận nhiệt đới trên Tây Bắc Thái Bình Dương và biển Đông theo các cách phân loại khác nhau, *Tạp chí KTTV* 559 (2007) 4-10.
- [19] Đinh Bá Duy, Ngô Đức Thành, Phan Văn Tân, Mối quan hệ giữa ENSO và số lượng, cấp độ Xoáy thuận Nhiệt đới trên khu vực Tây Bắc - Thái Bình Dương, *Biển Đông giai đoạn 1951-2015*, *VNU Journal of Science: Earth and Environmental Sciences* 32 (3S) (2016) 43-55.
- [20] Đinh Bá Duy, Ngô Đức Thành, Nguyễn Thị Tuyết, Phạm Thanh Hà, Phan Văn Tân, Đặc điểm hoạt động của Xoáy thuận Nhiệt đới trên khu vực Tây Bắc Thái Bình Dương, *Biển Đông và vùng trực tiếp chịu ảnh hưởng trên lãnh thổ Việt Nam giai đoạn 1978-2015*, *VNU Journal of Science: Earth and Environmental Sciences* (2) 32 (2016) 1-11.
- [21] Đinh Văn Ưu, Đánh giá quy luật biến động dài hạn và xu thế biến đổi số lượng bão và áp thấp nhiệt đới trên khu vực Tây Thái Bình Dương, *Biển Đông và ven biển Việt Nam*, *Tạp chí Khoa học ĐHQGHN, Khoa học Tự nhiên và Công nghệ* (3S) 25 (2009) 542-550.
- [22] Nguyễn Văn Hiệp và nnk, Đặc điểm hoạt động của bão ở Tây Bắc Thái Bình Dương và Biển Đông qua số liệu Ibracs, *Tuyên tập báo cáo tại Hội thảo khoa học năm 2016 của Viện Khoa học KTTV & BDKH*, (2006) 9-14.
- [23] Vũ Thanh Hằng, Ngô Thị Thanh Hương,, Phan Văn Tân, Đặc điểm hoạt động của bão ở vùng biển gần bờ Việt Nam giai đoạn 1945-2007, *Tạp chí Khoa học ĐHQGHN, Khoa học Tự nhiên và Công nghệ* (3S) 26 (2010) 344-353.
- [24] Nguyễn Văn Tuyên, Khả năng dự báo hoạt động mùa bão biển Đông Việt Nam: Phân tích các yếu tố dự báo và nhân tố dự báo có thể (Phần I), *Tạp chí KTTV*, 568 (2008) 1-8.
- [25] Nguyễn Văn Tuyên, 2008: Khả năng dự báo hoạt động mùa bão biển Đông Việt Nam: Phân tích các yếu tố dự báo và nhân tố dự báo có thể (Phần II). *Tạp chí KTTV*, 571 (2008) 1-11.
- [26] Phan Văn Tân, 2009-2010, Nghiên cứu tác động của biến đổi khí hậu toàn cầu đến các yếu tố và hiện tượng khí hậu cực đoan ở Việt Nam, khả năng dự báo và giải pháp chiến lược ứng phó. Đề tài cấp Nhà nước, mã số KC08.29/06-10.
- [27] <https://www.jma.go.jp/jma/jma-eng/jma-center/rsmc-hp-pub-eg/besttrack.html>.
- [28] <https://www.esrl.noaa.gov/psd/data/climateindices/list/>
- [29] T. Ngo-Duc, J. Matsumoto, H. Kamimera, H.H. Bui, Monthly adjustment of Global Satellite Mapping of Precipitation (GSMaP) data over the VuGia–ThuBon River Basin in Central Vietnam using an artificial neural network. *Hydrological Research Letters* 7 (4) (2013) 85-90. <https://doi.org/10.3178/hrl.7.85>.
- [30] J. C. L. Chan, J. E. Shi and C. M. Lam, Seasonal forecasting of tropical cyclone activity over the Western North Pacific and the South China Sea, *Wea. Forecasting* 13 (1998) 997-1004.
- [31] E. S. Blake, W. M. Gray, Prediction of August Atlantic Basin Hurricane Activity. *Wea. Forecasting* 19 (2004) 1044-1060.
- [32] P. J. Klotzbachi, W. M. Gray, Extended range forecast of Atlantic seasonal Hurricane activity and U. S. landfall strike probability for 2008, (2007) <http://hurricane.atmos.colostate.edu/Forecasts>.