

THỬ NGHIỆM DỰ BÁO TỔ HỢP CHO QUI ĐẠO BÃO BẰNG PHƯƠNG PHÁP THỐNG KÊ TỪ DỰ BÁO CỦA CÁC TRUNG TÂM QUỐC TẾ

ThS. Nguyễn Chi Mai, CN. Nguyễn Thu Hằng,
Nguyễn Phương Liên
Trung tâm Dự báo khí tượng thủy văn Trung ương

Như các phân tích tổng quan trong [1], phương pháp tổ hợp thống kê từ dự báo của các mô hình hay là từ các trung tâm khác nhau tỏ ra thích hợp và có triển vọng áp dụng khả quan với điều kiện thông tin và máy tính hiện nay của nước ta. Trong phần này, số liệu của mùa bão 2003 được thử nghiệm đánh giá sử dụng kết quả dự báo của các trung tâm phát báo quốc tế trong các mùa bão 2001 và 2002. Ba phương pháp thống kê được sử dụng để thử nghiệm là: phương pháp lấy trung bình đơn giản, phương pháp hồi qui tuyến tính đa biến và cuối cùng là phương pháp tổ hợp tuyến tính với trọng số dựa trên sai số dự báo của từng dự báo thành phần sử dụng. Các kết quả bước đầu cho thấy việc sử dụng tổ hợp có hiệu quả tốt, cho sai số trung bình của dự báo tổ hợp nhỏ hơn sai số trung bình của trung tâm dự báo tốt nhất. (Đề tài được hỗ trợ bởi chương trình NCKHCB, mã số 730902).

Mở đầu

Dự báo tổ hợp (DBTH) thống kê từ kết quả của nhiều mô hình khác nhau (còn được gọi là “siêu tổ hợp” (superensemble)) là phương pháp tương đối đơn giản và thích hợp với điều kiện thông tin và máy tính của Việt Nam hiện nay. Riêng đối với dự báo đường đi của bão, phương pháp này sử dụng ở nhiều các trung tâm dự báo nghiệp vụ trên thế giới cùng với những phương pháp dự báo tổ hợp phức tạp hơn như dùng các điều kiện ban đầu khác nhau hay chạy các phiên bản khác nhau của một mô hình.

Bản chất của phương pháp này là đưa ra được tổ hợp dự báo có độ tin cậy cao hơn dựa trên dự báo từ các nguồn khác nhau làm đầu vào. Như vậy, đây cũng có thể coi là một trong những phương pháp thống kê kết quả của mô hình (Model Output Statistics - MOS) tuy có điểm khác là ở đây sử dụng kết quả từ nhiều mô hình khác nhau thay vì sử dụng đặc tính thống kê của riêng một mô hình nào đó như trong MOS.

Trong bài này, mục 1 sẽ trình bày các phương pháp tổ hợp thống kê được sử dụng, kết quả thử nghiệm với mùa bão 2003 sẽ được nói đến ở phần 2.

1. Phương pháp tổ hợp thống kê

Đối với bài toán dự báo đường đi của bão, có ba phương pháp thống kê đang được sử dụng phổ biến hiện nay là:

1) Lấy trung bình đơn giản (TBĐG) của các dự báo. Phương pháp này do Carr và Elsebery [3] đề xuất và đang được sử dụng ở Trung tâm Dự báo Bão của Mỹ ở Guam (JTWC).

2) Thực hiện hồi qui tuyến tính (HQT) đa biến theo đề xuất của Krishnamurti và nnk [4].

3) Tạo tổ hợp tuyến tính của các dự báo với trọng số phụ thuộc vào sai số dự báo của mô hình.

Về bản chất, cả ba phương pháp trên đều có điểm chung là tạo tổ hợp tuyến tính. Khác nhau cơ bản giữa chúng chính là cách tính trọng số cho tổ hợp tuyến tính đó. Như vậy, công thức tổng quát của tổ hợp thống kê bằng cách tạo tổ hợp tuyến tính có thể viết dưới dạng:

$$F_{th} = \sum_{i=1}^N w_i F_i, \quad (1)$$

trong đó F_{th} là kết quả dự báo tổ hợp, F_i là dự báo thành phần của từng mô hình hay của các trung tâm dự báo, w_i là trọng số tương ứng với từng dự báo thành phần và N là số thành phần tham gia vào tổ hợp. Các phương pháp tính trọng số sẽ được trình bày dưới đây. Riêng phương pháp 2 sẽ được thực hiện hơi khác so với cách của Weber [5] do cơ sở dữ liệu không đủ lớn để phân loại trước khi xây dựng phương trình tính trọng số.

a. Phương pháp lấy trung bình đơn giản

Ở phương pháp này, trọng số của mọi dự báo thành phần đều như nhau và có dạng đơn giản như sau:

$$w = w_i = 1/N \quad (2)$$

Như vậy, mọi dự báo thành phần đều được coi là quan trọng như nhau. Phương pháp này đơn giản, không cần phải có số liệu lịch sử, không cần quan tâm đến đặc điểm hay tính chất của các nguồn số liệu. Tuy nhiên, cần phải lưu ý rằng chất lượng của dự báo tổ hợp theo cách này sẽ giảm sút đáng kể trong trường hợp có một vài dự báo thành phần không tốt, tách ra hẳn so với chòm các dự báo thành phần khác. Một biện pháp cải tiến cho những trường hợp này là lựa chọn dự báo trước khi đưa vào tập hợp (selected consensus). Đây chính là cách tiếp cận có hệ thống để giải quyết vấn đề dự báo bão mà Carr và Elsebery [3] đã đề xuất. Cần lưu ý, để thực hiện được điều này, người làm dự báo cũng phải rất có kinh nghiệm và nắm chắc các kiến thức synop ảnh hưởng đến đường đi của bão cũng như các đặc điểm dự báo của từng nguồn số liệu. Tuy nhiên, không phải lúc nào việc lựa chọn dự báo thành phần cũng cải thiện được chất lượng dự báo tổ hợp, mà ngược lại rất có thể lại lược bỏ những nguồn thông tin tốt.

b. Tính trọng số bằng phương pháp hồi qui tuyến tính đa biến

Công thức tổng quát của phương pháp HQT đa biến về cơ bản cũng có thể đưa về công thức (1), tuy cũng có điểm hơi khác là có thể thêm thành phần tự do và có thể viết dưới dạng như sau:

$$F_{th} = C + \sum_{i=1}^N w_i F_i \quad (3)$$

với C là số hạng tự do, w_i là trọng số của từng nhân tố và chính là các hệ số hồi qui được tính và F_i là các dự báo thành phần, N là số lượng các dự báo thành phần tham gia tổ hợp.

c. Phương pháp tính trọng số theo sai số

Ở phương pháp này, trọng số có thể được biểu diễn như sau:

$$w_i = \frac{1/e_i}{\sum_{i=1}^N 1/e_i} \quad (4)$$

Trong đó e_i là sai số của nguồn dự báo thành phần, các ký hiệu khác giống như đã trình bày ở trên. Cách biểu diễn này đảm bảo rằng tỷ trọng của từng dự báo thành phần tỷ lệ nghịch với sai số tương ứng và tổng các tỷ trọng phải bằng 1.

Điều có thể thay đổi hay cải tiến ở phương pháp này chính là cách xác định sai số của từng dự báo thành phần e_i . Cách đơn giản nhất có thể dùng là sử dụng sai số trung bình của nguồn dự báo đó trong một (hay vài) mùa bão trước. Phương pháp này đồng nghĩa với việc cho rằng kỹ năng dự báo của từng nguồn dự báo, về trung bình, không thay đổi từ năm trước qua năm sau. Chúng tôi sẽ thử nghiệm cách tính này ở phần sau (ký hiệu là phương pháp 3a).

Phương pháp của Weber [5] xác định sai số dự báo của từng nguồn tương ứng với các tính chất và đặc điểm của bão (ví dụ như cường độ, kích thước, tốc độ di chuyển vv...). Ở đây, giả thiết ngầm định là các dự báo có kỹ năng tương tự đối với các cơn bão có các tính chất tương đương. Tuy phương pháp này được thử nghiệm cho kết quả khả quan đối với các cơn bão ở Đại Tây Dương, nhưng khả năng áp dụng hiệu quả ở khu vực Tây Bắc Thái Bình Dương còn là vấn đề cần nghiên cứu. Mối quan tâm ở đây nảy sinh do các cơn bão ở Đại Tây Dương thường có nguồn gốc là các sóng Châu Phi và do đó thường có các đặc tính tương tự nhau về kích thước, tốc độ chuyển động, hay cường độ... Điều này dẫn đến sự khác nhau của các tính chất này trong các cơn bão rất có thể là do nguyên nhân khác (ví dụ như hiện trạng của các quá trình qui mô lớn hơn). Như vậy, sự tương quan giữa sai số dự báo và các đặc tính hình học của bão có thể chỉ là thể hiện bề ngoài của mối tương quan giữa sai số dự báo với những điều kiện khác. Tuy nhiên, đối với khu vực Thái Bình Dương, bão có thể phát sinh bởi nhiều cơ chế khác nhau, do đó, phương pháp như của Weber rất có thể sẽ cho các kết quả khác. Chúng tôi sẽ xem xét khả năng sử dụng của phương pháp này ở phần sau (ký hiệu là phương pháp 3b).

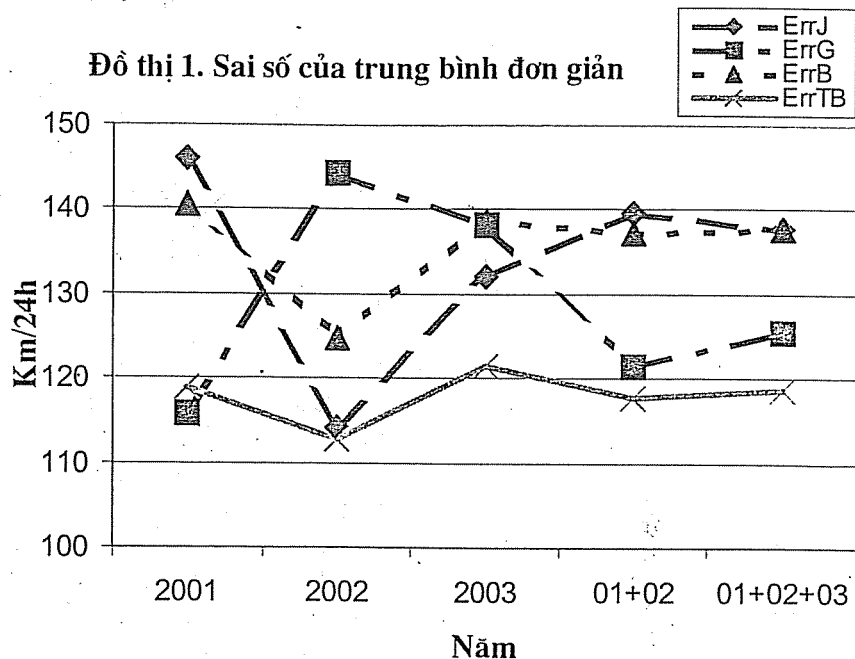
Giả thiết trên của Weber có thể bị phê phán là thiếu sự liên quan giữa kỹ năng dự báo với các hình thể synop hay là thực trạng qui mô lớn của khí quyển. Điều này phát sinh từ thực tế là mỗi mô hình dự báo thường có sai số hệ thống với một số hình thể synop nhất định. Do đó, sẽ là lý tưởng nếu như sai số của từng nguồn dự báo được xác định tương ứng với hình thể synop. Tuy nhiên, việc thiết lập được quan hệ này đòi hỏi phải có nhiều số liệu hơn, các hệ thống synop đặc trưng xung quanh bão cũng cần được xác định để phân loại hình thể synop chứ không phải bão. Do vậy, phương pháp này chưa được thực hiện ở đây.

2. Các kết quả thử nghiệm với mùa bão 2003

Để thử nghiệm cho các cơn bão ảnh hưởng đến Việt Nam trong mùa bão 2003, chúng tôi sử dụng bộ số liệu quỹ đạo đã chỉnh lý của Việt Nam và các dự báo của các trung tâm bão trong khu vực bao gồm Nhật Bản, Bắc Kinh và JTWC của 3 mùa bão từ 2001 đến 2003. Trong đó, số liệu của hai mùa bão 2001 và 2002 được sử dụng để tính hệ số hồi qui, năm 2003 là số liệu độc lập để thử nghiệm. Số lượng các bản ghi sử dụng cho các tính toán này đòi hỏi phải có đủ số liệu của cả 3 trung tâm và số liệu của 24 giờ sau để kiểm chứng, do đó, chỉ bao gồm 108 bản ghi cho năm 2001, 28 bản ghi cho năm 2002 và 44 bản ghi cho năm 2003.

a. Kết quả thử nghiệm với phương pháp lấy trung bình đơn giản

Đồ thị 1 biểu diễn sai số của các trung tâm Nhật (ErrJ), JTWC (ErrG)*, Bắc Kinh (ErrB) và tổ hợp theo TBĐG (ErrTB - đường liền nét) trong các năm 2001, 2002, 2003, tổng hợp của hai năm (2001, 2002) và cả ba năm (2001, 2002, 2003). Các kết quả này thể hiện rõ nhận định rằng sai số của TBĐG luôn nhỏ hơn sai số của dự báo thành phần. Riêng năm 2001 là một ngoại lệ với sai số của TBĐG hơi lớn hơn sai số của JTWC. Tuy nhiên, sự khác biệt này là rất nhỏ (~3km) và không đáng kể so với độ biến thiên của sai số dự báo 24h.



Kết quả trên đồ thị 1 cũng thể hiện một ưu thế của TBĐG là nó cho sai số nhỏ nhất ngay cả khi kỹ năng của các nguồn dự báo thay đổi rõ rệt theo từng năm.

b. Kết quả thử nghiệm với phương pháp hồi quy tuyến tính

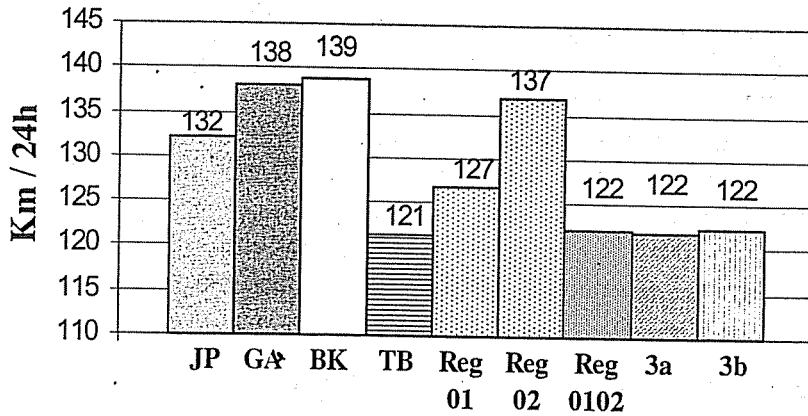
Trên đồ thị 2, các cột từ trái sang phải lần lượt biểu thị sai số dự báo 24h cho mùa bão năm 2003 của các trung tâm Nhật (JP), Guam-JTWC (GA), Bắc Kinh (BK), tổ hợp TBĐG (TB), các tổ hợp HQTТ với các hệ số được tính từ số liệu của mùa bão 2001 (Reg01), 2002 (Reg02) và cả hai năm (Reg0102), theo phương pháp 3a và 3b.

Đối với phương pháp HQTТ, có thể thấy rằng sai số của tổ hợp biến đổi tương đối mạnh khi sử dụng các số liệu nền khác nhau để tính hồi qui. Năm 2002 cho kết quả tối nhất trong khi tổng hợp của cả hai mùa bão 2001 và 2002 lại đưa ra sai số nhỏ

* Lưu ý rằng, theo tổng kết của Trung tâm Dự báo JTWC tại Guam (<https://metoc.npmoc.navy.mil>), sai số dự báo 24h của trung tâm này trong năm 2001 cho bão lớn (typhoon, với $V_{max} \geq 65kt$) và bão yếu hơn (tropical cyclone với $V_{max} \geq 35kt$) lần lượt là 67km và 73km. Sai số tương ứng của năm 2002 là 47km và 66km. Như vậy là nhỏ hơn nhiều so với số liệu tính được ở đây (sai số trung bình cho bão ở mọi cường độ khác nhau là 112 km cho năm 2001 và 144km cho năm 2002). Sự khác biệt này có thể được giải thích bởi sự khác nhau giữa các bộ số liệu đưa vào tính toán: bộ số liệu được dùng ở đây chỉ bao gồm những cơn bão có ảnh hưởng đến Việt Nam, như vậy, tạo thành một tập hợp con so với các tính toán thống kê của JTWC. Với sự giải thích như vậy, có thể suy ra rằng kỹ năng dự báo của JTWC cho khu vực của Việt Nam tương đối kém hơn so với kỹ năng chung của nguồn số liệu này.

hơn cả. Điều này cũng biểu hiện một qui luật là bộ số liệu để tính hệ số hồi qui càng lớn thì kết quả của tổ hợp HQTT càng tốt. Tuy nhiên, cũng phải nhận thấy rằng kết quả tốt nhất này cũng chỉ xấp xỉ và thậm chí có sai số còn hơi lớn hơn kết quả của TBDG nêu trên, mặc dù phương pháp TBDG đơn giản hơn nhiều.

Đồ thị 2. Sai số dự báo cho năm 2003

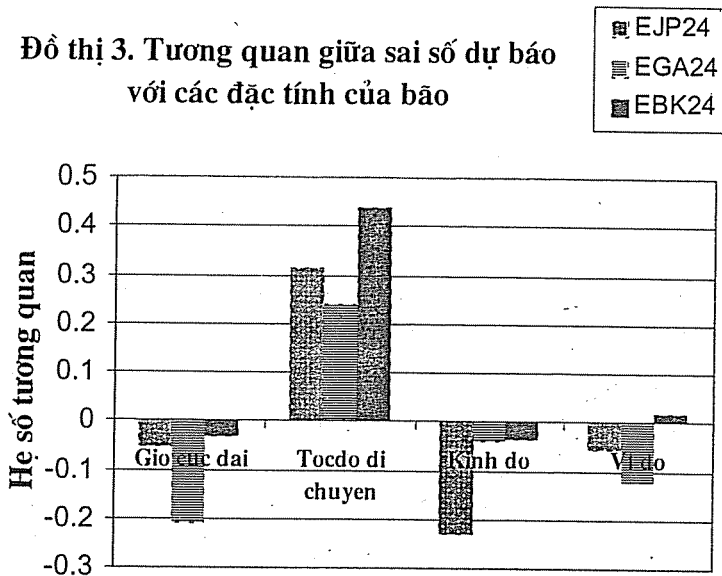


Các trung tâm và các hướng dự báo

c. Kết quả thử nghiệm với phương pháp trọng số phụ thuộc sai số

Với việc sử dụng sai số trung bình trong mùa bão trước để tính trọng số cho tổ hợp, phương pháp 3a cho kết quả tương đương (và hơi lớn hơn một chút) so với các phương pháp TBDG và HQTT nêu trên.

Đồ thị 3. Tương quan giữa sai số dự báo với các đặc tính của bão



Các đặc tính của bão

Để áp dụng phương pháp 3b, tức là xác định trọng số của các nguồn dự báo dựa trên sai số của chúng tương ứng với các đặc điểm của bão, các phân tích tương quan đã được thực hiện trên bộ số liệu phụ thuộc (2001-2002). Như đồ thị 3 biểu diễn, đối với cả 3 nguồn dự báo, tương quan mạnh nhất là giữa tốc độ di chuyển và sai số. Với giá trị dương, các hệ số tương quan này cho thấy các nguồn dự báo cho sai số càng lớn

khi bão di chuyển càng nhanh, trong đó trung tâm Bắc Kinh có vẻ nhạy cảm với yếu

tố này nhất. Đối với cường độ của bão thể hiện thông qua tốc độ gió cực đại V_{max} , tất cả các hệ số đều mang dấu âm, thể hiện rằng sai số càng nhỏ khi bão có cường độ càng mạnh. Trong đó, dự báo của Guam nhạy cảm hơn hẳn đối với yếu tố này. Điều này cũng phù hợp với báo cáo của JTWC [5] cho thấy sai số của trung tâm này đối với bão mạnh (typhoon, $V_{max} \geq 65kt$) nói chung lớn hơn hẳn so với bão yếu hơn (TC, $V_{max} < 65kt$). Ngoài ra, tham số về vị trí ban đầu của bão (kinh độ, vĩ độ) cũng được xem xét tương quan với sai số. Nhìn chung thì yếu tố này ít gây ảnh hưởng nhiều đến kỹ năng dự báo của các trung tâm, ngoại trừ kinh độ ban đầu đối với các dự báo của Nhật. Có thể rút ra rằng sai số dự báo của các trung tâm (đặc biệt là Nhật) càng lớn đối với kinh độ càng nhỏ (ở phía tây hơn). Tuy nhiên, cần phải lưu ý rằng với giá trị tuyệt đối của các hệ số tương quan nhỏ như ở đây, các qui luật được rút ra ở trên có thể thay đổi trong từng trường hợp cụ thể.

Với các phân tích tương quan như trên, chúng tôi thực hiện phân lớp số liệu theo hai yếu tố chính là tốc độ di chuyển (lớn hơn và nhỏ hơn 15km/h) và cường độ bão (lớn hơn và nhỏ hơn 50kt) và tính sai số trung bình (e_i) cho 4 trường hợp phân lớp trên. Sử dụng công thức (4) cho từng trường hợp phân lớp đối với số liệu năm 2003 cho kết quả trung bình của phương pháp tổ hợp này là 122km.

3. Kết luận

Với số liệu của 3 mùa bão từ 2001 đến 2003 để thử nghiệm cho năm 2002, cần phải lưu ý rằng kết quả thử nghiệm của một năm không đủ để thể hiện hết sự khác nhau giữa các mùa bão cũng như chất lượng dự báo từ năm này qua năm khác và hơn nữa đây là kết luận đối với giá trị trung bình chứ không phải cho từng trường hợp cụ thể. Tuy nhiên, các phương pháp tổ hợp thống kê được thử nghiệm cho năm 2003 cho thấy hiệu quả rõ rệt của phương pháp lấy trung bình đơn giản và có thể sử dụng tham khảo cho công tác dự báo bão nghiệp vụ. Mặt khác, việc tiếp tục thử nghiệm và nghiên cứu ảnh hưởng của từng điều kiện synop cụ thể đến sai số của từng dự báo thành phần cũng là một hướng cần được tiếp tục xem xét để sử dụng các thông tin này một cách hiệu quả hơn.

Tài liệu tham khảo

1. Nguyễn Chi Mai, Nguyễn Thu Hằng: "Phương pháp dự báo tổ hợp và khả năng ứng dụng". - *Tạp chí Khí tượng Thủy văn*, tháng 2-2004.
2. Báo cáo Bão hàng năm của JTWC (năm 2001 và 2002),
<https://metoc.npmoc.navy.mil/jtwc/atcr/2001atcr/>,
<https://metoc.npmoc.navy.mil/jtwc/atcr/2002atcr/>
3. Carr, L., E., III, and R. L. Elsebery. *Systematic and integrated approach to tropical cyclone track forecasting. Part I. Approach overview and description of meteorological basis*. Tech. Rep. NPS-MR-94-002, Naval Postgraduate school, Monterey, CA 93943-5114, 273 pp, 1994.
4. Krishnamurti, T. N., C. M. Kishtawal, Z. Zhang, T. LaRow, D. Bachiochi, E. Williford, S. Gadgil, and S. Surendran. Multimodel ensemble forecasts for weather and seasonal climate. - *J. Climate*, 13, 4196-4216, 2000.
5. Weber, H.. Hurricane Track Prediction Using a Statistical Ensemble of Numerical Models. - *Mon. Wea. Rev.*, submitted, 2001.