

CHỈ TIÊU ĐÁNH GIÁ VÀ LỰA CHỌN MÔ HÌNH DỰ BÁO KHÍ HẬU

PGS.TS. Trần Việt Liên
Trung tâm Nghiên cứu Khí hậu

Các mô hình dự báo thời tiết và khí hậu đang được phát triển khá mạnh mẽ trên thế giới. Đã có hàng trăm mô hình ở những quy mô khác nhau đang được sử dụng vào nghiệp vụ và cũng có hàng trăm mô hình khác đang được nghiên cứu, thử nghiệm. Vậy làm thế nào để có thể đánh giá và lựa chọn được một mô hình thích hợp cho đối tượng cần dự báo? Để giải đáp vấn đề này cũng đã có nhiều công trình đề cập đến. Trong "Thông báo kỹ thuật" số 8 của Tổ chức Khí tượng thế giới (WMO) do Henry R. Stanski, Laurence J. Wilson và William R. Burrows soạn thảo năm 1989 [10] đã nêu ra hàng loạt các chỉ tiêu để đánh giá, thẩm định mô hình dự báo thời tiết, khí hậu. Theo các tác giả này thì *độ chính xác (accuracy) của một mô hình dự báo được định nghĩa là mức độ phù hợp của dự báo thời tiết với thời tiết thực đã xảy ra, được quan trắc khí tượng ghi lại*. Mức độ sai khác giữa trị số quan trắc và trị số dự báo được gọi là sai số (error). Cùng với khái niệm trên, tài liệu này còn nêu ra một số khái niệm khác như *độ tin cậy (reliability)*; *độ tinh xảo (skill)*.. Có hai dạng quan trọng của các chỉ tiêu dùng cho mục tiêu đánh giá dự báo được tài liệu [10] nêu ra:

1. Chỉ tiêu đánh giá dựa trên sai số

Đối với các kết quả dự báo số, nhất là các biến liên tục, chỉ số đánh giá được tính cụ thể qua nhiều chỉ số như

+ Sai số trung bình ME

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

hoặc sai số quân phương là căn bậc hai của sai số bình phương trung bình - RMSE

$$RMES = \left[\frac{1}{N} \sum_{i=1}^N (F_i - O_i)^2 \right]^{1/2} \quad (2)$$

+ Sai số tuyệt đối trung bình (MAE)

$$MAE = \frac{1}{N} \left[\sum_{i=1}^N |F_i - O_i| \right] \quad (3)$$

+ Phương sai rút gọn (RV)

$$RV = 1 - \frac{\sum (F_i - O_i)^2}{\sum (M - O_i)^2} = \frac{MSE_c - MSE}{\overline{DME_c}} \quad (4)$$

ở đây F_i và O_i là giá trị dự báo và quan trắc ở lần dự báo thứ i ,

M là trung bình mẫu hoặc trung bình khí hậu.

Đối với các dự báo xác suất, Brier (1950) đề nghị chỉ số đánh giá PS (Brier Score) có dạng sau:

$$PS = \frac{1}{N} \sum_{i=1}^N (f_i - O_i)^2 \quad (5)$$

ở đây f_i là xác suất dự báo xuất hiện của hiện tượng ở lần dự báo thứ i ; O_i nhận 2 giá trị: 1 nếu nó xuất hiện và 0 nếu không xuất hiện ở dự báo thứ i .

2. Chỉ tiêu đánh giá dựa vào “bảng liên kết” (CONTINGENCY TABLE)

Trong bài báo của Harrison và Graham [7] đăng trên “Thông báo” số 3 (VII-2001) của WMO đã đưa ra thảo luận vấn đề đánh giá chất lượng, giá trị và ứng dụng của dự báo mùa. Để đánh giá chất lượng và giá trị của dự báo khí hậu đối với một hiện tượng nào đó (như hạn, mưa, lốc tố...) mà kết quả được yêu cầu trả lời chỉ là *có* hoặc *không*, các tác giả cũng sử dụng “bảng liên kết” giữa kết quả dự báo với thực tế đã xuất hiện để xây dựng các chỉ tiêu đánh giá (bảng 1):

Bảng 1. Bảng liên kết 2 pha

		Dự báo	
		Có	Không
Quan trắc	Có	N11-Thành công (Hit)	N12-Thất bại (Miss)
	Không	N21-Báo động sai (False alarm)	N22-Đúng, bị bỏ qua Correct rejection

Bốn trường hợp có thể xảy ra, biểu thị các tình huống, hiện tượng có được dự báo hay không và thực tế nó có xuất hiện hay không. Khi hiện tượng xuất hiện và được dự báo ta có “thành công”, hiện tượng xuất hiện nhưng không được dự báo có “thất bại”. Nếu hiện tượng không xuất hiện nhưng lại dự báo có, tức là đã “báo động sai” còn nếu hiện tượng không xuất hiện và cũng được dự báo không xuất hiện thì dự báo đúng nhưng kết quả bị bỏ qua. Đánh giá ý nghĩa kinh tế của dự báo khí hậu theo mô hình chi phí-tổn thất (cost-loss model) của Murphy-1994 được các tác giả mô tả ở bảng 2.

Bảng 2. Bảng liên kết 2 pha

		Dự báo	
		Có	Không
Quan trắc	Có	Tổn thất xuất hiện nhưng được giảm nhẹ do có phòng chống	Tổn thất xuất hiện nhưng không được giảm bớt do phòng chống
	Không	Mất chi phí phòng chống nhưng tổn thất không xuất hiện	Tổn thất không xuất hiện và cũng không mất chi phí phòng chống

Sau nhiều lần dự báo, ta có thể tổng kết để đánh giá mức độ chính xác cũng như hiệu quả kinh tế của mô hình. Để tiến hành đánh giá mô hình sau N lần dự báo, các tác giả cũng nêu ra các chỉ tiêu đánh giá sau:

- Tỷ lệ thành công (Hit Rate - HR)

$$HR = \frac{N11}{(N11+N12)} \quad (6)$$

- Tỷ lệ báo động sai (False Alarm Rate - FR) là tỷ số của số lần dự báo nhưng hiện tượng không xuất hiện với tổng số lần hiện tượng không xuất hiện:

$$FR = \frac{N21}{(N21+N22)} \quad (7)$$

Ở định nghĩa này ta thấy FR tương tự như chỉ số SDT (Signal detection theory) do Mason đề ra năm 1980 [10] nhưng không giống với FAR do Henry R. Stanski, Laurence J. Wilson và William R. Burrows nêu ra ở [10]. Hai chỉ tiêu HR và FR đặc trưng cho mô hình dự báo các hiện tượng. Bảo đảm cân bằng giữa chúng là đòi hỏi của bất kỳ mô hình dự báo nào và khi tỷ lệ "thành công" vượt tỷ lệ "báo động sai", mô hình được coi là "tinh xảo" (skill) [7,10].

Với quan niệm trên nhưng khi đối tượng dự báo A không phải là một hiện tượng với 2 khả năng xuất hiện như đã nêu mà là một đại lượng xuất hiện dưới dạng một chuỗi giá trị, được chia thành nhiều khoảng (pha) để dự báo, sẽ xử lý như thế nào?

Giả sử đối tượng dự báo (ĐTDB) được chia thành 2 pha có vai trò và ý nghĩa như nhau, có khả năng xuất hiện ngang nhau thì rõ ràng, kết quả "báo động sai" đối với pha này sẽ là "thất bại" đối với pha kia và "tỷ lệ báo động sai" của mô hình dự báo đối với pha này sẽ có ý nghĩa là "tỷ lệ thất bại" đối với pha còn lại. Khi đó, ngoài ý nghĩa cụ thể đối với mỗi pha dựa theo định nghĩa trên, ta sẽ phải có thêm định nghĩa khác đại diện cho cả mô hình. Đó là tỷ lệ "thành công" và "báo động sai" trung bình của các pha. Theo các ký hiệu đã nêu ta có các định nghĩa sau:

- Tỷ lệ thành công trung bình:

$$HR_{TB} = [(N11/(N/2) + N22)/(N/2)]/2 = (N11 + N22)/N \quad (8)$$

- Tỷ lệ báo động sai trung bình:

$$FR_{TB} = [(N12/N/2) + (N21/N/2)]/2 = (N12 + N21)/N \quad (9)$$

Với định nghĩa này, rõ ràng mô hình chỉ có ý nghĩa dự báo khi $HR > FR$ tức là $HR > 50\%$. Thực ra, đối với mô hình 2 pha, đây cũng mới chỉ là "xác suất khí hậu". Hiện nay, để được sử dụng trong nghiệp vụ dự báo, HR_{TB} thường phải đạt trên 65-70%. HR_{TB} ở đây cũng có thể coi là độ chính xác của mô hình dự báo theo quan niệm thông thường như dự báo dấu chuẩn sai.

Đối với dự báo 3 pha là loại khá phổ biến hiện nay trong dự báo khí hậu ở nhiều nước. Tuy nhiên, cách xác định các pha dựa vào giá trị của đối tượng cũng có những khác nhau. Guiling Wang and Elfatih A.B Eltahir -1999 xác định 3 pha dựa vào trung bình và phương sai của mẫu [9].

- Pha1: trên trung bình (vượt chuẩn) $A > MA + a$,
- Pha 2: xấp xỉ trung bình (gần chuẩn) $MA - a < A < MA + a$,
- Pha 3: dưới trung bình (hụt chuẩn) $A < MA - a$.

Việc xác định trị số a có thể xuất phát từ lý thuyết hoặc từ khảo sát các chuỗi số liệu thực. Cách chọn $a = \sigma/2$ của công trình [9] cũng là một trong những cách chọn được một số tác giả sử dụng hiện nay. Tuy nhiên, cách chọn này rất khó bảo đảm tính

đồng xác suất giữa các pha, nhất là khi chúng có phân bố không chuẩn. Để bảo đảm tính ngang nhau về xác suất xuất hiện, cách xác định các khoảng pha dựa vào các phân vị là cách xác định hợp lý nhất. Đối với trường hợp 3 pha, đó là các tam phân vị (tercile), là các điểm chia ứng với xác suất 33% và 67%.

Để xác lập chỉ tiêu đánh giá độ chính xác của mô hình dự báo, Guiling Wang và Eltahir [9] đề xuất một chỉ tiêu đánh giá được gọi là "chỉ số dự báo" (forecasting Index) ký hiệu là FI. Việc xác định FI như sau: gọi FA' và FA là hàm biểu thị xác suất dự báo và xác suất xuất hiện thực của biến cố A trong mô hình 3 pha, ta sẽ có:

$$\begin{aligned} PA_i &= \begin{cases} 1,0,0 & \text{--- nếu } A \in \text{Pha I} \\ 0,1,0 & \text{.... nếu } A \in \text{Pha II} \\ 0,0,1 & \text{..... nếu } A \in \text{Pha III} \end{cases} \\ PA'_i &= \begin{cases} 1,0,0 & \text{--- nếu } A' \in \text{Pha I} \\ 0,1,0 & \text{.... nếu } A' \in \text{Pha II} \\ 0,0,1 & \text{..... nếu } A' \in \text{Pha III} \end{cases} \end{aligned}$$

Ở lần dự báo thứ i xác suất dự báo FP_i đối với các pha sẽ là

$$FP_i = \sum_{j=1}^3 PA_{i,j} PA'_{i,j}$$

Trong đó $PA_{i,j}$, $PA'_{i,j}$ là xác suất xảy ra A và A' ở pha i trong lần thử nghiệm thứ j .

$$FP_j = \begin{cases} 1 & \text{nếu } A_{ij} \text{ và } A'_{ij} \in \text{Pha } i \\ 0 & \text{nếu } A_{ij} \text{ hoặc } A'_{ij} \notin \text{Pha } i \end{cases} \quad (i=1,3)$$

Như vậy, $FP_j=1$ khi dự báo đúng và $FP_j=0$ khi dự báo sai ở lần thử thứ j . Sau N lần thử nghiệm chỉ tiêu để đánh giá mô hình sẽ là:

$$FI = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^3 PA_{i,j} . PA'_{i,j} \quad (10)$$

Với công thức trên, FI chính là tỷ lệ thành công (HR) hay phần trăm đúng (PC) nêu ra trong công thức (9) ở trên và tương tự như công thức (13 và 16) với dự báo hiện tượng hay dự báo 2 pha. FI càng lớn mô hình càng có độ chính xác cao. Với dự báo 3 pha thì xác suất khí hậu là 0,333. Vậy, mô hình dự báo chỉ có hiệu quả khi $FI > 0,333$. Một cách tương tự, theo [10] thì "bảng liên kết" đối với mô hình 3 pha được nêu ở bảng 3.

Bảng 3. Bảng liên kết đối với mô hình 3 pha sau N lần dự báo

		Dự báo (A')			
		Pha I	Pha II	Pha III	Tổng số
Quan trắc (A)	Pha I	N11	N12	N13	N1A
	Pha II	N21	N22	N23	N2A
	Pha III	N31	N32	N33	N3A
	Tổng số	N1A'	N2A'	N3A'	N

Tương tự như trên, với giả thiết là các pha có cùng xác suất nên ta cũng đi đến khái niệm "tỷ lệ thành công" và "tỷ lệ báo động sai" trung bình như sau:

- Tỷ lệ thành công trung bình HR_{TB}

$$FI = HR_{TB} = \frac{1}{3} \sum_{i=1}^3 \frac{N_{ii}}{N} = \frac{1}{N} \sum_{i=1}^3 N_{ii} \quad (11)$$

- Tỷ lệ báo động sai trung bình FR_{TB}

$$FR_{TB} = \frac{1}{3} \sum_{i=1}^3 \sum_{j \neq i} \frac{N_{ij}}{N_{iA}} = \sum_{i=1}^3 \sum_{j \neq i} \frac{N_{ij}}{N} \quad (12)$$

Như vậy, mô hình được coi là "tinh xảo" nếu $HR_{TB} > FR_{TB}$. Trong công trình [10] các tác giả cũng đưa ra chỉ số HR tương tự như trên đối với mô hình 3 pha. Với mô hình 3 pha các tác giả còn đưa ra chỉ số Heidke (Heidke Skill Score - HSS)

$$HSS = \frac{(N_{11} + N_{22} + N_{33}) - \frac{N_{1A} \cdot N_{1A'} + N_{2A} \cdot N_{2A'} + N_{3A} \cdot N_{3A'}}{N} \quad (13)$$

Trong trường hợp tổng quát, mô hình dự báo cho ĐTDB có M pha ta sẽ lập được "bảng liên kết" có dạng:

Bảng 4. Bảng liên kết với M pha

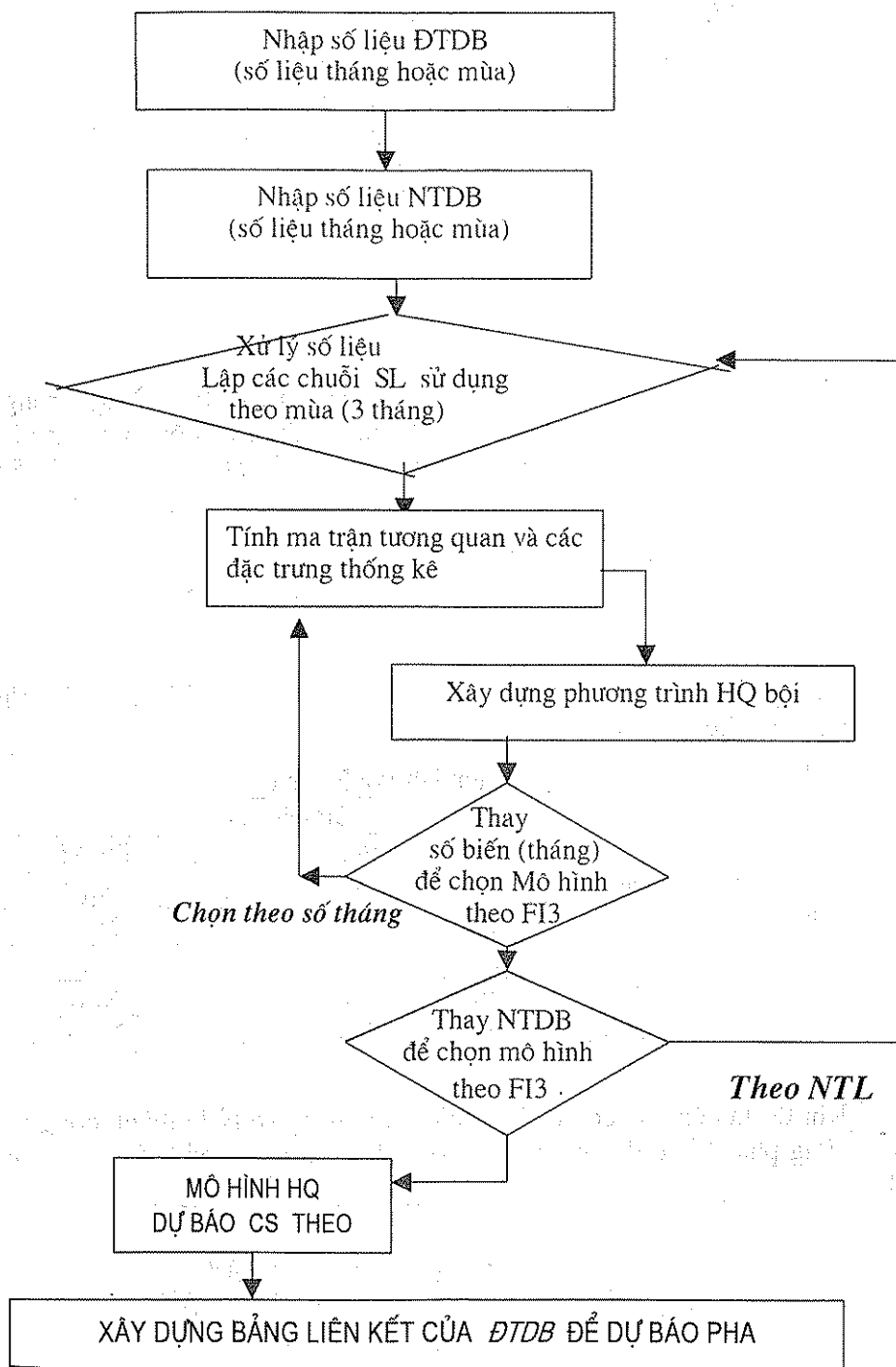
		Dự báo (A')				
Quan trắc (A)		Pha1	Pha2		PhaM	Tổng
	Pha1	N_{11}	N_{12}		N_{1M}	N_{1A}
	Pha2	N_{21}	N_{22}		N_{2M}	N_{2A}
						
	PhaM	N_{M1}	N_{M2}		N_{MM}	N_{MA}
	Tổng	$N_{1A'}$	$N_{2A'}$		$N_{MA'}$	N

Khi đó ta cũng sẽ có các khái niệm tương tự về tỷ lệ thành công và báo động sai cho từng pha, tỷ lệ thành công và báo động sai trung bình đại diện chung cho cả mô hình:

$$FI = HR_{TB} = \frac{1}{M} \sum_{i=1}^M \frac{N_{ii}}{N} = \frac{1}{N} \sum_{i=1}^M N_{ii} \quad (14)$$

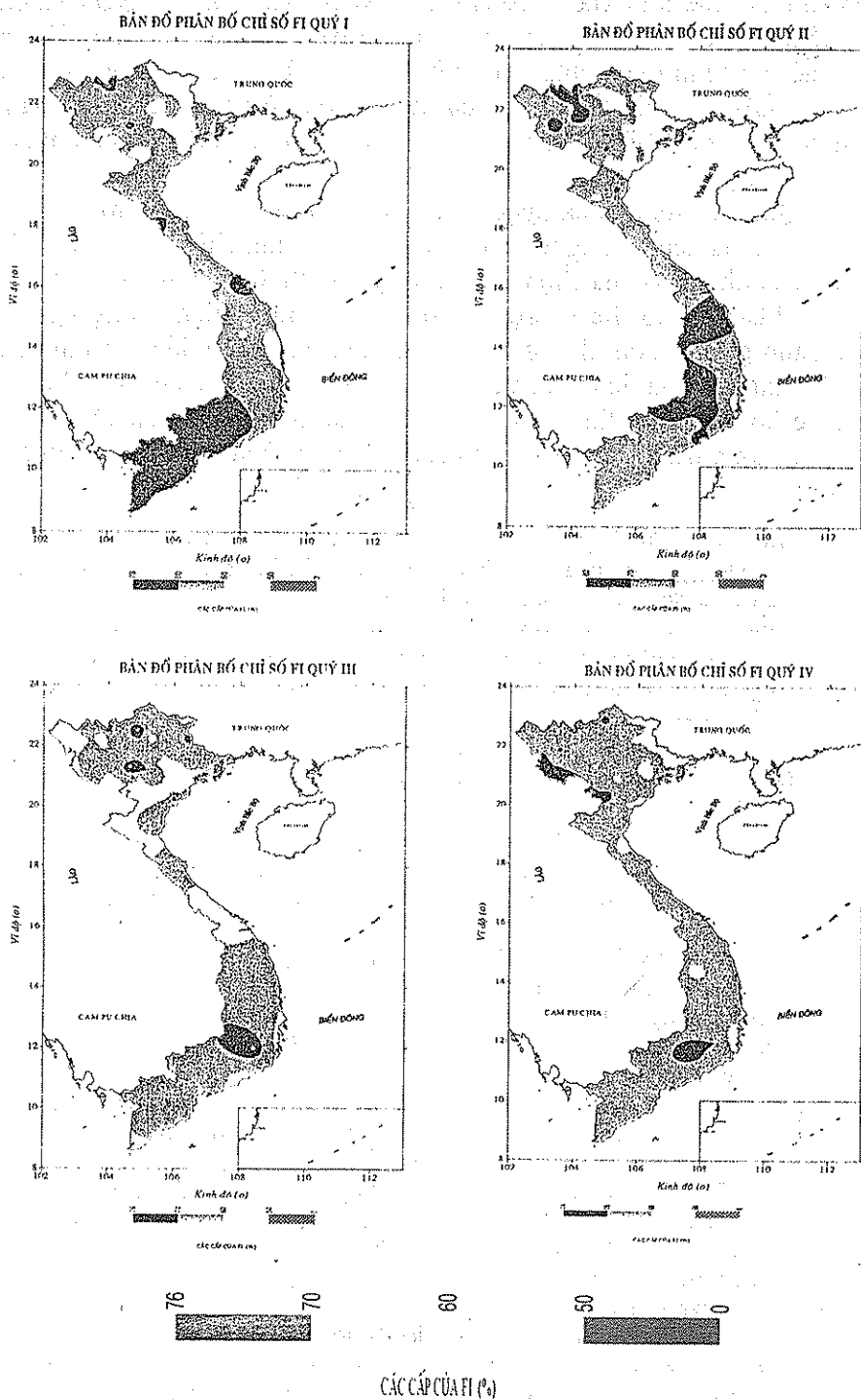
Như đã nêu ở trên, với mô hình M pha thì xác suất khí hậu đối với mỗi pha sẽ là $1/M$. Khi một pha nào đó có tỷ lệ thành công $HR > 1/M$ thì kết quả dự báo được coi là có ý nghĩa đối với pha đó. Song, nếu đánh giá cho cả mô hình thì phải có tỷ lệ thành công trung bình vượt giá trị trên mới coi là có ý nghĩa dự báo. Nhưng để được

coi là "tính xảo" để có thể ứng dụng được vào dự báo thì phải bảo đảm chỉ tiêu $HR_{TB} > FR_{TB}$ tức là $HR_{TB} > 0,5$ (hoặc 50%). Trong nghiên cứu và kiểm định mô hình dự báo hiện nay, người ta còn đưa ra một số chỉ tiêu và phương pháp khác mà trong



Hình 1. Sơ đồ tổng quát xây dựng mô hình dự báo khí hậu dựa trên phương pháp hồi quy nhiều biến

bài này chưa có điều kiện đề cập đến. Song, chỉ tiêu FI hay HR đã nêu ở trên được coi là chỉ tiêu cơ bản và được sử dụng phổ biến để đánh giá và lựa chọn các mô hình theo những phương án khác nhau.



Hình 2. Chỉ số FI của các mô hình dự báo chuẩn sai lượng mưa các quý

Trong công trình [3,4], chúng tôi sử dụng FI như là tiêu chuẩn để lựa chọn mô hình khi xây dựng mô hình dự báo mùa đối với chuẩn sai nhiệt độ và lượng mưa. Hình 1 cho ta sơ đồ thực hiện việc phân tích và tuyển chọn các NTDB trên máy vi tính bằng phương pháp hồi quy nhiều biến. Số biến được chọn sao cho kết quả dự báo theo mô hình đạt được FI lớn nhất. Với các NTDB đã chọn là chuẩn sai nhiệt độ mặt nước biển (SSTA) các khu vực NINO1+2, NINO3, NINO3.4, NINO4 và chỉ số dao động nam (SOI) của 12 tháng trước mùa cần dự báo, kết quả cho thấy FI đối với dự báo 3 pha, các mô hình được chọn hầu hết đều đạt 50-60%, trong đó có một tỷ lệ đáng kể đạt trên 60%. Độ dài chuỗi số liệu được chọn để đánh giá 35-50 năm. Hình 2 cho phân bố giá trị FI đối với các mô hình hồi quy nhiều biến dự báo mưa 4 quý của lưới trạm đã sử dụng.

Qua các kết quả đã được nghiên cứu và thử nghiệm ở một số nước cũng như ở nước ta như đã nêu, chúng tôi cho rằng việc sử dụng chỉ số dự báo FI=HR cho dự báo 3 pha để đánh giá và lựa chọn mô hình dự báo khí hậu là thích hợp và có hiệu quả. Mô hình có FI>50% có thể sử dụng vào dự báo thử nghiệm trong nghiệp vụ. Mô hình được đánh giá tốt hơn khi nó có FI lớn hơn. Nếu có nhiều mô hình cần lựa chọn thì FI=Max được coi là tiêu chuẩn để chọn có hiệu quả. Các chỉ số HSS, RMES có thể sử dụng để hỗ trợ thêm khi FI đạt được kết quả như nhau. Dựa vào các chỉ số vừa nêu, trong các công trình [3,4] đã đi đến các mô hình và kết quả dự báo thử nghiệm.

Tài liệu tham khảo

1. **Trần Việt Liễn 2000.** ENSO với dự báo khí hậu. Hội thảo khoa học lần II đề tài NCKH cấp nhà nước: " Tác động của ENSO đến thời tiết khí hậu, môi trường và kinh tế xã hội Việt Nam". Hà Nội, 29-VI-2000 (tr 50-68).
2. **Trần Việt Liễn 2001.** Một số kết quả nghiên cứu bước đầu về việc sử dụng thông tin ENSO vào dự báo khí hậu ở Việt Nam. Hội thảo khoa học lần III đề tài NCKH cấp nhà nước: " Tác động của ENSO đến thời tiết khí hậu, môi trường và kinh tế xã hội Việt Nam". Hà Nội, IV-2001 (tr 44-68).
3. **Trần Việt Liễn 2001.** Xây dựng mô hình dự báo mùa bằng phương pháp hồi quy nhiều biến. Báo cáo tại hội thảo của đề án "Nghiên cứu thử nghiệm dự báo khí hậu ở Việt Nam", IX-2001. (44 tr).
4. **Trần Việt Liễn và CTV 2001.** Ứng dụng kết quả nghiên cứu bước đầu về hiện tượng ENSO vào dự báo khí hậu ngắn hạn ở Việt Nam. Tổng kết đề tài nghiên cứu cấp Nhà nước "Tác động của hiện tượng ENSO đến khí hậu thời tiết và kinh tế xã hội ở Việt Nam". Viện KTTV (120 tr).
5. **D.J Carton, 1997.** Seasonal forecast. Climate Research technical note No.80. Harley Center for Climate Prediction and Research. London (27p).
6. **William Gray, 1994** The use of ENSO information in Hurricane forecasting. Report of Workshop " The Potential Use and Misuse of El Nino information in North America. 31/X-3/XI 1994 Boulder, Colorado.
7. **D.E. Harrison and Graham, 2001** Forecast quality, forecast application and forecast value: cases from southern African seasonal forecast. WMO Bulletin. Vol. 50 No 3, July 2001. (p 228-231).
8. **Tran Viet Lien, 1998.** Relation between ENSO and Tropical cyclone activity in the Northwest Pacific and Vietnam. Proceedings of "Asian Pacific Friend and Game joint workshop on ENSO, floods and droughts in the 1990's in Southeast Asia and the Pacific. Hanoi 23-26 March 1999 (p 118-125).
9. **Guiling Wang and Elfatih A.B Eltahir 1999** Use of ENSO information in Medium- and Long-range Forecasting of the Nile floods. Journal Climate. Vol 12 (p 1726-1737)
10. **WMO, 1989** Survey of common verification methods in meteorology. (by H.R Stanski, Laurence J. Wilson and William R. Burrows). World weather watch. Technical report No 8. WMO/TD NO 358. Ontario, Canada (114 p)