

THỬ NGHIỆM BƯỚC ĐẦU DỰ BÁO MƯA MÙA Ở VIỆT NAM

TS. Nguyễn Văn Thắng

Trung tâm Nghiên cứu Khí hậu, Viện KTTV

1. Mở đầu

Dự báo khí hậu, trong đó có dự báo mưa mùa đang được sự quan tâm rất lớn của Tổng cục Khí tượng Thủy văn (KTTV) thông qua triển khai đề án Dự báo khí hậu và các nội dung hợp tác về dự báo khí hậu giữa Việt Nam với Mỹ và Úc trong khuôn khổ hợp tác song phương về khí tượng thủy văn. Dự báo mưa mùa trên thế giới hiện tại có thể được thực hiện thông qua các cách sau:

- Trực tiếp từ phương pháp thống kê,
- Trực tiếp từ mô hình khí quyển với trường nhiệt độ mặt nước biển được dự báo,
- Trực tiếp từ mô hình kép khí quyển - đại dương,
- Gián tiếp từ kết quả các mô hình khí quyển và mô hình kép thông qua phương pháp hạ thấp phạm vi (downscaling).

Đối với các nước đang phát triển thì phương pháp thống kê thực nghiệm trong dự báo khí hậu được coi là ưu việt hơn cả vì tính đơn giản, dễ sử dụng, ít tốn kém, có kết quả dự báo nhanh,... Tuy nhiên, việc xác định các nhân tố độc lập đầu vào trong tập hợp các dãy số liệu khí hậu quan trắc được để xây dựng mô hình dự báo cho yếu tố cần dự báo (nhiệt, mưa mùa) mang tính quyết định. Đơn giản nhất là tìm kiếm mối quan hệ giữa các yếu tố khí hậu hoặc giữa yếu tố khí hậu cần dự báo với các nhân tố quan trắc được liên quan tới ENSO như nhiệt độ nước biển bề mặt vùng En Nino 1, 2, 3, 4, 1+2, 3+4; chỉ số SOI; chỉ số MEI;... để xây dựng mô hình dự báo. Nhưng nhiều khi không thể tìm thấy được mối quan hệ chặt chẽ để có thể xây dựng mô hình dự báo, nhất là chỉ sử dụng số liệu của riêng từng trạm, vì một trạm chỉ mang tính địa phương, không thể hiện tác động của các quá trình khí hậu phạm vi lớn như ENSO, ngoài ra còn bị ảnh hưởng của sai số thiết bị quan trắc. Do đó, để có bức tranh rõ nét về dao động khí hậu phải sử dụng các chỉ số trung bình khu vực của các vùng đồng nhất.

Hiện nay phương pháp dự báo khí hậu được sử dụng rộng rãi nhất là dùng các chỉ số khí hậu như mưa, nhiệt độ của các vùng đồng nhất làm yếu tố dự báo và các trường khí tượng có được từ kết quả của các mô hình động lực toàn cầu hoặc khu vực như nhiệt độ mặt nước biển (SST), khí áp mặt biển (SLP), tốc độ và hướng gió ở các mực..., làm nhân tố dự báo.

Tóm lại, ở các nước phát triển thường kết hợp chạy đồng thời mô hình khí hậu động lực và mô hình thống kê thực nghiệm để đưa ra bản tin dự báo khí hậu tốt hơn, còn ở các nước đang phát triển cần sử dụng các kết quả đầu ra của các mô hình động lực để xây dựng các mô hình dự báo thống kê thực nghiệm.

Trong nội dung bài báo này, tác giả sẽ trình bày về khả năng sử dụng các trường SST và SLP toàn cầu để thử nghiệm dự báo chỉ số mưa mùa cho miền Bắc và miền Trung Việt Nam.

2. Cơ sở dữ liệu và phương pháp

a. Cơ sở dữ liệu

- Số liệu trung bình tháng về lượng mưa và nhiệt độ của các trạm khí tượng khí hậu, từ năm 1951 đến năm 2000 (dưới dạng text *.txt);

- Số liệu trung bình tháng của tất cả các ô lưới về chuẩn sai SST, SLP, bức xạ sóng dài (ORL), hướng và tốc độ gió ở các mức, từ năm 1951 đến năm 2000 (dưới dạng nhị phân *.bin);

Trong nội dung thử nghiệm chỉ mới sử dụng số liệu mưa trung bình tháng từ năm 1960 đến 2000 của 7 trạm là Sơn La, Sa Pa, Hòn Dấu, Hà Nội, Thanh Hoá, Pleiku, Nha Trang; chuẩn sai SST và SLP.

b. Phương pháp

- Lựa chọn/ chia ra các vùng đồng nhất: mỗi vùng nên có ít nhất 5 trạm với đầy số liệu trên 30 năm (ít nhất cũng phải từ 25 năm). Có nhiều cách được sử dụng để chia ra các vùng đồng nhất, nhưng cách thường dùng nhất vẫn là sử dụng phương pháp phân tích thành phần chính (PCA),

- Tính toán chỉ số trung bình khu vực là giá trị trung bình các chỉ số của tất cả các trạm trong vùng theo công thức:

$$X_j = \frac{1}{N_j} \sum_{i=1}^{N_j} \frac{R_{ij} - \overline{R_i}}{\sigma_i}$$

Trong đó:

R_{ij} - lượng mưa mùa; R_i - lượng mưa mùa trung bình của trạm i theo thời kỳ chuẩn; σ_i - độ lệch chuẩn của lượng mưa mùa tại trạm i ; N_j - số trạm có sẵn trong vùng của năm thứ j .

- Dùng chương trình Grads.exe để tính và vẽ các đường đẳng trị hệ số tương quan giữa chỉ số mưa mùa với chuẩn sai SST hoặc SLP của các ô lưới theo các tháng khác nhau nhằm tìm kiếm và xác định vùng - thời gian có mối tương quan tốt nhất.

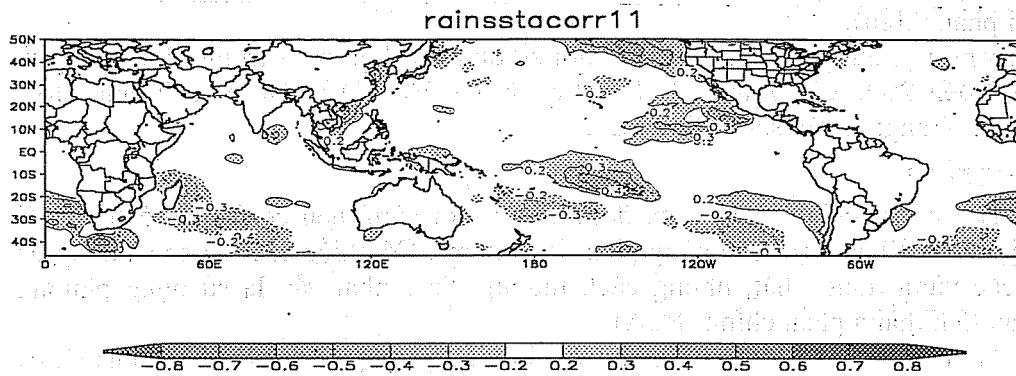
- Dùng chương trình Systat.10 để xây dựng mô hình dự báo hồi qui và đánh giá mô hình bằng phương pháp thẩm tra dự báo (forecast verification or Cross - Validation) thông qua bảng ngẫu nhiên (contingency table) [1,2,3].

- Truy cập số liệu trên Internet dưới dạng nhị phân và dùng chương trình Grads để lấy số liệu tháng năm đã chọn của nhân tố dự báo (chuẩn sai SST và SLP) tính giá trị của yếu tố cần dự báo (chỉ số mưa mùa) với xác suất xác định thông qua Bảng ngẫu nhiên.

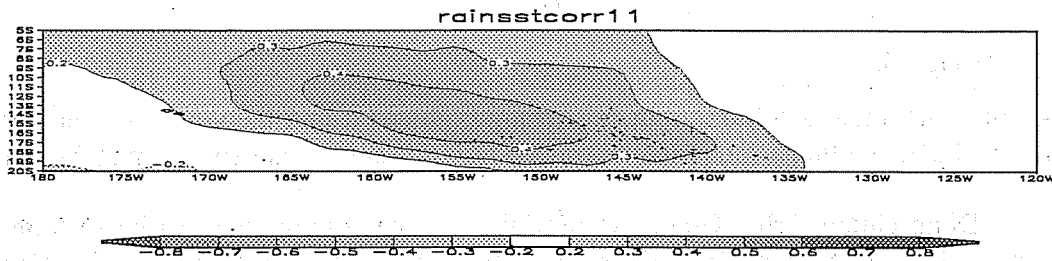
3. Quá trình thử nghiệm

Trong nội dung thử nghiệm, tác giả chỉ sử dụng số liệu mưa tháng từ năm 1960 — 2000 của 7 trạm và không tiến hành phân tích thành phần chính, mà chỉ đơn thuần phân thành 2 vùng: vùng miền Bắc gồm trạm Sơn La, Sa Pa, Hòn Dấu, Hà Nội; vùng miền Trung gồm Thanh Hoá, Pleiku, Nha Trang. Mùa mưa ở Việt Nam thường kéo dài từ tháng V đến hết tháng X, nhưng ở đây tác giả chỉ dừng lại việc xây dựng mô hình dự báo thử nghiệm cho 3 tháng đầu mùa mưa (tháng V, VI, VII) — nửa đầu mùa mưa và tính chỉ số mưa mùa cho từng vùng: miền Bắc và miền Trung từ năm 1960 đến năm 2000. Chạy chương trình Grads để xác định mối tương quan giữa chỉ số mưa mùa của từng vùng với chuẩn sai SST và SLP theo lưới toàn cầu của trung bình các tháng 1, 2, 3, 4 từ năm 1960 đến năm 2000 và chọn ra khu vực và thời gian có mối tương quan tốt nhất được thể hiện trên hình 1a, b; 2a, b (đối với vùng miền Bắc) và hình 3a, b; 4a, b (đối với vùng miền Trung).

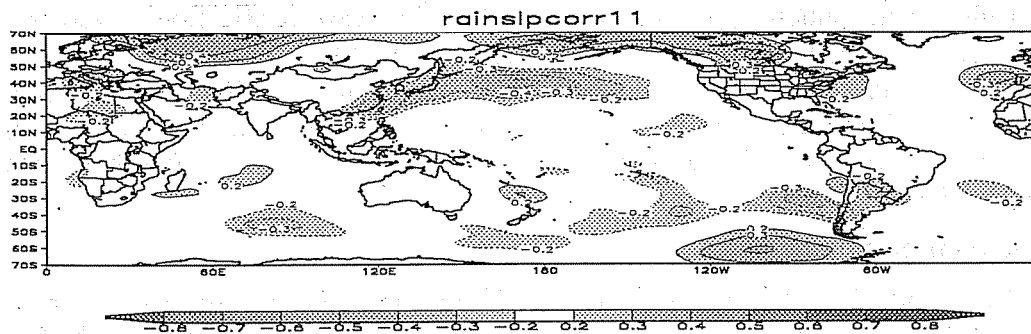
Hình 1. Bản đồ đường đẳng trị hệ số tương quan giữa chỉ số mưa mùa miền Bắc và chuẩn sai SST tháng IV từ 1960 đến 2000 (a)



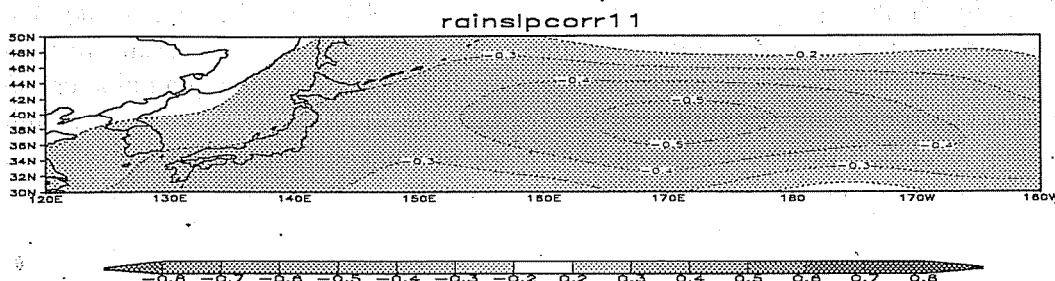
(b). Khu vực có mối tương quan tốt nhất



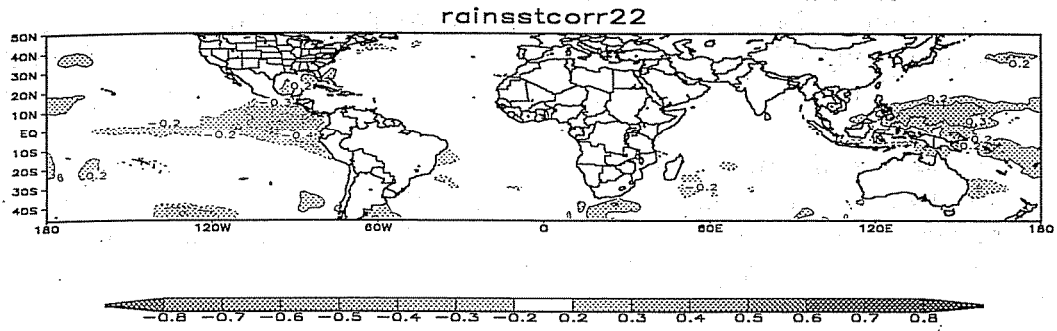
Hình 2. Bản đồ đường đẳng trị hệ số tương quan giữa chỉ số mưa mùa miền Bắc và chuẩn sai SLP tháng IV từ 1960 đến 2000 (a)



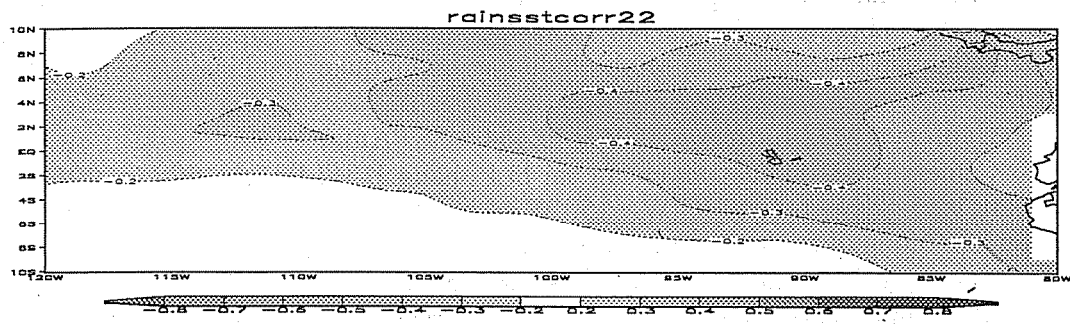
(b). Khu vực có mối tương quan tốt nhất



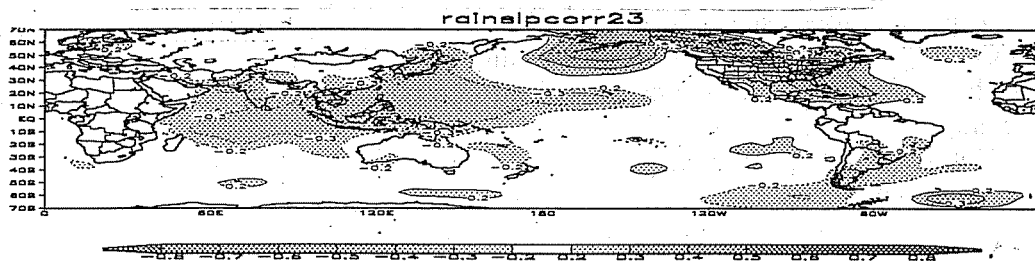
Hình 3. Bản đồ đường đẳng trị hệ số tương quan giữa chỉ số mưa mùa miền Trung và chuẩn sai SST tháng III từ 1960 đến 2000 (a)



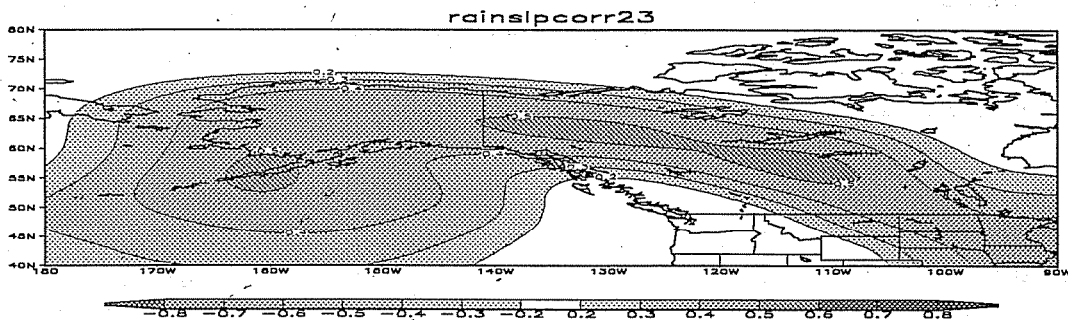
(b). Khu vực có mối tương quan tốt nhất



Hình 4. Bản đồ đường đẳng trị hệ số tương quan giữa chỉ số mưa mùa miền Trung và chuẩn sai SLP tháng II từ 1960 đến 2000 (a)



(b). Khu vực có mối tương quan tốt nhất



Lấy được số liệu chuẩn sai SST và SLP của khu vực và tháng đã chọn, dùng Chương trình Systat 10 để lập mô hình tương quan dự báo cho từng vùng.

Chỉ số mưa mùa tháng V-VII của vùng miền Bắc (Northvn MJJ indices) và chuẩn sai SST tháng IV (SSTAapr) từ 1960 - 2000 khu vực: 16°S - 12°S; 160°W - 147°W có hệ số tương quan $R = +0,428$ và Mô hình dự báo tương quan đơn giản 1.T (SRPM1.T) có dạng:

$$\text{Northvn MJJ indices} = -0,104 + 0,745 \times \text{SSTAapr}$$

Northvn MJJ rainfall indices và chuẩn sai SLP tháng IV (SLPAapr) từ 1960 — 2000 khu vực: 35°N - 42°N; 160°E - 180°E có hệ số tương quan $R = -0,519$ và Mô hình dự báo tương quan đơn giản 1.P (SRPM1.P) là:

$$\text{Northvn MJJ indices} = -0,089 - 0,001 \times \text{SLPAapr}$$

Mô hình dự báo tương quan đa biến 1.T,P (MRPM1.T,P) có hệ số tương quan bội $R = 0,602$ có dạng:

$$\text{Northvn MJJ indices} = -0,150 + 0,547 \times \text{SSTAapr} + 0,001 \times \text{SLPAapr}$$

Chỉ số mưa mùa tháng V-VII của vùng miền Trung (Centervn MJJ indices) và chuẩn sai SST tháng III (SSTamar) từ 1960 - 2000 khu vực: 0° - 5°N; 98°W - 87°W có hệ số tương quan $R = -0,447$ và Mô hình dự báo tương quan đơn giản 2.T (SRPM2.T) có dạng:

$$\text{Centervn MJJ indices} = 0,013 - 0,464 \times \text{SSTamar}$$

Centervn MJJ indices và chuẩn sai SLP tháng II (SLPAfeb) thời kỳ 1960 — 2000 khu vực: 47°N - 60°N; 165°W - 150°W có hệ số tương quan $R = +0,4873$ và Mô hình dự báo tương quan đơn giản 2.P (SRPM2.P) là:

$$\text{Centervn MJJ indices} = 0,0174 + 0,0005 \times \text{SLPAfeb}$$

Mô hình dự báo tương quan đa biến 2.T,P (MRPM2.T,P) có hệ số tương quan bội $R = 0,5617$ có dạng:

$$\text{Centervn MJJ indices} = 0,0244 - 0,3151 \times \text{SSTamar} + 0,0004 \times \text{SLPAfeb}$$

Để xác định chất lượng các mô hình dự báo, chúng tôi sử dụng phương pháp thẩm tra dự báo (forecast verification or Cross - Validation) thông qua Bảng ngẫu nhiên (Contingency table) bao gồm cả việc đánh giá mối quan hệ giữa kết quả dự báo hoặc tổ hợp kết quả dự báo (forecast) với dữ liệu quan trắc của yếu tố cần dự báo (predictant).

Kết quả tính toán theo bảng ngẫu nhiên như sau:

- Đối với SRPM1.T

	ForBelow	ForNor	ForAbove	Total
ObsBelow	8	2	4	14
ObsNor	4	5	4	13
ObsAbove	4	2	8	14
Total	16	9	16	41

- Đối với SRPM1.P

	ForBelow	ForNor	ForAbove	Total
ObsBelow	8	3	3	14
ObsNor	3	7	3	13
ObsAbove	3	5	6	14
Total	14	15	12	41

• Đối với MRPM1.T,P

	ForBelow	ForNor	ForAbove	Total
ObsBelow	8	3	3	14
ObsNor	4	4	5	13
ObsAbove	3	5	6	14
Total	15	12	14	41

• Đối với SRPM2.T

	ForBelow	ForNor	ForAbove	Total
ObsBelow	5	4	5	14
ObsNor	4	3	6	13
ObsAbove	0	4	10	14
Total	9	11	21	41

• Đối với SRPM2.P

	ForBelow	ForNor	ForAbove	Total
ObsBelow	6	5	3	14
ObsNor	5	3	5	13
ObsAbove	3	3	8	14
Total	14	11	16	41

• Đối với MRPM2.T,P

	ForBelow	ForNor	ForAbove	Total
ObsBelow	7	2	5	14
ObsNor	3	6	4	13
ObsAbove	2	3	9	14
Total	12	11	18	41

Các đánh giá so sánh giữa các mô hình được trình bày trong bảng 1 và 2.

Bảng1. Các chỉ tiêu đánh giá độ chính xác của dự báo
phân theo đa loại đối với Northvn MJJ indices

Chỉ tiêu	SRPM1.T	SRPM1.P	MRPM1.T,P	Ghi chú
Hit Rate (H)	51.2%	51.2%	44%	MRPM1.T,P Kém
Heidke hit skill score (HSS)	27%	27%	16.4%	
Probability of detection for each tercile:				SRPM1.T Tốt hơn
-Below Normal Category	57%	57%	57%	
-Above Normal Category	57%	43%	43%	
False Alarm Rate (FAR) for each tercile:				
-Below Normal Category	25%	21%	20%	
-Above Normal Category	25%	25%	21%	
Linear Error in Probability Space (LEPS)	28%	28%	24%	
Coefficient of correlation	+0,428	-0,519	0,602	

Bảng 2. Các chỉ tiêu đánh giá độ chính xác của dự báo phân theo đa loại đối với Centervn MJJ indices

Chỉ tiêu	SRPM2.T	SRPM2.P	MRPM2.T,P	Ghi chú
Hit Rate (H)	44%	41,5%	53,7%	MRPM2.T,P Tốt nhất
Heidke hit skill score (HSS)	16,4%	12,7%	30,9%	
Probability of detection for each tercile:				
-Below Normal Category	35,7%	42,9%	50,0%	
-Above Normal Category	71,4%	57,1%	64,3%	
False Alarm Rate (FAR) for each tercile:				
-Below Normal Category	0%	21,4%	16,7%	
-Above Normal Category	23,8%	18,8%	27,8%	
Linear Error in Probability Space (LEPS)	29,3%	23,3%	31,6%	
Coefficient of correlation	-0,447	+0,4873	0,5617	

Các chỉ tiêu so sánh từ bảng 1 và 2 cho thấy mô hình MRPM1.T,P có thể không sử dụng được cho dự báo, mặc dù hệ số tương quan bội $R = 0,602$, vì chỉ số Hit rate thấp ($H = 44\%$). Mô hình SRPM1.T và SRPM1.P có thể sử dụng cho dự báo, nhưng mô hình SRPM1.T có chỉ số Probability of detection for Above Normal Category (57%) cao hơn so với mô hình SRPM1.P (43%). Chỉ số False Alarm Rate (FAR) for the Below Normal Category của mô hình SRPM1.P là 21% , thấp hơn FAR của mô hình SRPM1.T (25%).

Theo bảng 2, chúng tôi chọn mô hình MRPM2.T,P có chỉ số Hit Rate (H) cao nhất ($53,7\%$) và Linear Error in Probability Space (LEPS) là $31,6\%$. Các mô hình SRPM2.T và SRPM2.P có thể không sử dụng vì chỉ số Hit Rates rất thấp, thậm chí hệ số tương quan tốt, tương ứng là $-0,447$ và $+0,487$.

Như vậy, chúng ta tìm được mối tương quan của chỉ số mưa mùa (tháng V, VI, VII) của vùng miền Bắc với chuẩn sai SST tháng IV khu vực: $16^{\circ}\text{S} - 12^{\circ}\text{S}$; $160^{\circ}\text{W} - 147^{\circ}\text{W}$ (hệ số tương quan $R = +0,428$) và với chuẩn sai SLP tháng IV khu vực: $35^{\circ}\text{N} - 42^{\circ}\text{N}$; $160^{\circ}\text{E} - 180^{\circ}\text{E}$ (hệ số tương quan âm $R = -0,519$), và kiến nghị sử dụng mô hình SRPM1.T để tiến hành thử nghiệm dự báo mưa mùa miền Bắc theo phương trình:

$$\text{Northvn MJJ indices} = -0,104 + 0,745 \times \text{SSTA}_{\text{Apr}}$$

Thêm vào đó có thể sử dụng mô hình SRPM1.P để tham khảo và xem xét bổ sung kết quả dự báo của mô hình SRPM1.T

Đối với vùng miền Trung thì quan hệ của chỉ số mưa mùa cùng thời kỳ với chuẩn sai SST tháng III khu vực: $0^{\circ} - 5^{\circ}\text{N}$; $98^{\circ}\text{W} - 87^{\circ}\text{W}$ (hệ số tương quan âm $R = -0,447$); và với chuẩn sai SLP tháng II khu vực: $47^{\circ}\text{N} - 60^{\circ}\text{N}$; $165^{\circ}\text{W} - 150^{\circ}\text{W}$ (hệ số tương quan $R = +0,4873$) được sử dụng lập mô hình dự báo MRPM2.T,P để thử nghiệm theo phương trình:

$$\text{Centervn MJJ indices} = 0,0244 - 0,3151 \times \text{SSTA}_{\text{Mar}} + 0,0004 \times \text{SLPA}_{\text{Feb}}$$

Tuy nhiên, phải thừa nhận việc dự báo mưa mùa ở Việt Nam rất phức tạp bởi ảnh hưởng của cả hệ thống hoàn lưu hành tinh và các sự kiện phạm vi lớn như gió mùa/ hoàn lưu Haddley, ENSO/ hoàn lưu Walker, xoáy thuận nhiệt đới, không khí lạnh, dải hội tụ nhiệt đới,.... Trong phạm vi bài này chỉ trình bày thời gian (tháng) và khu vực có mối quan hệ tốt nhất giữa yếu tố cần dự báo được chọn và các nhân tố dự

báo, chứ chưa đưa ra những diễn giải về hình thế synóp và cơ sở vật lý của mối tương quan tìm thấy được ở trên.

Tất nhiên các trạm và số liệu thử nghiệm ở đây cũng không đặc trưng được cho toàn vùng miền Bắc và miền Trung, do đó các chỉ số mưa mùa và tiếp theo là giá trị tương quan và mô hình dự báo có thể cần được xem xét và điều chỉnh nhiều. Cố gắng của chúng tôi ở đây là mong muốn đưa ra thử nghiệm bước đầu lập mô hình dự báo bằng phương pháp thống kê dựa theo kết quả đầu ra — các trường khí tượng của mô hình toàn cầu và khu vực.

4. Kết quả thử nghiệm dự báo

Để dự báo chỉ số mưa mùa (tháng V, VI, VII) năm 2001 cho vùng miền Bắc và miền Trung theo các mô hình dự báo trên, chúng ta cần biết được giá trị chuẩn sai SST và SLP tháng IV năm 2001 đối với miền Bắc và chuẩn sai SST tháng III-2001 và SLP tháng II-2001 đối với miền Trung.

Truy cập số liệu từ mạng Internet (<http://www.cptec.inpe.br/>) và chạy Chương trình Grads để cập nhật số liệu chuẩn sai SST và SLP cho khu vực và thời gian đã chọn. Sử dụng số liệu chuẩn sai SST và SLP tháng IV năm 2001 tính được chỉ số mưa mùa tháng V, VI, VII vùng miền Bắc theo mô hình SRPM1.T:

$$\begin{aligned}\text{NorthvnMJJ indices} &= -0,104 + 0,745x(\text{SSTA}_{\text{Apr}}) = \\ &= -0,104 + 0,745x(-0,01092) = -0,857.\end{aligned}$$

Dựa vào Bảng ngẫu nhiên (Contingency table) đưa ra bản tin dự báo về lượng mưa mùa tháng V, VI, VII thấp hơn trung bình nhiều năm với suất bảo đảm 50%, gần trung bình là 25% và trên trung bình: 25% (hình 5).

Tương tự đối với vùng miền Trung, mô hình MRPM2.T,P được sử dụng:

$$\begin{aligned}\text{Centervn MJJ indices} &= \\ &= 0,0244 - 0,3151x\text{SSTA}_{\text{Mar}} + \\ &+ 0,0004x\text{SLPA}_{\text{Feb}} = 0,0244 - \\ &- 0,3151x0,0401 + 0,0004x4,0665 = \\ &= 0,0134.\end{aligned}$$

Có nghĩa là lượng mưa mùa (tháng V, VI, VII) của miền Trung gần với trung bình nhiều năm với suất bảo đảm 55%; khoảng 27% trên trung bình và 18% dưới trung bình (hình 5).

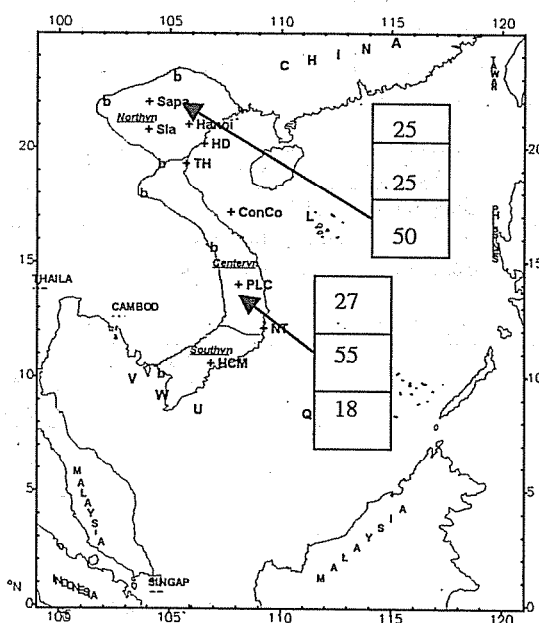


Figure 5

Tài liệu tham khảo

1. Chương trình Systat 10 và Grads gồm bộ sách hướng dẫn sử dụng và đĩa cài đặt. Copyright c 2000 by SPSS Inc., printed in USA.
2. Guidelines on performance assessment of public weather services, WMO/TD, No. 1023 (chapter 4).
3. Report on model - skill by Mouhamadou issa lele, ACMAD in the third workshop on Regional climate prediction and applications- Tropical Pacific Islands and Rim, Oklahoma, 30/4-8/6/2001.