

Dự án Danida

Nghiên cứu thủy tai do biến đổi khí hậu và xây dựng hệ thống thông tin nhiều bên tham gia nhằm giảm thiểu tính dễ bị tổn thương ở Bắc Trung Bộ Việt Nam (CPIS)

Mã số . 11-P04-VIE

Tên đề tài:

Dự án Nghiên cứu thủy tai do biến đổi khí hậu và xây dựng hệ thống thông tin nhiều bên tham gia nhằm giảm thiểu tính dễ bị tổn thương ở Bắc Trung Bộ Việt Nam

Chủ nhiệm dự án: GS. TS. Phan Văn Tân

Báo cáo WP3:

**BÁO CÁO KHOA HỌC VỀ KẾT QUẢ DỰ TÍNH KHÍ HẬU
TƯƠNG LAI, TÍNH BẤT ĐỊNH VÀ PHƯƠNG PHÁP
ĐÁNH GIÁ, XỬ LÝ**

Người thực hiện:

Ngô Đức Thành

Nội dung 2.6:

Báo cáo kĩ thuật phân tích độ bất định

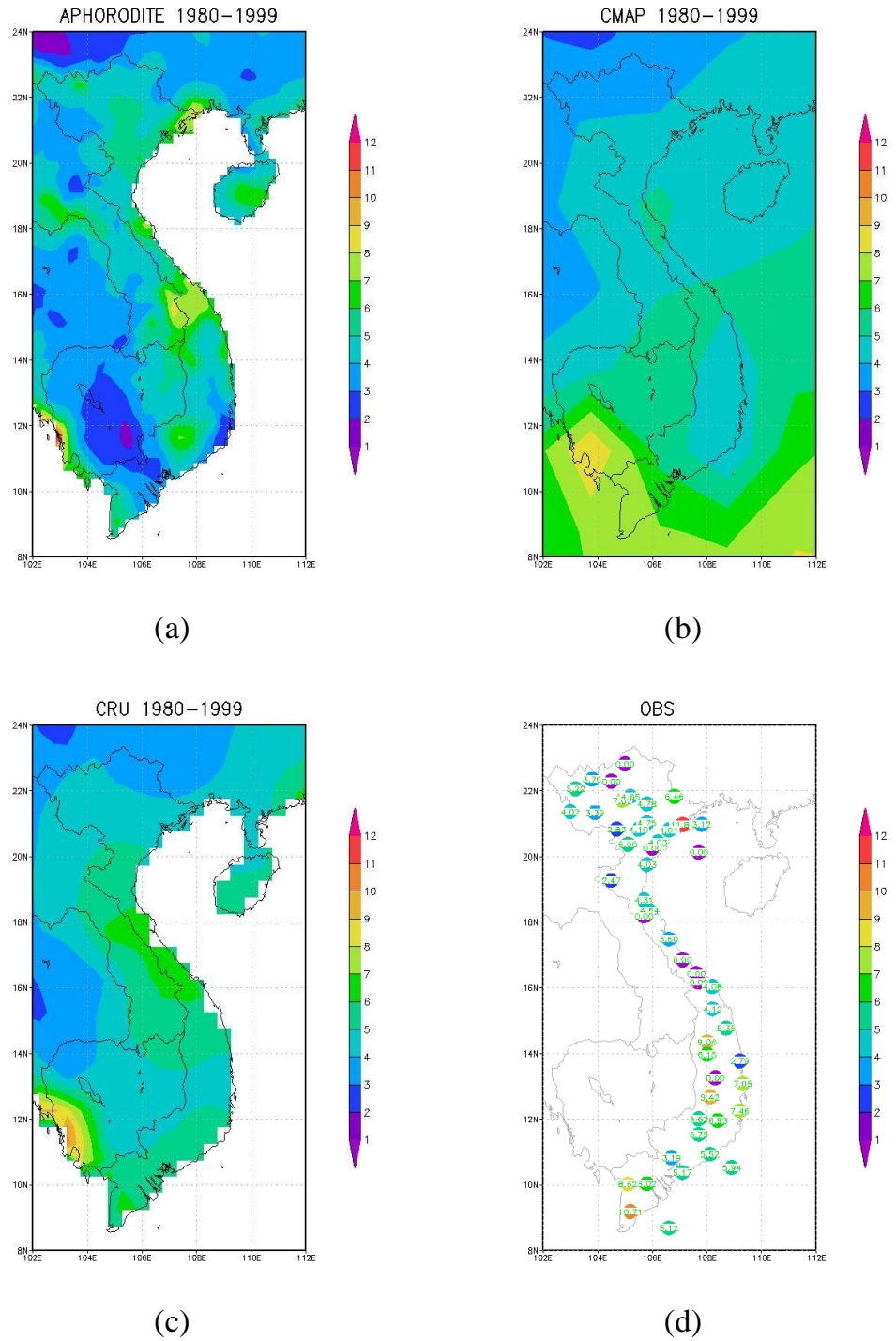
Người thực hiện: Nguyễn Quang Trung

1. Tổng quan về kết quả mô hình

Ngày nay, mô hình khí hậu là một công cụ quan trọng trong nghiên cứu mô phỏng khí hậu cũng như dự tính khí hậu tương lai. Cùng với sự phát triển của khoa học tính toán, các mô hình ngày càng phát triển và trở nên hoàn thiện hơn. Tuy nhiên có một thực tế là không một mô hình nào có thể mô phỏng cũng như dự tính một cách chính xác, mô hình này có thể dự tính tốt cho khu vực này và không tốt cho khu vực khác, hoặc tốt cho mùa này nhưng lại không tốt cho khu vực khác.

Đối với bài toán dự tính khí hậu, nhiều mô hình có thể cho kết quả dự tính khác nhau và trong tương lai thì chưa thể biết mô hình nào hay kết quả nào tin cậy hơn, do mô hình luôn ẩn chứa độ bất định. Đó là lý do vì sao cần tiến hành đánh giá, nghiên cứu ảnh hưởng của độ bất định đến kết quả mô hình. Dưới đây là kĩ thuật phân tích độ bất định của nhiều mô hình với các kết quả khác nhau. Kết quả mô phỏng lượng mưa của 5 mô hình CCAM, MRI, REMO, MM5 và RegCM sẽ được phân tích và so sánh với số liệu trạm trung bình tháng trong giai đoạn 1980-1999. Sau đó, bằng cách xây dựng phương trình hồi quy tuyến tính dựa trên số liệu quan trắc tại trạm và số liệu 5 mô hình được đưa về trạm, một kết quả tổ hợp sẽ được so sánh và đánh giá với số liệu quan trắc.

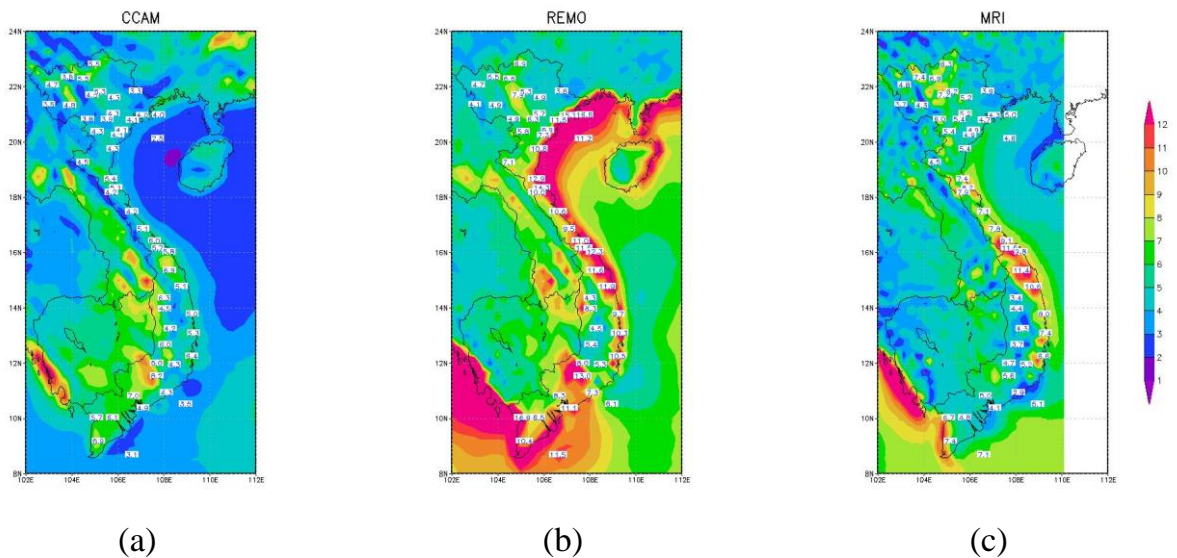
- a. So sánh số liệu mưa trạm với một số sản phẩm dữ liệu mưa

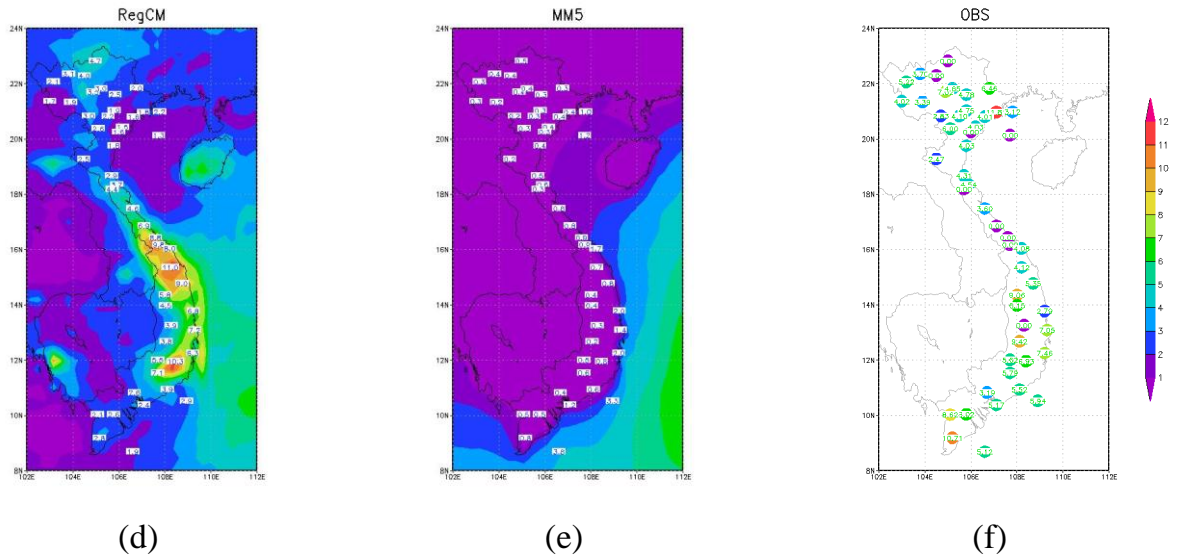


Hình 1: Số liệu mưa trên lưới (a) APHORODITE, (b) CMAP, (c) CRU, (d) quan trắc tại trạm Trước hết, để khảo sát số liệu trạm, có thể dùng một số sản phẩm dữ liệu mưa trên lưới (APHORODITE, CMAP, CRU) để so sánh và phân tích phân bố vùng mưa và lượng mưa trung bình trên các trạm và các khu vực tại Việt Nam.

Kết quả trung bình 20 năm 1980-1999 trên miền Việt Nam cho thấy các kết quả lượng mưa không có sự khác biệt nhiều. Về phương diện không gian, lượng mưa phân bố chủ yếu ở khu vực đồng bằng duyên hải miền Trung. Ở sản phẩm CMAP và CRU do độ phân giải kém hơn nên vùng mưa phân bố không rõ rệt, lượng mưa chủ yếu trong khoảng 5-7mm/ngày, vùng mưa phân bố chủ yếu theo vĩ độ và có vùng cực đại ở Nam Bộ (CMAP) và Trung Bộ (CRU). Ở bộ số liệu APHORODITE với độ phân giải cao hơn, cho lượng mưa phân bố theo vùng rõ rệt hơn. So với hai bộ số liệu còn lại, APHORODITE cho kết quả gần với số liệu mưa trạm nhất, lượng mưa cao hơn ở khu vực đồng bằng ven biển miền Trung và vùng Nam Bộ với lượng mưa phổ biến trong khoảng 7-8mm/ngày và lượng mưa phổ biến 4-6 mm/ngày ở hầu hết các khu vực còn lại. Ở bộ số liệu APHORODITE còn cho thấy rõ lượng mưa cực tiểu ở khu vực Bình Thuận – Phan Thiết, nơi có lượng mưa thấp nhất trên số liệu trạm quan trắc.

b. So sánh số liệu mưa trạm với các số liệu mô hình



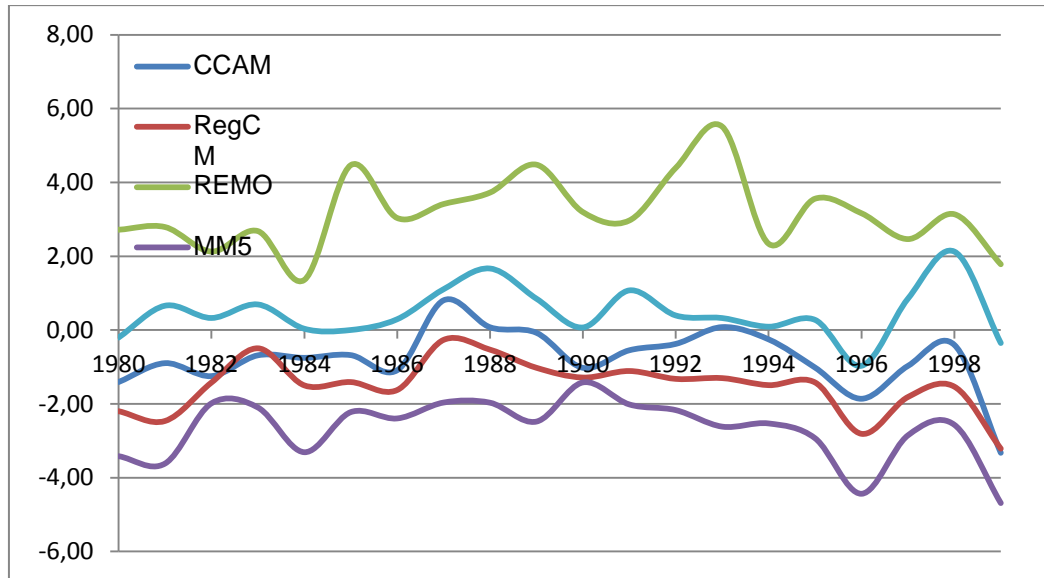


Hình 2. Tổng quan về số liệu mưa các mô hình và số liệu trạm trung bình giai đoạn 1980-1999

Với giá trị trung bình mô phỏng cùng chuỗi thời gian là giai đoạn 1980-1999, kết quả mô phỏng như trên (hình 2) cho thấy kết quả mô phỏng của mô hình REMO và MRI cho giá trị lượng mưa rất lớn ở khu vực ven biển, và dọc khu vực duyên hải miền Trung với lượng phổ biến từ 11-12mm/ngày. Đặc biệt mô hình REMO cho mô phỏng lượng mưa rất lớn (trên 12mm/ngày) ở khu vực Vịnh Bắc Bộ và vùng ven biển Bắc Trung Bộ, những nơi không có số liệu quan trắc để kiểm chứng.

Ngược lại, ở mô hình RegCM và mô hình MM5 lại cho mô phỏng lượng mưa khá thấp, ở mô hình MM5 cho kết quả mô phỏng lượng mưa toàn Việt Nam dưới 2mm/ngày. Ở mô hình RegCM cho kết quả mô phỏng khả quan hơn, nhờ khả năng bắt được vùng mưa ở duyên hải miền Trung, nơi có sự bổ sung lượng mưa lớn do địa hình.

Mô hình CCAM cho kết quả mô phỏng lượng mưa tốt nhất, các đặc trưng vùng và các cực đại địa phương được thể hiện khá rõ nét trên bản đồ. Sự phân vùng mưa trên bản đồ cũng khá rõ ràng và chi tiết.



Hình 1: Giá trị sai số trung bình tất cả các trạm cho 5 mô hình

2. Tổ hợp kết quả bằng phương pháp xây dựng phương trình hồi quy tuyến tính

Phương pháp dự báo sử dụng ở đây là hồi qui tuyến tính, yếu tố dự báo Y được biểu diễn bởi tổ hợp tuyến tính của các nhân tố dự báo X_1, X_2, \dots, X_m bằng hệ thức:

$$Y = a_0 + \sum_{j=1}^m a_j X_j$$

trong đó a_0, a_1, \dots, a_m là các hệ số hồi qui, được xác định bởi:

$$a_0 = \bar{y} - \sum_{j=1}^m a_j \bar{x}_j$$

$$\sum_{j=1}^m a_j R_{jk} = R_{yk}, k = 1, 2, \dots, m$$

với R_{jk} là mômen tương quan giữa các nhân tố X_j và X_k ($j, k=1, 2, \dots, m$), R_{yk} là mômen tương quan giữa yếu tố dự báo Y và nhân tố X_k ($k=1, 2, \dots, m$); \bar{y}, \bar{x}_j là giá trị trung bình của Y và các X_j .

Các nhân tố dự báo thường có quan hệ với nhau, và trong nhiều trường hợp, sự có mặt của nhân tố này dẫn đến sự không cần thiết có mặt của những nhân tố

khác. Vậy cần chọn số nhân tố bằng bao nhiêu và chọn những nhân tố nào trong số các nhân tố dự tuyển làm nhân tố dự báo. Để thực hiện vấn đề này, ở đây em sử dụng phép lọc nhân tố bằng phương pháp hồi qui từng bước.

Xét hồi qui tuyến tính giữa biến phụ thuộc Y và m biến độc lập X_1, \dots, X_m . Để xây dựng phương trình hồi qui tuyến tính giữa Y và các X_i bằng phương pháp hồi qui từng bước tiến hành theo thứ tự sau:

Bước 1: Tính các hệ số tương quan toàn phần r_{yi} giữa Y và các X_i ($i=1..m$) và chọn trong chúng hệ số có giá trị tuyệt đối lớn nhất. Giả sử:

$$|r_{y1}| = \max_{1 \leq i \leq m} \{|r_{yi}|\}$$

khi đó, biến X_1 là biến có tác động chính lên Y và ta xây dựng phương trình hồi qui:

$$Y^{(1)} = a_0^{(1)} + a_1^{(1)} x_1.$$

Tương ứng với phương trình trên ta tính được chuẩn sai thặng dư $s^{(1)}$ là sai số bình phương trung bình của giá trị dự báo theo phương trình hồi qui so với giá trị quan trắc:

$$s^{(k)} = \sqrt{\frac{1}{n-k-1} \sum_{i=1}^n (Y_i - Y_i^{(k)})^2}, k = 1, 2, \dots$$

trong đó k là số nhân tố có mặt trong phương trình hồi qui, n là dung lượng mẫu, $Y_i^{(k)}$ là giá trị dự báo ứng với mẫu thứ i.

Bước 2: Tính các hệ số tương quan riêng $r_{yi,1}$ ($i=2..m$) và chọn hệ số có giá trị lớn nhất trong chúng. Giả sử:

$$|r_{y2}| = \max_{2 \leq i \leq m} \{|r_{yi,1}|\}$$

Khi đó ta chọn tiếp biến X_2 và xây dựng phương trình hồi qui:

$$Y^{(2)} = a_0^{(2)} + a_1^{(2)} x_1 + a_2^{(2)} x_2 \quad (*)$$

Tương ứng với nó ta cũng tính được chuẩn sai thặng dư $s^{(2)}$. Kết thúc bước này ta có phương trình hồi qui hai biến (*) mà độ chính xác của nó được đánh giá bởi $s^{(2)}$.

Bước 3: Bây giờ ta đem so sánh giá trị chuẩn sai thặng dư $s^{(2)}$ với $s^{(1)}$:

Nếu
$$\left| \frac{s^{(2)} - s^{(1)}}{s^{(2)}} \right| < \varepsilon$$

thì biến X_2 sẽ bị bỏ qua. Quá trình tính sẽ chấm dứt. ở đây, ε là một số dương tùy ý ta đưa vào để đánh giá xem nếu tăng thêm biến cho phương trình hồi qui thì độ chính xác có tăng lên đáng kể hay không. Hay nói cách khác, khi thêm vào phương trình hồi qui một biến mới thì đóng góp thông tin của nó làm giảm sai số được bao nhiêu phần trăm; nếu mức độ giảm không vượt quá ε thì có thể bỏ qua nó.

Nếu
$$\left| \frac{s^{(2)} - s^{(1)}}{s^{(2)}} \right| \geq \varepsilon$$
 thì biến X_2 sẽ được chọn. Ta lại tính tiếp các hệ số tương

quan riêng $r_{yi.12}$ với $i=3..m$ và qui trình được lặp lại bắt đầu như bước 2. Quá trình cứ tiếp tục như vậy cho đến khi hết tất cả các biến hoặc tự kết thúc như đã trình bày.

Như vậy ở bước thứ k ta có chuẩn sai thặng dư $s^{(k)}$ tương ứng với phương trình hồi qui:

$$Y^{(k)} = a_0^{(k)} + a_1^{(k)} x_1 + \dots + a_k^{(k)} x_k$$

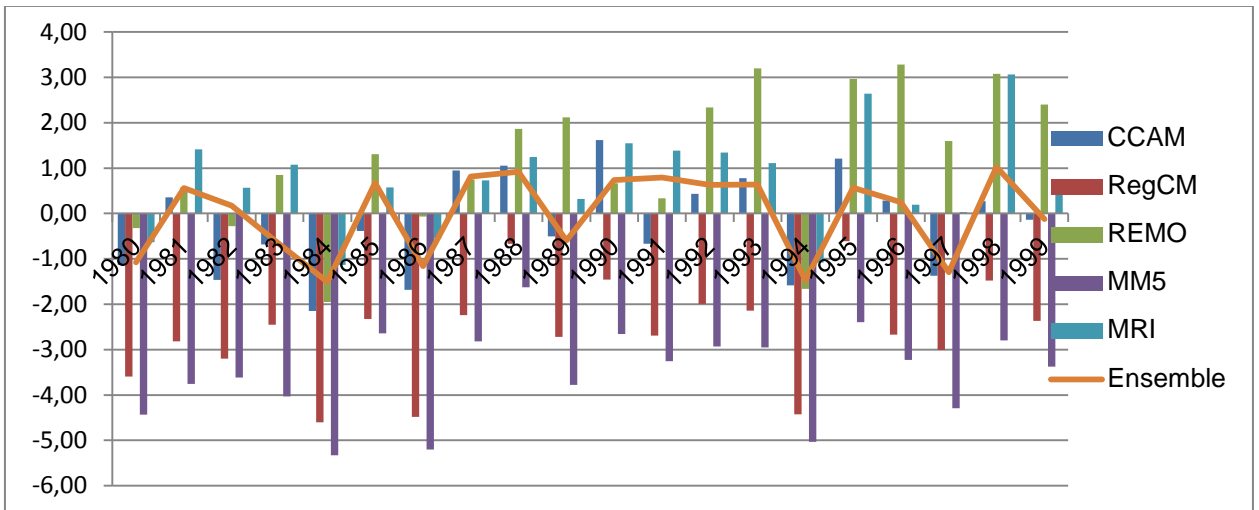
với điều kiện lựa chọn:

$$\left| \frac{s^{(k)} - s^{(k-1)}}{s^{(k)}} \right| < \varepsilon$$

3. Phân tích và so sánh kết quả tổ hợp với giá trị các mô hình.

Tại mỗi một trạm, dựa vào chuỗi số liệu quan trắc và 5 chuỗi số liệu mô hình đã được đưa về giá trị trạm ta tiến hành xây dựng phương trình hồi quy tuyến tính. Dựa vào phương trình hồi quy tuyến tính này, ta sẽ tìm ra giá trị tối ưu dựa vào thông tin của 5 mô hình. Áp dụng mẫu đối với trạm Hà Nội. (Phương pháp tính tương tự đối với các trạm khác)

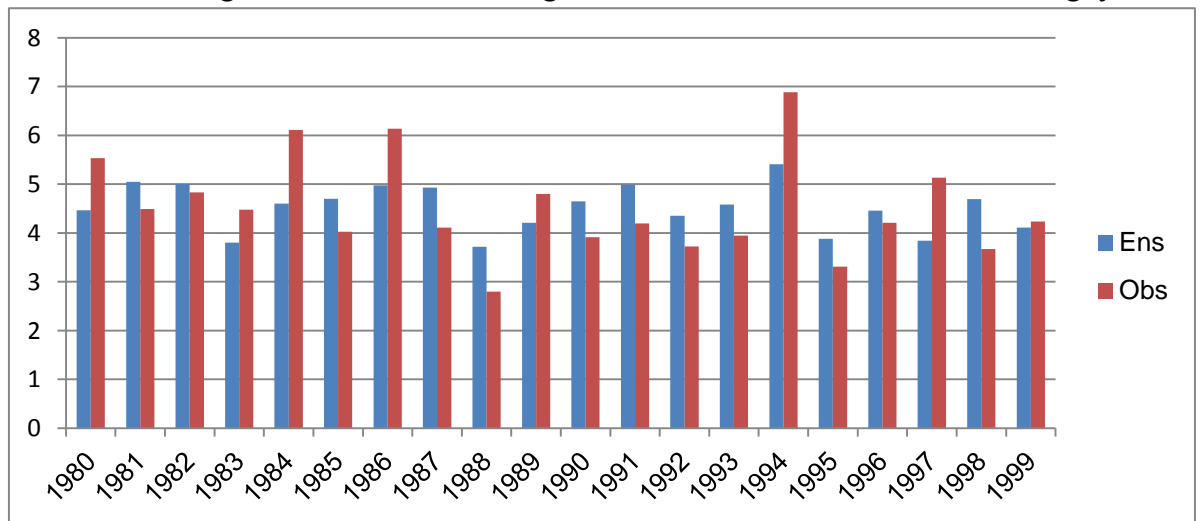
Phân tích kết quả tại trạm Hà Nội



Đồ thị : 1 Giá trị sai số lượng mưa của 5 mô hình tại trạm Hanoi

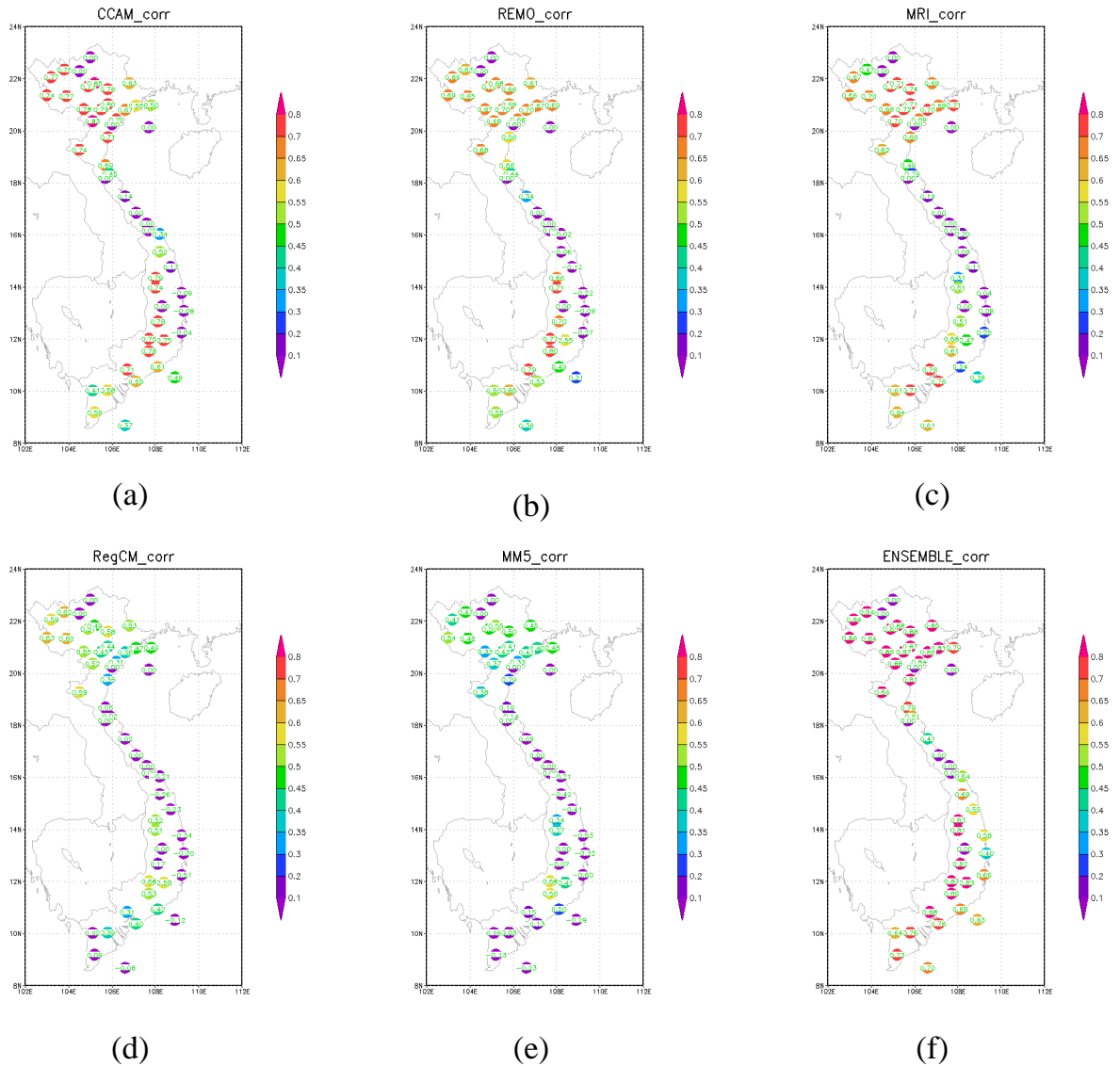
Tại trạm Hà Nội, giá trị sai số lượng mưa của 5 mô hình cho thấy một số nhận xét sau:

- RegCM và MM5 luôn cho giá trị mô phỏng thiên thấp, MM5 cho giá trị trung bình thấp hơn giá trị quan trắc khoảng 2.5mm/ngày, RegCM cho giá trị trung bình thấp hơn giá trị quan trắc khoảng 2mm/ngày. Tuy nhiên sai số của RegCM mang tính hệ thống, để biết được kỹ năng của mô hình nào mô phỏng tốt hơn, chúng ta có thể xem xét đến giá trị tương quan giữa mô hình và quan trắc.
- Ngược lại với RegCM và MM5, mô hình REMO cho giá trị mô phỏng mưa thiên cao, giai đoạn 1980-1990 cho mô phỏng khá tốt, tuy nhiên giai đoạn 1991-1999 kết quả mô phỏng cho giá trị lớn hơn trung bình quan trắc khoảng 2mm/ngày.
- CCAM và nhất là MRI mô phỏng tại trạm Hà Nội khá gần với các giá trị trạm. Sai số trung bình so với giá trị tại trạm $\sim 0.2\text{mm/ngày}$.



Đồ thị : 2 Giá trị lượng mưa so sánh với quan trắc sau khi đã tiến hành tổ hợp

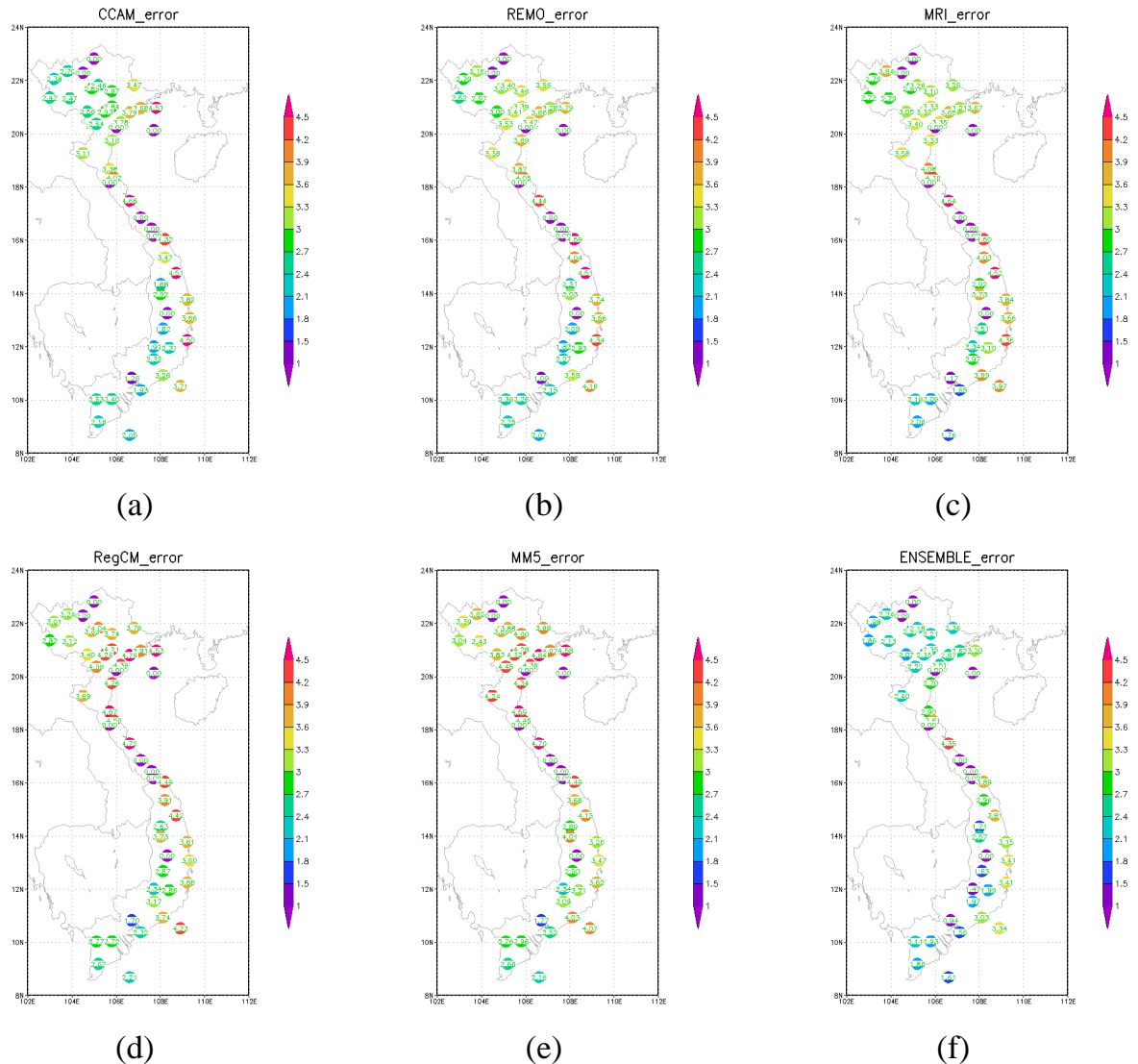
Sau khi tiến hành xây dựng phương trình hồi quy, giá trị xây dựng được đã cho kết quả tốt hơn khá nhiều so với kết quả của các mô hình. Để đánh giá được khả năng nắm bắt của mô hình ta có thể dựa vào hệ số tương quan giữa giá trị quan trắc và kết quả 5 mô hình, kết quả tổ hợp.



Hình 4. Hệ số tương quan giữa các mô hình, kết quả tổ hợp và quan trắc CCAM(a), REMO(b), MRI(c), RegCM(d), MM5(e), tổ hợp(f)

Dựa vào hệ số tương quan giữa quan trắc và 5 mô hình như trên hình 4 có thể nhận thấy giá trị tại trạm của mô hình CCAM, MRI và REMO có tương quan khá tốt với quan trắc ở khu vực phía bắc, khu vực Tây Nguyên và Nam Bộ. Điều này cho thấy, kĩ năng mô phỏng tốt của các mô hình tại khu vực đó. Mô hình MM5 và

RegCM cho tương quan thấp ở khu vực miền Trung và Tây Nguyên. Kết quả tổ hợp cho thấy tương quan khá cao tại tất cả các khu vực. Điều đó cho thấy kết quả tổ hợp đã nắm bắt khá tốt các đặc trưng của trường mưa.



Hình 5. Độ lệch chuẩn các mô hình, kết quả tổ hợp và quan trắc CCAM(a), REMO(b), MRI(c), RegCM(d), MM5(e), tổ hợp(f)

Ngược lại với hệ số tương quan, độ lệch chuẩn càng nhỏ tức là chuỗi giá trị biến thiên càng nhỏ, kết quả mô hình thì được tập trung hơn. Trên hình 5 cho thấy độ lệch chuẩn của REMO, CCAM và MRI ở khu vực Bắc Bộ và Nam Bộ khá nhỏ, và khá lớn ở khu vực miền Trung. Độ lệch chuẩn tại các trạm của mô hình RegCM và MM5 tại hầu hết các trạm đều khá lớn trừ khu vực Nam Bộ và Tây Bắc Bộ. Kết

qua tổ hợp cho thấy giá trị độ lệch chuẩn khá nhỏ, cho thấy mức độ biến thiên của sai số là khá nhỏ, kết quả đáng tin cậy hơn so với các trường hợp còn lại.

4. Kết luận

Dựa trên các kết quả phân tích có thể thấy rằng các mô phỏng đối với trường mưa vẫn còn tồn tại nhiều hạn chế, trong cả 5 mô hình chưa có mô hình nào cho kết quả mô phỏng tốt cả về phương diện không gian và định lượng. Nhìn tổng quan chung thì CCAM và MRI là cho mô phỏng tốt hơn cả, MM5 thì mô phỏng định lượng rất thấp so với thực tế. REMO tuy cho mô phỏng cao hơn thực tế nhưng độ tin cậy khá cao, có thể dùng phương pháp hiệu chỉnh đối với mô hình này sẽ cho kết quả tốt hơn.

Một phương pháp tổ hợp kết quả dựa trên phương trình hồi quy tuyến tính cho kết quả tốt hơn hẳn so với các mô hình còn lại cả về phương diện không gian cũng như định lượng.