

SỬ DỤNG HÀM PHÂN LỚP ĐỂ DỰ BÁO ĐỒNG NHIỆT THỜI HẠN 6-12 GIỜ CHO VÙNG ĐỒNG BẰNG BẮC BỘ TRONG NỬA ĐẦU MÙA HÈ

NCS. Nguyễn Viết Lành
Trường Cán bộ KTTV, Hà Nội

Đặt vấn đề: Các phương pháp thống kê toán học được ứng dụng rộng rãi trong nghiên cứu khí tượng, khí hậu nói chung và xây dựng phương trình dự báo thời tiết nói riêng. Đối với bài toán dự báo lựa chọn hay dự báo pha hiện tượng thời tiết, một trong những phương pháp thống kê được ứng dụng rộng rãi là phương pháp hàm phân lớp. Phương pháp này đã được Fisher [6] sử dụng lần đầu tiên vào năm 1938. Trần Tân Tiến [3] cũng đã chỉ ra toàn bộ thuật toán để xác định các hệ số của hàm phân lớp. Ở Việt Nam, hàm phân lớp cũng đã được một số tác giả ứng dụng trong dự báo pha hiện tượng thời tiết.

1. Cơ sở số liệu và phương pháp nghiên cứu

a. Cơ sở số liệu

Số liệu ban đầu là một trong những nền tảng của các phương pháp thống kê. Số liệu ban đầu bao gồm các yếu tố cấu thành độ dài của chuỗi. Trong điều kiện nước ta, khi chưa có được những tập số liệu khảo sát đầy đủ có tính chất chuyên sâu, việc khai thác số liệu cho mục đích nghiên cứu bài toán dự báo đồng nhiệt phải xuất phát từ tình hình thực tế của các nguồn số liệu và chất lượng của số liệu. Nguyên tắc chung của phương thức khai thác số liệu là khai thác triệt để các nguồn số liệu sẵn có và có khả năng đáp ứng được những yêu cầu của bài toán đặt ra.

Để thực hiện bài viết này, chúng tôi đã tiến hành khai thác số liệu của các tháng V, VI và VII trong 13 năm (1982-1994) từ các nguồn như đã lấy để tính chỉ số bất ổn định của khí quyển [2]. Phương pháp phân tích hình thể trên bản đồ synop được dẫn ra trong [1], [2].

b. Phương pháp nghiên cứu

Lý thuyết của hàm phân lớp có thể được trình bày một cách ngắn gọn như sau:

Giả sử có hai pha thời tiết là ϕ_0 và ϕ_1 cùng với x_1, x_2, \dots, x_m là các nhân tố dự báo. Các nhân tố dự báo được chia thành hai lớp tương ứng với hai pha thời tiết là ϕ_1 và ϕ_2 . Như vậy, từ tập số liệu lưu trữ, ta biết được mỗi vectơ nhân tố dự báo $X = \{x_1, x_2, \dots, x_m\}$ thuộc lớp nào của một trong hai pha thời tiết là ϕ_0 và ϕ_1 . Nhiệm vụ đặt ra là phải tìm cách giải để khi có bất kỳ một vectơ nhân tố dự báo nào nằm trong bộ số liệu lưu trữ ta cũng chỉ ra được thời tiết thuộc pha ϕ_0 hay ϕ_1 .

Mỗi một vectơ nhân tố ảnh hưởng là một điểm trong không gian m chiều. Những điểm ứng với những vectơ nhân tố dự báo thuộc một lớp sẽ tạo nên một miền

nào đó trong không gian m chiều. Giữa hai miền này có thể xảy ra các trường hợp sau:

- Hai miền này riêng biệt nhau, các vectơ nhân tố dự báo chia thành hai lớp. Đây là trường hợp phân lớp lý tưởng. Trong trường hợp này sẽ có một mặt cong phân chia hai lớp.
- Hai miền đan xen vào nhau mà ta không thể phân chia các vectơ nhân tố dự báo thành các lớp được. Như vậy, bài toán không thể giải được bằng phương pháp này trên cơ sở các nhân tố đã chọn.
- Hai miền có những phần chung nào đó. Trường hợp này đã đưa đến bài toán dự báo có điều kiện.

Quy tắc giải cho bài toán dự báo lựa chọn có dạng:

$$I = \text{sign} (P (X) - P_0) \quad (1)$$

Trong đó: P_0 là một hằng số.

$P (X)$ là hàm phụ thuộc vào vectơ nhân tố dự báo $X = \{ x_1, x_2, \dots, x_m \}$.

Nếu dấu của biểu thức là dương thì ta dự báo thời tiết ứng với pha ϕ_1 , còn dấu của biểu thức âm thì ta dự báo thời tiết ứng với pha ϕ_0 .

Giả sử, hàm mật độ phân bố của X thoả mãn phân bố chuẩn trong từng lớp, khi đó, hàm phân bố trong lớp 1:

$$f_1(x_1, x_2, \dots, x_m) = (2\pi)^{-m/2} |R|^{-1/2} \exp \left[-\frac{1}{2|R|} \sum_{i,j} R^{-1}_{ij} (x_i - \mu_{i1})(x_j - \mu_{j1}) \right]$$

và hàm phân bố trong lớp 0:

$$f_0(x_1, x_2, \dots, x_m) = (2\pi)^{-m/2} |R|^{-1/2} \exp \left[-\frac{1}{2|R|} \sum_{i,j} R^{-1}_{ij} (x_i - \mu_{i0})(x_j - \mu_{j0}) \right]$$

Trong đó: R_{ij} là các phần tử của ma trận tương quan.

$$R_{ij} = M [(x_i - \mu_i)(x_j - \mu_j)]$$

$|R|$ là định thức của ma trận tương quan.

Những phần tử ở lớp 1 thuộc hàm phân bố f_1 còn những phần tử ở lớp 0 thuộc hàm phân bố f_0 . Để đơn giản, ta giả thiết R_{ij} của lớp 1 bằng R_{ij} của lớp 2. Lấy logarit của tỉ số hai hàm này, ta có:

$$\begin{aligned} \ln \frac{f_1(X)}{f_0(X)} &= \frac{1}{2} \sum_{i,j=1}^m R^{-1}_{ij} [(x_i - \mu_{i1})(x_j - \mu_{j1}) - (x_i - \mu_{i0})(x_j - \mu_{j0})] \\ &= -\frac{1}{2} \sum_{i,j=1}^m R^{-1}_{ij} (\mu_{i0} - \mu_{i1}) x_i + \text{const} \end{aligned} \quad (2)$$

Ta ký hiệu:

$$C_i = -\frac{1}{2} \sum_{j=1}^m R_{ij} (\mu_{0j} - \mu_{1j})$$

thì biểu thức trên viết được về dạng của hàm phân lớp:

$$I = \ln \frac{f_1(X)}{f_0(X)} - \text{const} = \sum_{i=1}^m C_i x_i$$

Hay: $I = C_0 + C_1 x_1 + C_2 x_2 + \dots + C_m x_m$ (3)

Khi có vectơ nhân tố $\{X\}$ ta tính I theo (3). Nếu $I > 0$ pha thời tiết là pha 1, ngược lại, $I < 0$ pha thời tiết là pha 0.

Thực tế tính toán hàm phân lớp người ta chia số liệu lưu trữ làm 2 lớp ứng với 2 pha thời tiết. Tập mẫu trong mỗi lớp là N_1 và N_2 . Để thuận tiện, ta gọi lớp thứ nhất là A, lớp thứ hai là B, khi đó:

$$d_i = x_i(A) - x_i(B)$$

Nếu $d_i = 0$ thì nhân tố x_i bị loại ra khỏi tập nhân tố dự báo.

Ta có hệ phương trình:

$$R_{11} C_1 + R_{12} C_2 + \dots + R_{1m} C_m = d_1$$

$$R_{21} C_1 + R_{22} C_2 + \dots + R_{2m} C_m = d_2$$

$$R_{m1} C_1 + R_{m2} C_2 + \dots + R_{mm} C_m = d_m$$

Giải hệ phương trình trên ta tìm được C_i .

2. xây dựng phương trình dự báo

a. Các nhân tố ban đầu

Như đã thấy, quá trình hình thành và phát triển của dông nhiệt, ngoài nhân tố hình thể synop, liên quan với:

-Độ bất ổn định thẳng đứng của khí quyển,

-Nhóm nhiệt độ và độ ẩm,

-Nhóm hướng gió và tốc độ gió,

-Nhóm khí áp và biến thiên khí áp,

-Điều kiện địa phương và vị trí địa lý.

Mối quan hệ giữa các nhân tố này với nhau cũng như giữa chúng với quá trình hình thành và phát triển dông nhiệt nói chung rất phức tạp và khá chặt chẽ. Theo Gruza [3], trong thống kê dự báo khí tượng, việc lựa chọn nhân tố ban đầu dự báo thường không có một nguyên tắc chung. Tùy theo từng hiện tượng, cơ chế vật lý của quá trình hình thành mà người nghiên cứu phải lựa chọn những nhân tố nào. Trên quan điểm đó, việc lựa chọn các nhân tố ban đầu cho các mô hình dự báo khả năng xuất hiện dông nhiệt được thực hiện theo nguyên tắc:

a) Căn cứ vào cơ chế vật lý của quá trình hình thành dông nhiệt,

b) Căn cứ vào nội dung và chất lượng của số liệu ban đầu.

Như kết quả đã dẫn [1], đông nhiệt hoạt động chủ yếu trong khoảng thời gian từ 13 đến 23 giờ với tần suất xuất hiện lớn nhất tập trung vào lúc 19 giờ. Cho nên, yếu tố dự báo (đông nhiệt) được lấy tại trạm Láng trong khoảng thời gian này.

Do bản chất của hiện tượng, độ bất ổn định của khí quyển có thể được xem là một trong những nhân tố có tính quyết định đối với quá trình hình thành và phát triển đông nhiệt, đặc trưng cho nhóm này gồm các chỉ số độ bất ổn định của khí quyển [2]. Khí áp và biến thiên khí áp phản ánh áp thấp nóng đang trong giai đoạn nào của quá trình phát triển. Gió trên các mực 850, 700 và 500mb cũng có vai trò quan trọng đối với quá trình hoạt động của hiện tượng này [1]. Bên cạnh đó, sự biến thiên theo thời gian của nhiệt độ và độ ẩm trước khi đông nhiệt xuất hiện cũng cần phải tính đến.

Trên cơ sở đó, các nhân tố ban đầu được chọn là:

1. Chỉ số độ bất ổn định của khí quyển: SI, LI, Cox_1, Cox_2, SWEAT, TI, WEMEX_L (WEMEX lớp dưới), WEMEX_U (WEMEX lớp trên) và KI,
2. Lượng mây tích lúc 10 giờ ở Hà Nội (MHN10) và Hoà Bình (MHB10),
3. Tốc độ gió lúc 7 giờ trên các mực: 850mb (f85), 700mb (f70) và mực 500mb (f50),
4. Hiệu nhiệt độ không khí 7 giờ giữa mặt đất với mực 850mb (T7_85),
5. Hiệu nhiệt độ không khí 7 giờ giữa mặt đất với mực 700mb (T7_70),
6. Hiệu nhiệt độ không khí 7 giờ giữa mặt đất với mực 500mb (T7_50),
7. Hiệu nhiệt độ không khí giữa mặt đất 10 giờ với mực 850mb 7 giờ (T10_85),
8. Hiệu nhiệt độ không khí giữa mặt đất 10 giờ với mực 700mb 7 giờ (T10_70),
9. Hiệu nhiệt độ không khí giữa mặt đất 10 giờ với mực 500mb 7 giờ (T10_50),
10. Trung bình nhiệt độ mặt đất 7 và 10 giờ trừ đi nhiệt độ mực 850mb 7 giờ (T710_85),
11. Trung bình nhiệt độ mặt đất 7 và 10 giờ trừ đi nhiệt độ mực 700mb 7 giờ (T710_70),
12. Trung bình nhiệt độ mặt đất 7 và 10 giờ trừ đi nhiệt độ mực 500mb 7 giờ (T710_50),
13. Độ hụt điểm sương mặt đất 7 giờ (TTd07) và 10 giờ (TTd010),
14. Độ hụt điểm sương 7 giờ mực 850mb (TTd85) và mực 700mb (TTd70),
15. Trung bình độ hụt điểm sương giữa hai mực: mặt đất và 850 (TBTTd2),
16. Trung bình độ hụt điểm sương giữa ba mực: mặt đất, 850 và 700mb (TBTTd3),
17. Hiệu nhiệt độ điểm sương mặt đất lúc 10 giờ và 7 giờ (Td10_7),
18. Biến thiên khí áp 24 giờ tại mặt đất lúc 7 giờ (ap7) và lúc 10 giờ (ap10).

b. Xây dựng phương trình dự báo

Khác với phương pháp hàm hồi quy từng bước, ở đây, khi xây dựng các phương trình dự báo khả năng xuất hiện đông nhiệt theo phương pháp hàm phân lớp, số biến dự tuyển cũng đồng thời là số nhân tố dự tuyển. Vì vậy, nếu trong phương pháp hồi quy từng bước, tùy thuộc vào bậc đa thức của hàm xấp xỉ và số nhân tố dự báo mà số biến dự tuyển có thể khác nhau giữa trường hợp này với trường hợp khác, còn ở đây, số biến dự tuyển luôn không đổi và bằng 32 biến. Các biến này được chia thành 2 lớp có và không có đông nhiệt. Tuy nhiên, việc xây dựng các phương trình dự báo trong đó kết hợp cả việc lựa chọn các biến dự tuyển đã được thực hiện theo quy trình đã trình bày trong mục 1, và sơ đồ tính hình 1. Điều đó cho phép loại bỏ được

những nhân tố mang lượng thông tin không đáng kể và giữ lại những nhân tố cho hiệu quả dự báo cao.

Kết quả tính toán xây dựng phương trình dự báo đồng nhiệt cho vùng đồng bằng Bắc Bộ được biểu diễn:

$$y_{BB} = 0,97328 - 0,02702 * TBTTd3 + 0,00982 * Td10_7 - 0,02576 * f85 + 0,00796 * TTd85 \quad (4)$$

Từ kết quả này ta nhận thấy rằng, số biến tham gia dự tuyến là 32 nhưng số biến tham gia vào các phương trình dự báo là 4. Trong phương trình (4), các biến TBTTd3, Td10_7, f85, TTd85 đều tham gia vào phương trình dự báo.

Như vậy, độ ẩm trung bình của cả 3 mực khí quyển tầng thấp, sự tăng trở lượng ẩm tầng thấp trong thời gian từ 7 đến 10 giờ, tốc độ gió và độ hụt điểm sương trên mực 850mb được xem là những nhân tố có vai trò quyết định đối với quá trình hình thành đồng nhiệt trên vùng đồng bằng Bắc Bộ.

c. Đánh giá hiệu quả dự báo

Khi xây dựng phương trình dự báo pha thời tiết, việc đánh giá được dựa trên cơ sở đánh giá hiệu quả của phương pháp dự báo lựa chọn. Đối với bài toán dự báo đồng nhiệt, việc đánh giá sẽ căn cứ vào mức độ đúng hay sai của bản tin dự báo khi sử dụng phương trình.

Ký hiệu F_1 và F_0 là các pha thời tiết tương ứng với hiện tượng đồng nhiệt có xảy ra và không xảy ra trong thực tế, P_{21} và P_0 là dự báo các pha độ. Khi đó, nếu tiến hành dự báo N lần, ta sẽ nhận được kết quả sau:

Có n_{00} là số lần dự báo không đồng và thực tế cũng không đồng.

n_{11} là số lần dự báo có đồng và thực tế cũng có đồng.

n_{01} là số lần dự báo không đồng nhưng thực tế lại có đồng.

n_{10} là số lần dự báo có đồng nhưng thực tế lại không đồng.

N_{00} là số lần dự báo không đồng.

N_{10} là số lần dự báo có đồng.

N_0 là số lần thực tế không đồng.

N_1 là số lần thực tế có đồng.

Hiệu quả dự báo hay độ chính xác toàn phần sẽ là:

$$U = \frac{N_{00} + n_{11}}{N} \quad (5)$$

Ngoài ra, Bagrov N. A. [6] đã đưa ra một tiêu chuẩn nữa gọi là tiêu chuẩn độ tin cậy H . Độ tin cậy H được xác định như sau:

$$H = \frac{U - U_0}{1 - U_0} \quad (6)$$

Trong đó, U_0 được gọi là độ chính xác toàn phần của dự báo ngẫu nhiên và được xác định theo công thức sau:

$$U_0 = \frac{1}{N} \left(\frac{N_0}{N} N_{00} + \frac{N_1}{N} N_{10} \right) \quad (7)$$

Trên chuỗi số liệu chúng tôi sử dụng để tính ở đây, thì:

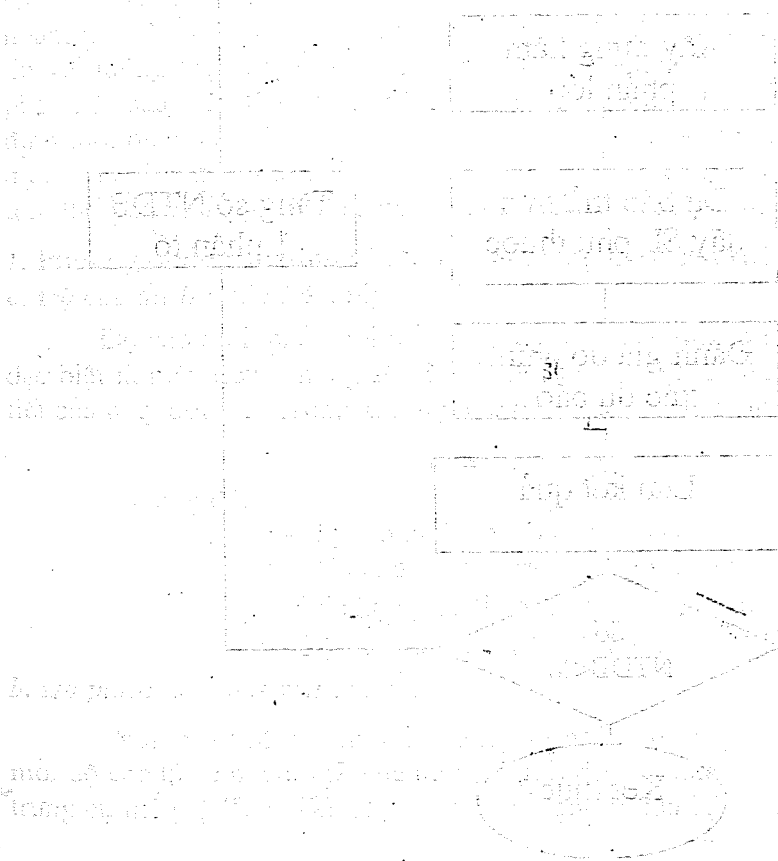
$$n_{00} = 50; n_{11} = 144; n_{01} = 12; n_{10} = 32;$$

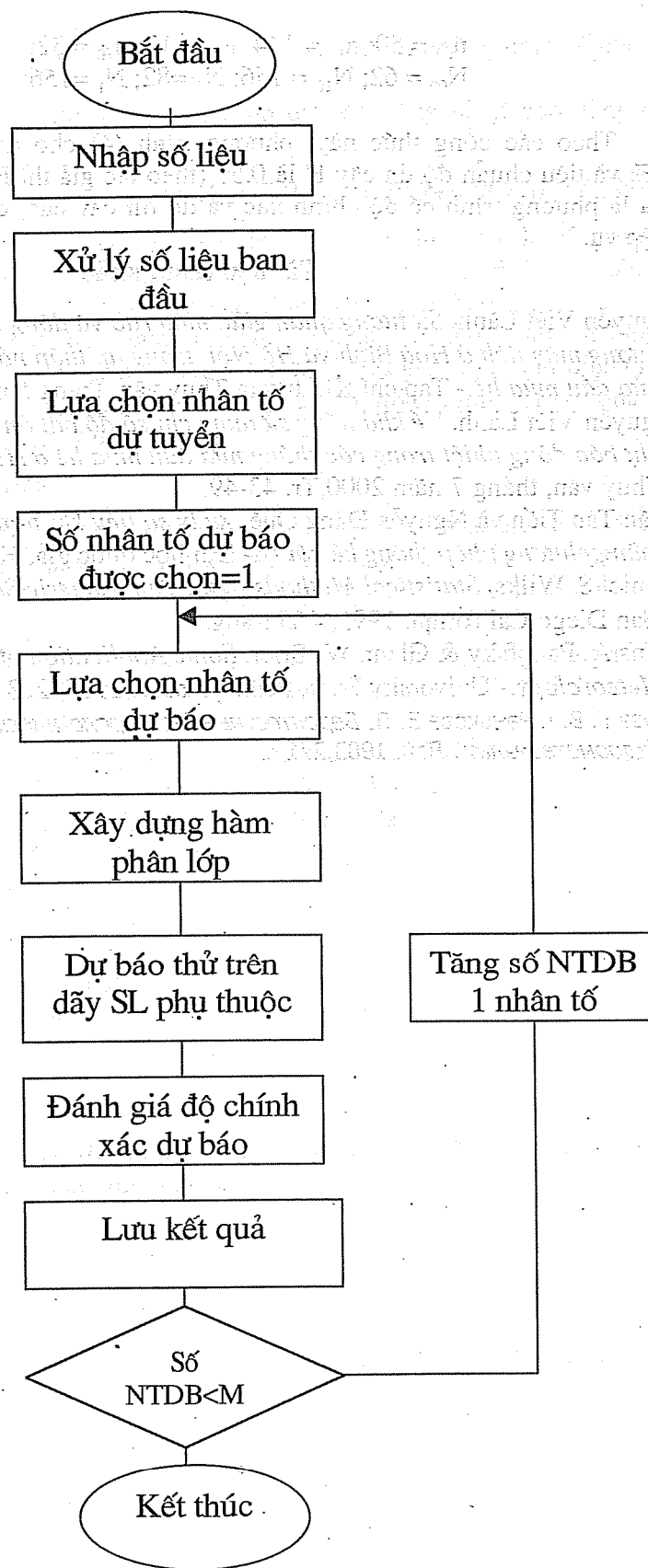
$$N_{00} = 62; N_{10} = 156; N_0 = 82; N_1 = 156;$$

Theo các công thức này, phương trình (4) cho độ chính xác toàn phần là 81,5% và tiêu chuẩn độ tin cậy H là 0,57 (theo tác giả thì $H > 0,2$ là đạt yêu cầu). Có nghĩa là phương trình có độ chính xác và độ tin cậy cao, có thể sử dụng được trong nghiệp vụ.

Tài liệu tham khảo

1. Nguyễn Việt Lành. Sự tương quan giữa mưa rào và đông nhiệt ở Hà Nội với lượng mây tích ở Hoà Bình và Hà Nội trong áp thấp nóng phía tây vào thời kỳ nửa đầu mùa hè.- Tạp chí Khí tượng Thủy văn, tháng 5 năm 2000, Tr. 43-46.
2. Nguyễn Việt Lành. Về khả năng sử dụng chỉ số độ bất ổn định của khí quyển để dự báo đông nhiệt trong các tháng nửa đầu mùa hè ở Hà Nội.- Tạp chí Khí tượng Thủy văn, tháng 7 năm 2000, Tr. 43-49.
3. Trần Tân Tiến và Nguyễn Đăng Quế. Xử lý số liệu khí tượng và dự báo thời tiết bằng phương pháp thống kê vật lý.- Đại học quốc gia, Hà Nội, 1997.
4. Daniel S. Wilks. *Statistical Methods in the Atmospheric Sciences*.- Academic Press San Diego California, 1995, 465 trang.
5. Hans A. Panofsky & Glenn W. Brier. *Some Application of Statistics to Meteorology*.- University Park, Pennsylvania, 1965, 223 P.
6. Груза Г. В. и Ранькова Э. Я. *Вероятностные Метеорологические Прогнозы* Гидрометеиздат. Лен. 1983, 271 с.





Hình 1: Sơ đồ xử lý và tính toán hàm phân lớp