

SỬ DỤNG HÀM HỒI QUY TỪNG BƯỚC ĐỂ DỰ BÁO ĐÔNG NHIỆT THỜI HẠN 6-12 GIỜ CHO KHU VỰC HÀ NỘI TRONG CÁC THÁNG NỬA ĐẦU MÙA HÈ

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

1. Đặt vấn đề

Việc nghiên cứu, dự báo đông nói chung, đông nhiệt nói riêng, ở nước ta hiện nay là một trong những vấn đề rất cấp bách. Nhu cầu thực tế của một số ngành kinh tế quốc dân đòi hỏi những bản tin dự báo đông có độ chính xác cao mới có thể tránh được những thiệt hại về người và của cũng như kế hoạch hoá được trong sản xuất, đặc biệt là trong các ngành giao thông đường không, đường thủy.

Nhìn chung, trong công tác nghiệp vụ dự báo thời tiết ở nước ta cho đến nay, phương pháp synop vẫn là phương pháp chủ đạo, chưa có được một phương pháp khác thay thế. Do đặc thù của phương pháp này mà việc dự báo đông nhiệt cho một địa phương cụ thể gặp phải những khó khăn đáng kể. Như đã biết, đông nhiệt là một hiện tượng khí tượng có quy mô nhỏ, việc dự báo nó bằng phương pháp synop khó có thể tính hết được những nhân tố ảnh hưởng đến các quá trình hình thành và phát triển của chúng. Bởi vậy, để dự báo hiện tượng này được chính xác hơn, bên cạnh phương pháp synop cần phải có những phương pháp khách quan khác bổ trợ.

Một trong những phương pháp bổ trợ được sử dụng rộng rãi là phương pháp thống kê vật lý. Phương pháp thống kê vật lý dựa trên cơ sở xử lý tính toán tập số liệu quan trắc, nghiên cứu mối quan hệ thống kê giữa hiện tượng cần dự báo với tập các nhân tố dự báo. Từ đó rút ra những quy luật thống kê mô tả quá trình hình thành, phát triển và tan rã của đông nhiệt, xây dựng các công thức dự báo hiện tượng.

2. 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 đông 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

Phương pháp hồi quy nói chung và hồi quy từng bước nói riêng được rất nhiều nhà khí tượng trong và ngoài nước ứng dụng để xây dựng các phương trình dự báo hiện tượng. Bài toán này có thể được trình bày như sau:

Giả sử có K pha thời tiết: $\phi_1, \phi_2, \dots, \phi_K$, xác suất để pha thứ i xảy ra là $P_i = P(\phi_i)$, ($i = 1, \dots, K$). Gọi X là vector các nhân tố dự báo, $X = \{x_1, x_2, \dots, x_m\}$, yếu tố dự báo phụ thuộc vào m nhân tố. Ký hiệu P^* là vector ước lượng của P ($P = \{P_1, P_2, \dots, P_K\}$). Khi đó ta có:

$$P^* = A \cdot X + A_0 \quad (2.1)$$

Trong đó, ma trận K hàng m cột A chứa các hệ số hồi quy và vector các hệ số tự do m chiều A_0 được xác định bằng phương pháp bình phương tối thiểu. Điều kiện ràng buộc khi xác định các hệ số trong A và A_0 là, tại mọi thời điểm, chúng cần phải thoả mãn hệ thức:

$$\sum_{i=1}^K P_i = 1 \quad (2.2)$$

Vector nhân tố dự báo X nói chung có thể là những đại lượng nhị phân, rời rạc hoặc liên tục [4].

Trên cơ sở những nguyên tắc này, để ứng dụng cho bài toán dự báo đông nhiệt, ta có bài toán dự báo lựa chọn giữa 2 pha thời tiết là có và không có đông nhiệt xuất hiện.

Gọi ϕ là sự kiện hiện tượng đông nhiệt xuất hiện và P là xác suất xuất hiện sự kiện ϕ . Khi đó P sẽ nhận giá trị bằng 1 nếu đông nhiệt xuất hiện và P nhận giá trị bằng 0 nếu đông nhiệt không xuất hiện. Như vậy, xác suất để sự kiện ϕ xuất hiện sẽ là:

$$P(\phi) = P(P = 1) = M[P] \quad (2.3)$$

Trong đó M là toán tử kỳ vọng. Như vậy, để xác định được xác suất xuất hiện hiện tượng ϕ , ta cần phải xác định kỳ vọng toán của P.

Giả thiết rằng, xác suất xuất hiện ϕ phụ thuộc vào m nhân tố dự báo $x_1, x_2, \dots, x_m$. Khi đó, thay cho $P(\phi)$, ta có xác suất có điều kiện $P(\phi/X)$.

$$P(\phi/X) = M[P/X] \quad (2.4)$$

Trong đó: X là vector các nhân tố dự báo $X = \{x_1, x_2, \dots, x_m\}$.

$M[P/X]$ là kỳ vọng có điều kiện của biến ngẫu nhiên P.

Như vậy, để dự báo xác suất xuất hiện đông nhiệt theo vector các nhân tố dự báo $X = \{x_1, x_2, \dots, x_m\}$ đã chọn, chúng ta cần phải xác định được kỳ vọng có điều kiện $M[P/X]$. Thông thường, $M[P/X]$ được xấp xỉ bằng một tổ hợp tuyến tính các nhân tố dự báo:

$$P = M[P/X] = a_0 + a_1x_1 + a_2x_2 + \dots + a_mx_m \quad (2.5)$$

Trong đó, các hệ số a_i ($i = 0, 1, \dots, m$) là các hệ số hồi quy được xác định bằng phương pháp bình phương tối thiểu.

Ta nhận thấy rằng, việc xấp xỉ P bằng một tổ hợp tuyến tính của các biến như trên có thể chưa phản ánh hết được mối quan hệ phụ thuộc giữa P với các nhân tố dự báo x_i . Bởi vì mối quan hệ giữa chúng rất có thể không hoàn toàn là tuyến tính. Do đó, thay cho (2.5), ta cần biểu diễn P dưới dạng một hàm nào đó của các nhân tố này:

$$P = M[P/X] = f(x_1, x_2, \dots, x_m) \quad (2.6)$$

Như vậy, để xác định P, ta phải xây dựng được hàm $f(x_i)$, tức là phải xác định được mối quan hệ thống kê giữa P với các nhân tố dự báo x_i :

$$P_i = f_i(x_i), (i=1, 2, \dots, m) \quad (2.7)$$

Dạng của $f_i(x_i)$ được chọn trên nguyên tắc bình phương tối thiểu. Từ đồ thị biểu diễn mối quan hệ (2.7), một lớp hàm nào đó sẽ được chọn làm hàm xấp xỉ. Sau

phép hồi quy, lớp hàm cho giá trị chuẩn sai bé nhất sẽ là lớp hàm được chọn làm dạng hàm $f_i(x_i)$.

Nếu hàm $f_i(x_i)$ được chọn xấp xỉ trên lớp hàm đa thức bậc k , khi đó ta có:

$$P_i = f_i(x_i) = a_{i0} + a_{i1}x_i + a_{i2}x_i^2 + \dots + a_{ik}x_i^k$$

Kết quả của bước này cho ta một tập m_1 hàm ($m_1 \leq m$) khác nhau dạng (2.7). Hàm $f(x_i)$ sẽ được xác định trên cơ sở các hàm này bằng cách tác dụng một toán tử L nào đó:

$$f(x_i) = L\{f_1, f_2, \dots, f_{m_1}\} \quad (2.8)$$

Toán tử L có thể là tuyến tính hoặc phi tuyến. Trong trường hợp L là tuyến tính ta có:

$$\begin{aligned} f(x_i) &= \sum_{i=1}^{m_1} \alpha_i f_i \\ &= \alpha_1 (a_{10} + a_{11}x_1 + a_{12}x_1^2 + \dots + a_{1k_1}x_1^{k_1}) \\ &\quad + \alpha_2 (a_{20} + a_{21}x_2 + a_{22}x_2^2 + \dots + a_{2k_2}x_2^{k_2}) \\ &\quad \dots \dots \dots \\ &\quad + \alpha_{m_1} (a_{m_10} + a_{m_11}x_{m_1} + a_{m_12}x_{m_1}^2 + \dots + a_{m_1k_{m_1}}x_{m_1}^{k_{m_1}}) \end{aligned} \quad (2.9)$$

Trong đó α_i ($i = 1, 2, \dots, m_1$) là các hệ số. Sau khi tuyến tính hoá các thành phần phi tuyến trong (2.9) ta nhận được:

$$P = f(x_i) = \sum_{i=1}^M A_i Z_i + A_0 \quad (2.10)$$

Ở đây $M \geq m_1$ là số các hạng tử, Z_i ($i = 1, 2, \dots, M$) là các hạng tử đã được tuyến tính hoá ($Z_1 = x_1, Z_2 = x_1^2, \dots$), A_i ($i = 1, 2, \dots, M$) là các hệ số hồi quy. Nếu dạng các đường cong (2.7) càng phức tạp thì M lớn hơn m_1 rất nhiều và do đó dạng thức (2.10) sẽ rất cồng kềnh. Có thể nhận thấy rằng trong (2.10) sẽ chứa một số hạng tử thừa nào đó không mang ý nghĩa đóng góp thông tin cho phương trình dự báo, bởi vì giữa các nhân tố dự báo có thể tồn tại mối quan hệ lẫn nhau nào đó. Do đó cần phải tiến hành loại bỏ những hạng tử không có ý nghĩa và chỉ giữ lại những hạng tử cần thiết. Quá trình loại bỏ đó được gọi là lọc nhân tố.

Có nhiều phương pháp lọc nhân tố. Phương pháp hồi quy từng bước [3] là một trong những phương pháp thường được sử dụng. Ưu điểm của phương pháp này là lựa chọn được tập những nhân tố tương quan với nhau kém nhưng lại quan hệ chặt chẽ với yếu tố dự báo. Phương pháp này được tiến hành như sau:

Tính các hệ số tương quan r_{yi} giữa yếu tố y với các nhân tố dự báo x_i ($i = 1, 2, \dots, m$) sau đó chọn trong chúng hệ số tương quan nào có giá trị tuyệt đối lớn nhất. Giả sử:

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

Khi đó x_1 là nhân tố có tác động chính lên y và ta xây dựng phương trình hồi quy:

$$y^{(1)} = a_0^{(1)} + a_1^{(1)}x_1 \quad (2.11)$$

Tương ứng với (2.11) ta tính được chuẩn sai thặng dư $s^{(1)}$, tiếp theo tính hệ số tương quan $r_{yi,1}$ ($i = 2, 3, \dots, m$) và cũng chọn hệ số có giá trị lớn nhất trong chúng, giả sử:

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

Khi đó ta chọn tiếp biến x_2 và xây dựng phương trình hồi quy:

$$y^{(2)} = a_0^{(2)} + a_1^{(2)} x_1 + a_2^{(2)} x_2 \quad (2.12)$$

Tương ứng với nó ta cũng tính được chuẩn sai thặng dư $s^{(2)}$. Đến đây ta có phương trình hồi quy hai biến (2.12) mà độ chính xác của nó được đánh giá bởi $s^{(2)}$.

Nếu:

$$\left| \frac{s^{(2)} - s^{(1)}}{s^{(2)}} \right| < \varepsilon \quad (2.13)$$

thì biến x_2 sẽ bị bỏ qua và một biến khác trong số các biến còn lại sẽ được lựa chọn để xây dựng phương trình (2.12). Ở đây, ε là một số dương tùy ý ta đưa vào để đánh giá xem nếu khi ta tăng thêm biến cho phương trình thì độ chính xác của nó có tăng lên đáng kể hay không.

Tuy nhiên, $|r_{y2.1}|$ có giá trị lớn nhất trong số các $|r_{yi.1}|$ do đó nhân tố sẽ được đưa vào tiếp theo thay thế x_2 sẽ là nhân tố thoả mãn điều kiện:

$$\max_{3 \leq i \leq m} \{ |r_{yi.1}| \}$$

Nếu tất cả các nhân tố còn lại đều thoả mãn (2.13) thì quá trình hồi quy sẽ kết thúc và phương trình hồi quy (2.11) là kết quả cuối cùng.

Nếu:

$$\left| \frac{s^{(2)} - s^{(1)}}{s^{(2)}} \right| \geq \varepsilon \quad (2.14)$$

thì nhân tố x_2 sẽ được chọn. Khi đó ta lại tính tiếp các hệ số tương quan $r_{yi.12}$ ($i = 3, 4, \dots, m$) và quy trình được lặp lại.

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

$$y^{(k)} = a_0^{(k)} + a_1^{(k)} x_1 + \dots + a_k^{(k)} x_k \quad (2.15)$$

Và điều kiện lựa chọn:

$$\left| \frac{s^{(k)} - s^{(k-1)}}{s^{(k)}} \right| < \varepsilon \quad (2.16)$$

Cùng với việc lọc những nhân tố thừa trong (2.10) còn có một vấn đề nữa là xác định được ngưỡng dự báo P_0 để: khi $P > P_0$ sẽ dự báo có dông nhiệt xuất hiện và ngược lại.

Như vậy, từ (2.10) ta có M biến tham gia dự tuyển. Cần phải tìm một tổ hợp k trong số M biến đó sao cho ứng với giá trị ngưỡng dự báo P_0 ta tìm được phương trình dự báo:

$$P = \sum_{i=1}^k a_i Z_i + a_0 \quad (2.17)$$

cho hiệu quả dự báo cao nhất. Muốn vậy, ta lần lượt thực hiện các bước hồi quy theo sơ đồ trên hình 1. Tức là cần phải tính $C_M^1 + C_M^2 + \dots + C_M^M$ hệ thức (2.17) và các giá trị ngưỡng P_0 tương ứng. Kết thúc một lần tính theo sơ đồ này, các giá trị hệ số hồi quy, ngưỡng dự báo P_0 và hiệu quả dự báo được lưu lại cho đến khi gặp một tổ hợp biến khác cho hiệu quả dự báo hơn thì chúng sẽ được thay thế bởi các giá trị mới.

3. 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], dông nhiệt hoạt động chủ yếu trong khoảng thời gian từ 13 đến 22 giờ với tần suất xuất hiện dông nhiệt lớn nhất tập trung vào lúc 19 giờ. Cho nên, yếu tố dự báo (dông 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 dông 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 dông nhiệt xuất hiện cũng cần phải tính đến.

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

1. Chín 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).

Theo kết quả tính chỉ số LI cho khu vực Hà Nội [2], chỉ số này lớn ($>-2,91$) hoặc rất nhỏ ($<-5,03$) thì xác suất không đồng đều lớn. Chính vì vậy, ta có thể hiệu chỉnh chỉ số này bằng một phép cộng đại số sao cho tất cả những giá trị của LI $>-2,91$ đều sẽ nhận giá trị $<-5,03$. Đến đây ta có thêm một nhân tố dự báo nữa, đó là chỉ số LI đã hiệu chỉnh, nhân tố ChLI.

b. Sự phụ thuộc của xác suất xuất hiện đông nhiệt vào các nhân tố dự báo

Xác định sự phụ thuộc của xác suất xuất hiện đông nhiệt vào từng nhân tố dự báo là tìm mối quan hệ giữa P với các x_i theo công thức (2.7). Đây là quá trình nhận dạng sự phụ thuộc của hàm P_i theo các x_i . Để nhận dạng xấp xỉ hàm P_i , cần tiến hành theo các bước sau:

- a) Vẽ đồ thị hàm (2.7) với trục tung là P_i và trục hoành là x_i . Loại bỏ những nhân tố không biểu hiện rõ qui luật phụ thuộc hoặc đường biểu diễn quá phức tạp.
- b) Chọn lớp hàm xấp xỉ là lớp hàm đa thức với bậc được lấy cao nhất là bậc 4.
- c) Thuật lựa chọn được thực hiện theo sơ đồ trên hình 2.

Theo sơ đồ này, trước khi xét cho nhân tố x_i , ta cần ấn định hai tham số lọc S_0 và S_c . Trong đó, S_0 là mức độ suy giảm của sai số chuẩn của phương trình hồi quy, S_c là sai số cho phép. Ta thấy bậc của đa thức càng tăng thì S_0 càng giảm nhưng sự suy giảm đó cần đạt đến mức đáng kể để cho việc nâng bậc đa thức trở nên có ý nghĩa. Nếu bậc của đa thức đã tăng lên đến tối đa (bậc 4) mà sai số chuẩn vẫn còn quá lớn thì loại bỏ nhân tố đang xét. Vì vậy, cần phải đưa vào tham số S_c là sai số cho phép của phép xấp xỉ. Kết quả tính theo nguyên tắc này cho thấy, trong số 33 biến trên, không có biến nào có quan hệ hàm bậc 4 với yếu tố dự báo.

c. Xây dựng các phương trình dự báo

Để xây dựng phương trình dự báo ta phải thực hiện quy trình tính toán được trình bày trong mục II với các nhân tố dự báo được chọn theo tiểu mục 3.2.

Kết quả cuối cùng, phương trình dự báo đông nhiệt cho khu vực Hà Nội là:

$$y_{HN} = 0,70037 - 0,02091 * TBTTd3 + 0,07919 * ChLI + 0,00977 * Td10_7 + 0,00445 * T710_70 \quad (3.1)$$

Phương trình (3.1) có ngưỡng dự báo $P_0 = 0,530$.

Từ các kết quả này ta nhận thấy rằng, số biến tham gia dự tuyển là 33 (kể cả biến bậc 2 và bậc 3) nhưng số biến tham gia vào các phương trình dự báo là 4 và trong cả 4 biến này đều là bậc nhất. Trong quá trình tính toán theo phương pháp hồi qui từng bước, TBTTd3 là biến xuất hiện đầu tiên. Như vậy, độ ẩm trung bình của cả 3 mực tầng thấp có thể được xem là nhân tố có vai trò quan trọng nhất đối với quá trình hình thành đông nhiệt trên khu vực Hà Nội. Các nhân tố quan trọng tiếp theo đối với sự hình thành đông nhiệt trong khu vực này là độ chỉ số độ bất ổn của khí quyển ChLI, sự tăng trữ lượng ẩm tầng thấp trong thời gian từ 7 đến 10 giờ và hiệu nhiệt độ giữa mực mặt đất và 700mb.

d. 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 đông 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_2 và F_1 là các pha thời tiết tương ứng với hiện tượng dông nhiệt có xảy ra và không xảy ra trong thực tế, P_2 và P_1 là dự báo các pha đó. Khi đó, nếu tiến hành dự báo N lần, ta sẽ nhận được kết quả sau:

Có n_{11} là số lần dự báo không dông và thực tế cũng không dông,
 n_{22} là số lần dự báo có dông và thực tế cũng có dông,
 n_{12} là số lần dự báo không dông nhưng thực tế lại có dông,
 n_{21} là số lần dự báo có dông nhưng thực tế lại không dông,
 N_{10} là số lần dự báo không dông,
 N_{20} là số lần dự báo có dông,
 N_{01} là số lần thực tế không dông,
 N_{02} là số lần thực tế có dông.

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

$$U = \frac{n_{11} + n_{22}}{N} \quad (3.2)$$

Ngoài ra, Bagrov N. A. [3] đã đưa ra một tiêu chuẩn nữa gọi là tiêu chuẩn độ tin cậy H . Độ tin cậy H được xác định như sau:

$$H = \frac{U - U_0}{1 - U_0} \quad (3.3)$$

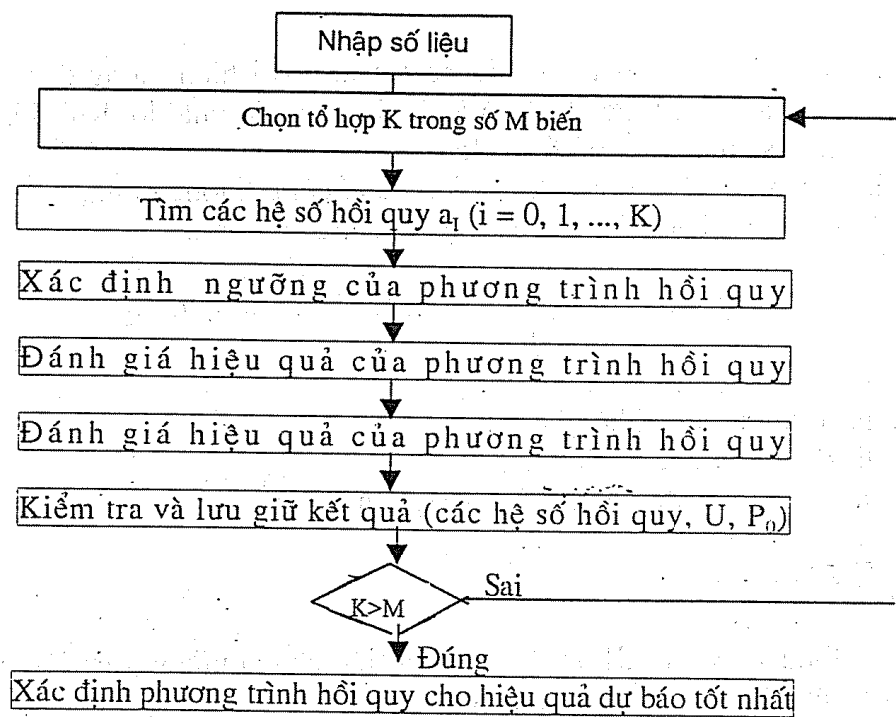
Trong đó, U_0 được gọi là độ chính xác toàn phần của dự báo ngẫu nhiên và được xác định theo công thức sau:

$$U_0 = \frac{1}{N} \left(\frac{N_{01}}{N} N_{10} + \frac{N_{02}}{N} N_{20} \right) \quad (3.4)$$

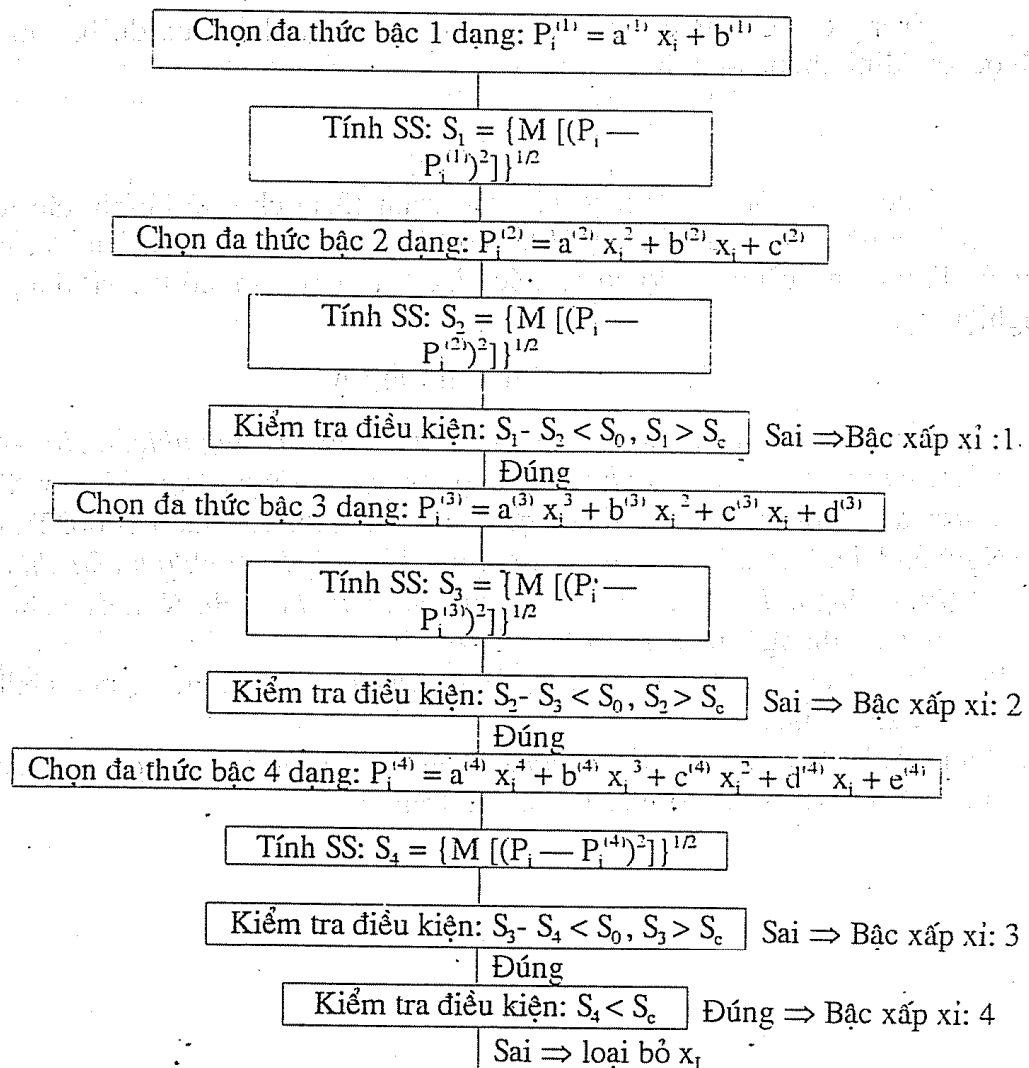
Theo các công thức này, phương trình (3.1) cho độ chính xác toàn phần là 81,93% và tiêu chuẩn độ tin cậy H là 0,6 (theo tác giả thì $H > 0,2$ là đạt yêu cầu). Có nghĩa là phương trình có độ chính xác và độ tin cậy cao, có thể sử dụng được trong nghiệp vụ.

Tài liệu tham khảo

1. Nguyễn Viết Lành. *Sự tương quan giữa mưa rào và dông 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 Viết Lành. *Về khả năng sử dụng chỉ số độ bất ổn định của khí quyển để dự báo dông 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. Phan Văn Tân. *Phương pháp thống kê trong khí hậu*. Đại học Quốc gia Hà Nội, 1999. 237 trang.
4. Daniel S. Wilks. *Statistical Methods in the Atmospheric Sciences*. Academic Press San Diego California, 1995. 465 trang.



Hình 1: Sơ đồ biểu diễn thuật toán lọc nhân tố khi xây dựng phương trình hồi quy



Hình 2: Sơ đồ khởi lựa chọn dạng hàm phụ thuộc của P vào các x_i