

QUY TẮC BAYES VÀ ỨNG DỤNG NÓ TRONG DỰ BÁO SƯƠNG MÙ Ở VỊNH BẮC BỘ

PGS-PTS Trần Tân Tiến
Đại học quốc gia, Hà Nội

Sương mù trên biển và ven bờ vịnh Bắc Bộ thường quan trắc được vào các tháng XI - IV. Trên các đảo và các trạm khí tượng ven biển sương mù có tần suất lớn nhất vào tháng III. Theo số liệu thống kê 10 năm (1976-1985), Hòn Ngư là trạm có sương mù nhiều nhất, trung bình tháng III có 13,4 ngày, cả năm 37,1 ngày có sương mù; Bãi Cháy là trạm có sương mù ít nhất, tháng III có 5,2 ngày và toàn năm 9,5 ngày. Sương mù ở trên biển và vùng ven biển ảnh hưởng đến giao thông, quốc phòng và phát triển kinh tế của nước ta. Vì vậy, dự báo sương mù biển là vấn đề được nhiều ngành kinh tế quan tâm. Trong công trình này chúng tôi sử dụng số liệu quan trắc ở các trạm khí tượng ở các đảo và ven bờ vịnh Bắc Bộ để xây dựng chỉ tiêu dự báo sự xuất hiện sương mù theo quy tắc Bayes.

1. Quy tắc Bayes

Giả sử ta có hai pha thời tiết là 1 và 2. Vectơ nhân tố dự báo hai pha thời tiết này là $X = \{x_1, x_2, \dots, x_n\}$, ở đây x_j là nhân tố dự báo. Các nhân tố này quan trắc N lần, khi đó ta ký hiệu

$$X_j = \{x_{1j}, x_{2j}, \dots, x_{nj}\}$$

($j = 1, N$). Với tập số liệu trên ta tính được $P(X/1)$ và $P(X/2)$. Bài toán đặt ra là phải chia miền xác định của X thành 2 miền A_1 và A_2 sao cho $X \in A_1$ dự báo thời tiết là pha 1, $X \in A_2 \rightarrow 2$.

Xác suất dự báo hai pha thời tiết 1 và 2 được xác định theo các miền A_1 và A_2 như sau:

$$\begin{aligned} P(2/1) &= \int_{A_2} P(X/1) dx = 1 - \int_{A_1} P(X/1) dx \\ P(1/2) &= \int_{A_1} P(X/2) dx \end{aligned} \quad (1)$$

Hàm số đặc trưng cho mức độ sai số trung bình của dự báo là:

$$\bar{L} = L_1 P(2/1) P_1 + L_2 P(1/2) P_2 \quad (2)$$

ở đây L_1, L_2 là hiệu của hiệu quả dự báo đúng và sai của pha thời tiết 1 và 2.

Thay (1) và (2) ta được

$$\bar{L} = \int_{A_1} [-L_1 P_1 P(X/1) + L_2 P_2 P(X/2)] dx + L_1 P_1 \quad (3)$$

Để \bar{L} đạt giá trị nhỏ nhất thì miền A_1 phải là những điểm thỏa mãn điều kiện:

$$L_1 P_1 P(X/1) \geq L_2 P_2 P(X/2) \quad (4)$$

Trong thực tế dự báo ta lấy $L_1 P_1 = L_2 P_2$, khi đó từ (4) ta có qui tắc dự báo của Bayes:

$$\frac{P(X/1)}{P(X/2)} \geq 1 \text{ dự báo pha thời tiết 1} \quad (5)$$

$$\frac{P(X/1)}{P(X/2)} < 1 \text{ dự báo pha thời tiết 2}$$

2. Phương pháp xác định xác suất $P(X/1)$ và $P(X/2)$

Theo số liệu lưu trữ ta chia vectơ x thành hai lớp có sương mù và không có sương mù. Mỗi nhân tố dự báo x_i được chia thành k_i khoảng. Khi đó ta được tổ hợp B_t khoảng

$$B_t = \{b_{11}, b_{12}, \dots, b_{k_1 k_1}; b_{21}, b_{22}, \dots, b_{2 k_2}, \dots, b_{n1}, b_{n2}, \dots, b_{k_n k_n}\} \quad (6)$$

Trong từng trường hợp mỗi nhân tố chỉ nằm trong một khoảng chia nên ta ký hiệu:

$$B_i = (b_{1k_1}, b_{2k_2}, \dots, b_{nk_n})_i \quad (7)$$

Tập hợp B_i này được sử dụng để tìm xác suất của nhân tố X rơi vào khoảng B_i trong từng lớp: có sương mù $P(B_i/1)$ và không có sương mù $P(B_i/2)$. Để tìm xác suất này ta sử dụng công thức của Bakhaduz:

Đối với lớp sương mù:

$$P(X/1) = P(b_{1k_1}, b_{2k_2}, \dots, b_{nk_n}) = P_1(X) \varphi(X) \quad (8)$$

ở đây $P_1(X)$ là xác suất khi các đại lượng b_i là độc lập.

$\varphi(X)$ là hàm tính đến sự tương quan của các đại lượng b_i . Các hàm này được tính như sau:

$$P_1(X) = \prod_{i=1}^n P(b_i k_i) = \prod_{i=1}^n m_i k_i$$

$$\varphi(x) = 1 + \sum_{i < j} \tau_{k_1 k_j} \cdot Z_{1k_i} Z_{2k_j} +$$

$$\sum_{i < j < e} \tau_{k_i k_j k_e} Z_{1k_i} Z_{2k_j} Z_{3k_e} +$$

$$\tau_{k_1 k_2 \dots k_n} Z_{1k_1} Z_{1k_2} \dots Z_{nk_n} \quad (9)$$

ở đây

$$Z_{nk_i} = (1 - m_{nk_i}) / \sqrt{mnKi (1 - mnKi)}$$

$$m_{k_i} = p (b_{nk_i})$$

$$\tau_{k_i k_j} = \sqrt{z_{1k_i} z_{2k_j}}$$

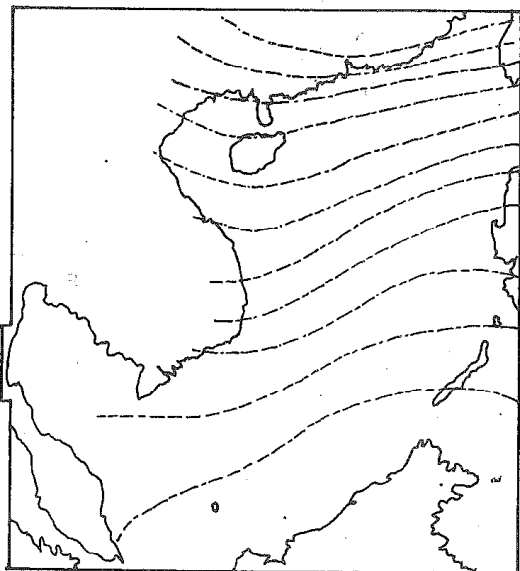
$$\tau_{k_i k_j k_e} = \sqrt[3]{z_{1k_i} z_{2k_j} z_{3k_e}}$$

Tương tự, ta có công thức để tính cho trường hợp không có sương mù $P(X/2)$. Biết $P(X/1)$ và $P(X/2)$ sử dụng (5) sẽ dự báo được sự xuất hiện sương mù.

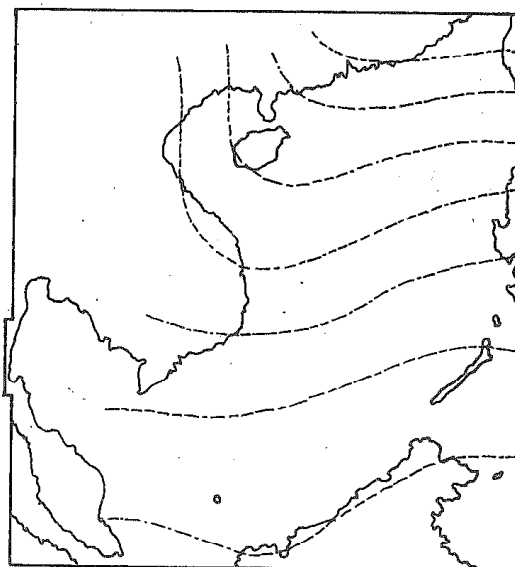
3. Dự báo sương mù ở các đảo và ven bờ tây bắc vịnh Bắc Bộ

Sử dụng số liệu 13 năm (1976-1988) ở 7 trạm khí tượng trên biển và ven bờ tây bắc vịnh Bắc Bộ để dự báo sương mù từ 1 giờ đến 8 giờ sáng theo số liệu 13 và 19 giờ ngày hôm trước. Các nhân tố dự báo được chọn là nhiệt độ T , điểm sương T_d , độ ẩm tương đối R , tốc độ gió V_{13} giờ và 19 giờ. Đã tiến hành tuyển chọn các nhân tố theo phương pháp đánh giá lượng thông tin về sự hình thành sương mù của từng nhân tố.

Đã tìm được qui tắc dự báo sương mù cho 7 trạm Cô Tô, Hòn Dấu, Bạch Long Vĩ, Hòn Ngư, Bãi Cháy, Cửa Ông và Văn Lý trong trường hợp không phân loại hình thể khí áp mặt đất và trong trường hợp hình thể 1 và hình thể 2 là các hình thể có sương mù (hình 1, 2) theo phân loại của Nguyễn Vũ Thi.



HÌNH 1: HÌNH THỂ 1



HÌNH 2: HÌNH THỂ 2

Trong khuôn khổ bài báo này chúng tôi chỉ trình bày kết quả tính toán ở trạm Côtô:

Tháng XI-I (Nếu nằm trong các khoảng sau sẽ dự báo suất hiện sương mù).

Nhân tố	T13	R19	V19
1	10,2 — 12,8	91 — 100	2,6 — 7,9
2	15,4 — 18,1	91 — 100	0 — 5,3
3	20,7 — 23,3	91 — 100	0 — 2,6
4	25,9 — 28,6	91 — 100	2,6 — 5,3

Đánh giá kết quả dự báo trên dãy số liệu phụ thuộc (10 năm) và độc lập (3 năm):

Số liệu phụ thuộc				Số liệu độc lập			
TT/DB	0	1	Tổng	TT/DB	0	1	Tổng
0	91,1	8,9	80,9	0	89,8	10,2	24,6
1	27,8	72,2	18,0	1	66,7	33,3	6
Tổng	74,2	85,0	82,7	Tổng	22,5	27,0	25,2

Hiệu quả chung : 90,7% 88,5%

Dự báo thử cho tháng XII- 1994 và tháng I-1995 có 8 lần quan trắc có sương mù, dự báo đúng 6 lần.

Tháng II và tháng III

Hình thế 1

Nhân tố	Td 19	R13	V19
1	15 — 19	71 — 79	0 — 1,3
2	15 — 19	86 — 100	0 — 1,3

Đánh giá kết quả dự báo trên dãy số liệu phụ thuộc đạt 86,9%, trên dãy số liệu độc lập 94 %.

Hình thế 2

Nhân tố	T13	Td19	R13
1	13,2 — 17,4	14,4 — 22,0	88 — 100
2	21,6 — 25,8	10,6 — 14,4	88 — 100
3	21,6 — 25,8	14,4 — 18,2	75 — 88
4	21,6 — 25,8	14,4 — 22,0	88 — 100

Đánh giá kết quả dự báo trên dãy số liệu phụ thuộc đạt 83,4 %, độc lập đạt 80,5%.

Kết quả tính toán cho các trạm khác có thể tham khảo trong [1]. Kết quả tính toán cho thấy có thể sử dụng phương pháp này để dự báo sự xuất hiện sương mù ở các trạm khí tượng.

Tài liệu tham khảo

1. Trần Tân Tiến. Thử nghiệm dự báo sương mù ở vịnh Bắc Bộ. Báo cáo khoa học đề tài KT03 - 04, năm 1993.