

PHÂN VÙNG KHÁCH QUAN KHÍ HẬU VIỆT NAM TRÊN CƠ SỞ CHỈ SỐ GIỐNG NHAU GIỮA CÁC ĐÀI TRẠM

NGUYỄN DUY CHÍNH

Viện Khí tượng Thủy văn

Sự xuất hiện và sử dụng ngày càng nhiều những hệ máy tính điện tử có tốc độ tính toán cao, bộ nhớ lớn (và do vậy các mô hình toán học cũng được áp dụng rộng rãi trong thực tế) đánh dấu một bước phát triển quan trọng và một triển vọng khả quan của môn toán thống kê. Môn toán thống kê đã đưa ra được nhiều mô hình sử dụng giúp cho việc mô tả những hiện tượng tự nhiên một cách định lượng và khách quan. Trong đó phương pháp phân tích nhân tố, phân tích liên kết, v.v... đã và đang được áp dụng vào việc phân vùng và phân loại, chẳng hạn trong ngành tâm lý, ngành y, giao thông, quy hoạch thành phố, kinh tế kế hoạch, ... Trong ngành Khí tượng, địa lý, thủy văn phương pháp phân tích nhân tố cũng được áp dụng có kết quả vào nhiều mục đích nghiên cứu. Một trong những ứng dụng có hiệu quả của phương pháp này là việc phân vùng và phân loại khí hậu dựa trên tính đặc trưng khách quan của các nhân tố. Theo chúng tôi biết D.Steiner [4,5] — giáo sư người Áo — trong chuyến đi tham dự Hội nghị chuyên đề toán thống kê «Quantitative methods and Regional analysis» ở nước Mỹ năm 1965 đã áp dụng phương pháp phân tích nhân tố và phân tích phân biệt vào việc phân vùng khách quan khí hậu nước Mỹ.

Trong công trình của mình Steiner đã chọn 16 tham biến của 4 yếu tố nhiệt độ, giáng thủy, độ nắng và độ độ ẩm ở 67 đài trạm khí tượng nước Mỹ có độ dài chuỗi quan trắc thống nhất là 30 năm cho việc phân tích của mình. Bằng phương pháp rút ra nhân tố ông đã thu được 4 nhân tố đặc trưng cho 16 tham biến nói trên. Các tải trọng nhân tố được quy về đài trạm để thiết lập không gian nhân tố trực giao. Steiner đã thu được không gian nhân tố trực giao bé hơn 4 lần so với không gian số liệu ban đầu, mà trong ý nghĩa toán thống kê lượng thông tin không bị hao hụt. Thông qua phương pháp xác định chỉ số giống nhau bằng khoảng cách giữa các đài trạm trong không gian nhân tố và thông qua phân tích phân biệt ông đã chia nước Mỹ ra 10 vùng khí hậu khác nhau một cách khách quan. Sự phân vùng này đã được giới các nhà khí tượng công nhận và đánh giá tốt. 5 năm sau (1970) R.Stelimacher [6] (CHDC Đức) cũng đã áp dụng phương pháp phân tích nhân tố và phân tích liên kết vào việc phân vùng khí hậu CHDC Đức, tuy trước đó không lâu W.Boer [1] đã phân vùng CHDC Đức bằng hệ thống chỉ tiêu kinh nghiệm. Bằng những kết quả thu được Stelimacher đã khẳng định rằng: «Nếu người ta giới hạn ở việc áp dụng các giá trị trung bình nhiều năm thì phương pháp phân tích liên kết, kỹ thuật R hoặc kỹ thuật Q của phân tích nhân tố có thể áp dụng được vào việc phân vùng và phân loại khí hậu sau khi đã kiểm tra mức tin

cặp của các kiểu và các vùng khí hậu thu được bằng phân tích phương sai nhiều biến» (trích dịch).

Ở nước ta trên quy mô lớn có sự phân hóa khí hậu tương đối rõ rệt và trên quy mô vừa trong điều kiện địa hình phức tạp cũng có sự phân hóa theo khu vực, cho nên vấn đề phân vùng khí hậu nước ta đã được đặt ra trong chương trình nghiên cứu khí hậu từ lâu. Công trình về phân vùng khí hậu Việt nam bằng phương pháp chỉ tiêu kinh nghiệm của Phạm Ngọc Toàn và Phan Tất Đắc [3] đã báo cáo và thảo luận ở Hội nghị khoa học Tổng cục Khí tượng thủy văn. Tuy vậy, ở đây chúng tôi vẫn mạnh dạn áp dụng phương pháp phân tích nhân tố và phân tích liên kết vào việc xác định khách quan ranh giới các vùng khí hậu nước ta và xem đó là một thử nghiệm cần thiết cho các ứng dụng phong phú khác, chẳng hạn cho việc phân vùng khách quan khí hậu xây dựng Việt nam, v.v..

2. Vài nét về phương pháp phân tích nhân tố

Chọn m tham biến của N đại trạm khí tượng, như vậy ma trận số liệu đưa vào phân tích là một ma trận gồm N hàng và m cột. Các giá trị của ma trận này trước hết được chuyển thành các giá trị chuẩn hóa, sau đó xác định ma trận tương quan tuyến tính để tìm các biến thiên giữa các cặp tham biến. Phương trình biểu diễn mô hình toán học cơ bản của phương pháp phân tích nhân tố được viết ở dạng chuyển vị và hoán vị các ma trận, có dạng:

$$Z(N, m) = F(N, p) \cdot A(p, m) \quad (1)$$

trong đó p là nhân tố được rút ra và thường nhỏ hơn số tham biến m nhiều. Vấn đề đầu tiên khi giải phương trình trên là vấn đề xác định ma trận tải trong nhân tố A , bởi vì ma trận các tải trong nhân tố quy về đại trạm F là ma trận kết quả chưa biết và chỉ có thể xác định được khi các ma trận Z và A đã tính được (Z là ma trận các giá trị chuẩn hóa). Các tải trong nhân tố chính là những thành phần vectơ riêng của ma trận tương quan và chính vì thế việc giải bài toán tìm các tải trọng nhân tố chính là giải bài toán tìm các giá trị riêng quen biết. Việc rút ra nhân tố được tiến hành từng bước theo quy tắc lập tổng bình phương các tải trọng nhân tố lớn nhất. Việc chấm dứt rút ra nhân tố khi tổng các giá trị riêng (tổng phương sai) nhỏ hơn 1. một cách lý thuyết thì các nhân tố được rút ra không tương quan lẫn nhau, nhưng trong thực tế tính toán thì chúng ít nhiều vẫn tương quan với nhau. Để cải tiến chất lượng các nhân tố người ta thường thực hiện thêm phép xoay trục giao các nhân tố, mà thông dụng nhất là phép xoay Varimax. Bước cuối của phân tích các nhân tố sau khi đã xác định được ma trận nhân tố là việc quy các giá trị nhân tố về đại trạm, thiết lập không gian các nhân tố trục giao.

Quá trình thực hiện phương pháp phân tích nhân tố trong thực tế bây giờ có thể nêu vắn tắt như sau:

- 1) Vạch kế hoạch thực hiện và mục đích áp dụng của bài toán, nhằm áp dụng được các kỹ thuật một cách thích hợp và loại trừ những kết quả bất hợp lý.

- 2) Thống kê và chỉnh lý ma trận số liệu ban đầu. Số tham biến và đại trạm cần phải thống nhất trong cả ma trận và cần phải lớn và chính xác đến mức cho phép.

- 3) Tính ma trận tương quan tuyến tính từ số liệu «thô» hoặc từ số liệu

đã chuẩn hóa. Dĩ nhiên, nếu ma trận tương quan được xác định từ số liệu chuẩn hóa, thì trước đó phải chuẩn hóa số liệu

4) Tính các tải trọng nhân tố bằng cách quy về bài toán xác định giá trị riêng.

5) Xoay trục giao cấu trúc nhân tố, chẳng hạn bằng quy tắc Varimax.

6) Đánh giá các nhân tố được rút ra và lập nhóm biến trong từng nhân tố, nhằm rút ra được các đại lượng ảnh hưởng quan trọng, đồng thời loại bỏ các đại lượng không quan trọng trong khi phân tích kết quả.

7) Xác định không gian các nhân tố trục giao.

3. Phân nhóm các đài trạm bằng phân tích liên kết trên cơ sở các khoảng cách trong không gian nhân tố.

2.1. Phân tích liên kết.

Cơ sở của việc phân nhóm bằng phân tích liên kết là chỉ số giống nhau giữa các quan sát (đài trạm). Các đài trạm càng giống nhau khi hệ số tương quan giữa chúng càng lớn hoặc khi khoảng cách giữa chúng càng bé. Như vậy các chỉ số giống nhau cực trị được ghép thành một nhóm. Chẳng hạn, Nếu A_{ij} là chỉ số cực trị của ma trận phân tích, nghĩa là hệ số tương quan là cực đại và khoảng cách giữa chúng là cực tiểu, thì các đài trạm mang chỉ số i và j thuộc vào một nhóm. Nếu T_k cũng là một chỉ số lớn tiếp theo thì trạm mang chỉ số k cũng thuộc vào một nhóm với i và j . Phương pháp này cho phép liên kết đến khi không còn xuất hiện một chỉ số lớn tiếp theo nào nữa giữa một trạm đã phân nhóm và một trạm chưa phân nhóm.

Các bước thực hiện của phương pháp phân tích liên kết có thể nêu tóm tắt như sau:

1) Gạch dưới giá trị cực trị trong mỗi cột của ma trận đưa vào phân tích (giá trị lớn nhất trong ma trận tương quan và bé nhất trong ma trận khoảng cách).

2) Tìm giá trị cực trị của ma trận. Hai đài trạm phụ thuộc được xếp vào một nhóm, chẳng hạn là nhóm A.

3. Tìm các đài trạm khác giống nhóm A, trong đó người ta kiểm tra liệu còn có một giá trị cực trị tiếp theo thuộc vào một trong hai trạm đó không?

4. Chọn các đài trạm giống các đài trạm tìm được ở bước 3.

5. Phương pháp này thực hiện đến khi không còn đài trạm phân nhóm nào có chỉ số gần giống nhất với một trạm chưa phân nhóm khác.

6. Các đài trạm thuộc các nhóm khác, chẳng hạn B, C, D, v.v... được xác định tương tự như trình tự xác định ở nhóm A.

2.2. Ma trận khoảng cách

Để đạt được mục đích phân vùng và phân loại khí hậu bằng phân tích liên kết trên cơ sở các đài trạm đã chọn, trước hết phải xác định ma trận các chỉ số giống nhau (ma trận khoảng cách trong không gian nhân tố trục giao). Các giá trị f_{ij} cho biết trị số đặc trưng của các đài trạm trong không gian nhân tố. Khoảng cách giữa các đài trạm như đã nói ở trên là thước đo của sự tương tự giống nhau giữa chúng. Chính vì thế mà việc thành lập nhóm của N đài trạm trên cơ sở các chỉ số giống nhau này dẫn đến sự phân vùng và

phân loại khách quan. Để thu được ma trận khoảng cách mong muốn, trước hết chúng ta xác định tất cả các khoảng cách cho phép theo từng cặp d_{ik} với phương trình:

$$d_{ik} = \sqrt{(f_{i1} - f_{k1})^2 + (f_{i2} - f_{k2})^2 + \dots + (f_{ip} - f_{kp})^2} \quad (2)$$

Trong đó $i, k = 1, 2, \dots, N$; p là số nhân tố được rút ra. Ma trận khoảng cách D có bậc (N, N) , trong đó các giá trị trên đường chéo chính d_{ii} phải bằng 0, vì mỗi trạm chỉ có $N-1$ khoảng cách trong số N đài trạm đưa vào phân tích và không có khoảng cách với chính nó:

$$D(N, N) = \begin{pmatrix} 0 & d_{12} & \dots & d_{1N} \\ d_{21} & 0 & \dots & d_{2N} \\ \dots & \dots & \dots & \dots \\ d_{N1} & d_{N2} & \dots & 0 \end{pmatrix} \quad (3)$$

3. Áp dụng phân tích nhân tố và phân tích liên kết vào việc phân vùng khí hậu Việt nam:

Như đã biết, số liệu đưa vào phân tích nhân tố là có giới hạn và cần phải thống nhất về cách xác lập. Việc xác lập và chọn các tham biến và đài trạm quan sát có ảnh hưởng nhất định đến kết quả cuối cùng. Đó cũng là yếu tố chủ quan đầu tiên của bài toán phân tích khách quan này. Để hạn chế bớt sai số do yếu tố chủ quan gây ra, người ta thường chọn nhiều tham biến và nhiều đài trạm đưa vào phân tích. Chọn nhiều quá không có hại, vì những đặc trưng không có phần đóng góp quan trọng sẽ bị loại bỏ, ngược lại khi chọn quá ít thì hai hoặc nhiều kiểu khí hậu sẽ lẫn vào nhau. Tuy vậy, khi chọn số tham biến và đài trạm cũng phải cân nhắc cho hợp lý, đồng thời cũng cần phải tính đến yếu tố kinh tế và khả năng bộ nhớ máy tính hiện có. Giả thiết quan trọng đối với bài toán phân vùng và phân loại khí hậu là phải chọn khoảng thời gian phân tích giống nhau đối với tất cả các đài trạm được chọn. Trên cơ sở đó chúng ta có thể giả thiết rằng các kết quả tìm được có giá trị trong cùng khoảng thời gian lựa chọn.

Chúng tôi chọn ma trận số liệu ban đầu gồm 76 đài trạm và 21 tham biến của 4 yếu tố: Nhiệt độ, lượng mưa, độ ẩm và độ nắng (xem bảng 1 và 2). Dựa trên các kết quả khảo sát và nghiên cứu, các tác giả phân vùng trong và ngoài nước cho rằng các tương quan nhiệt, ẩm mùa đông và mùa hè đóng vai trò quan trọng nhất trong việc thể hiện sự phân hóa khí hậu. Do vậy, trong các công trình của mình, các tác giả đều chọn các đặc trưng chỉ thị giữa mùa đông và giữa mùa hè, các đặc trưng trung bình năm cũng như các tỷ số giữa các đặc trưng đó vào việc phân vùng và phân loại khí hậu. Chúng tôi cho rằng việc lựa chọn như vậy là hợp lý và có thể áp dụng được trong điều kiện khí hậu nước ta. Sự phân bố của 76 đài trạm trên không được đều, phần miền trung quá thưa thớt (xem hình 1). Thực tế này mang lại cho việc phân tích kết quả những khó khăn và hạn chế nhất định. Độ dài chuỗi quan trắc của các đài trạm lựa chọn phần lớn là từ 1960 đến 1975 và các giá trị trong ma trận là số liệu trung bình tính theo độ dài chuỗi quan trắc đã chọn.

Bảng 1: Danh sách các đài trạm chọn vào phân tích
Vị trí đài trạm xem hình 1).

Số thứ tự	Tên trạm	Vĩ độ °N	Kinh độ °E	Số thứ tự	Tên trạm	Vĩ độ °N	Kinh độ °E
1	Côn đảo	8°42'	106°35'	39	Hòa bình	20°50'	105°20'
2	An xuyên	9,10	105,10	40	Hải dương	20,56	106,19
3	Ba xuyên	9,36	105,58	41	Quảng ninh	20,57	107,04
4	Rạch giá	10,00	105,05	42	Cô tô	20,58	107,46
5	Phủ quốc	10,13	103,58	43	Láng	21,01	105,48
6	Sài gòn	10,49	106,40	44	Yên châu	21,03	104,17
7	Hiệp hòa	10,55	106,18	45	Sông mã	21,04	103,44
8	Phan thiết	10,56	108,06	46	Sơn tây	21,08	105,30
9	Dầu tiếng	11,20	106,20	47	Phủ yên	21,13	104,01
10	Bảo lộc	11,28	107,48	48	Bắc giang	21,17	106,13
11	Liên phương	11,45	108,23	49	Vĩnh yên	21,19	105,36
12	Nha trang	12,15	109,12	50	Sơn động	21,20	106,51
13	Buôn mê thuật	12,41	108,05	51	Sơn la	21,20	103,54
14	Quy nhơn	13,46	109,46	52	Điện biên	21,22	103,00
15	Plâyku	13,59	108,00	53	Lục ngạn	21,22	106,33
16	Quảng Ngãi	15,08	108,47	54	Tam đảo	21,27	105,38
17	Đà nẵng	16,03	108,12	55	Móng cái	21,31	107,59
18	Huế	16,24	107,41	56	Tuần giáo	21,35	105,50
19	Quảng trị	16,44	107,11	57	Thái nguyên	21,35	103,25
20	Đồng hới	17,29	106,36	58	Văn chấn	21,36	104,30
21	Kỳ anh	18,05	106,17	59	Yên Bái	21,42	104,52
22	Hà tĩnh	18,21	105,15	60	Tuyên quang	21,49	105,12
23	Kim cương	18,27	105,15	61	Lạng sơn	21,50	106,46
24	Vĩnh	18,41	105,40	62	Quỳnh nhai	21,50	103,34
25	Đô lương	18,54	105,18	63	Thần uyển	22,01	103,55
26	Con cuông	19,02	104,54	44	Lai châu	22,03	103,09
27	Quỳnh lưu	19,08	105,38	65	Lục yên	22,05	104,43
28	Tương dương	19,16	104,27	66	Bắc cạn	22,09	105,49
29	Tây hiệu	19,19	105,25	67	Mường nhé	22,11	102,28
30	Quý châu	19,34	105,06	68	Thất Khê	22,15	106,28
31	Thanh hóa	19,48	105,46	69	Sapa	22,20	103,50
32	Văn lý	20,07	106,18	70	Sìn hồ	22,21	103,15
33	Nho quan	20,19	105,44	71	Mường tè	22,22	102,50
34	Thái bình	20,26	106,20	72	Lào cai	22,30	103,57
35	Nam định	20,26	106,09	73	Bắc hà	22,32	104,17
36	Hưng yên	20,40	106,06	74	Hà giang	22,49	104,59
37	Phủ liên	20,49	106,38	75	Trùng khánh	22,50	106,31
38	Mộc châu	20,49	104,41	76	Phổ bình	23,15	105,11

(Xem hình 1 ở trang sau)

Bảng 2: Các trạm tham biến lựa chọn.

X_1 : Nhiệt độ trung bình năm

X_2 : Nhiệt độ trung bình tháng I

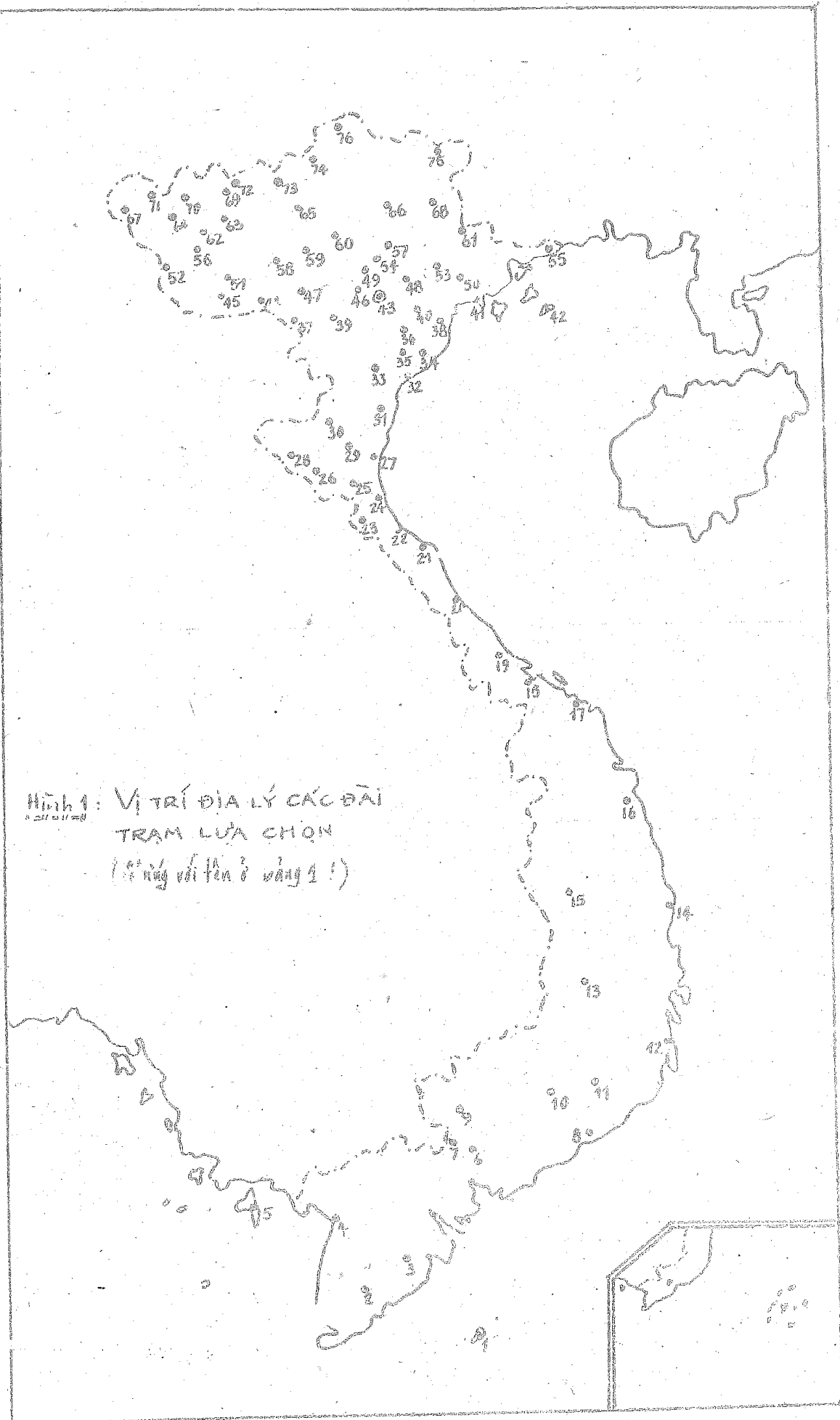
X_3 : Nhiệt độ trung bình tháng VII

X_4 : Nhiệt độ trung bình 3 tháng mùa hè

X_5 : Nhiệt độ 3 tháng mùa đông

X_6 : Biên độ nhiệt độ trung bình năm

- X₇ : Tỷ nhiệt trung bình tháng VII/I
- X₈ : Lượng mưa trung bình năm
- X₉ : Lượng mưa trung bình tháng I
- X₁₀ : Lượng mưa trung bình tháng VII
- X₁₁ : Lượng mưa trung bình 3 tháng mùa hè
- X₁₂ : Lượng mưa trung bình 3 tháng mùa đông
- X₁₃ : Tỷ số lượng mưa trung bình tháng VII/I
- X₁₄ : Độ ẩm tương đối trung bình năm.
- X₁₅ : Độ ẩm tương đối trung bình tháng I
- X₁₆ : Độ ẩm trung bình tháng VII
- X₁₇ : Tỷ số độ ẩm tương đối trung bình tháng VII/I
- X₁₈ : Độ nắng trung bình năm
- X₁₉ : Độ nắng trung bình tháng I
- X₂₀ : Độ nắng trung bình tháng VII
- X₂₁ : Tỷ số phần trăm độ nắng trung bình tháng VII/I

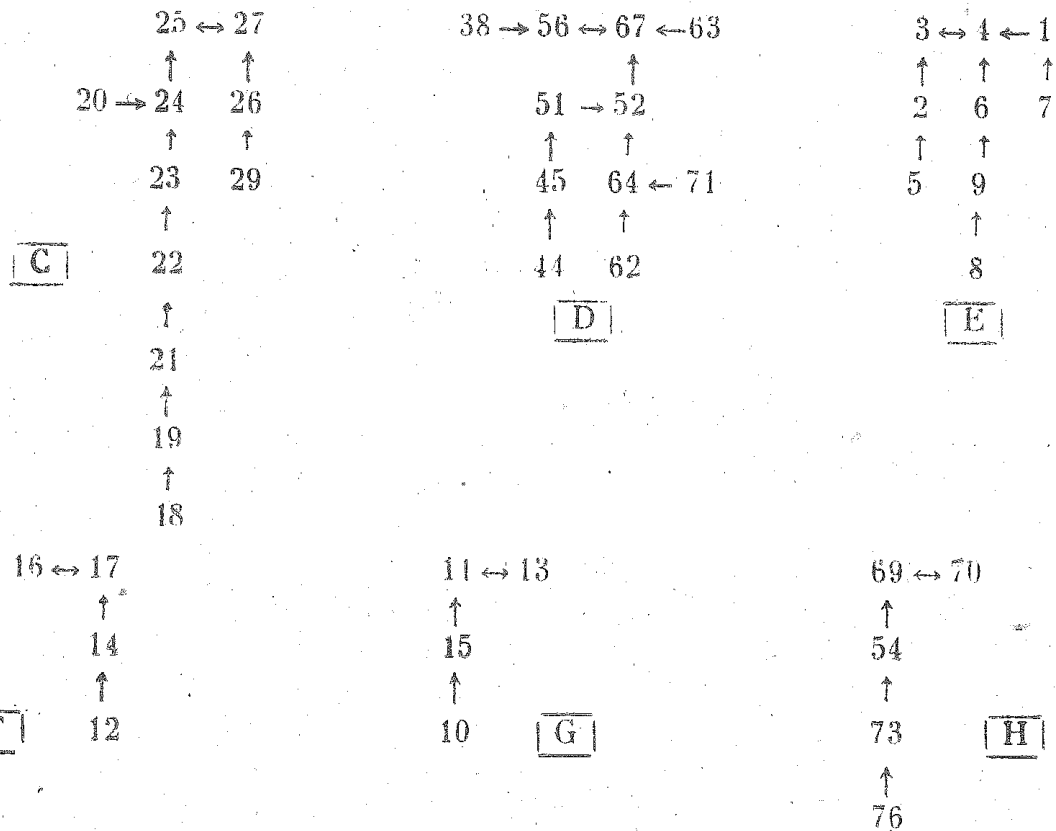
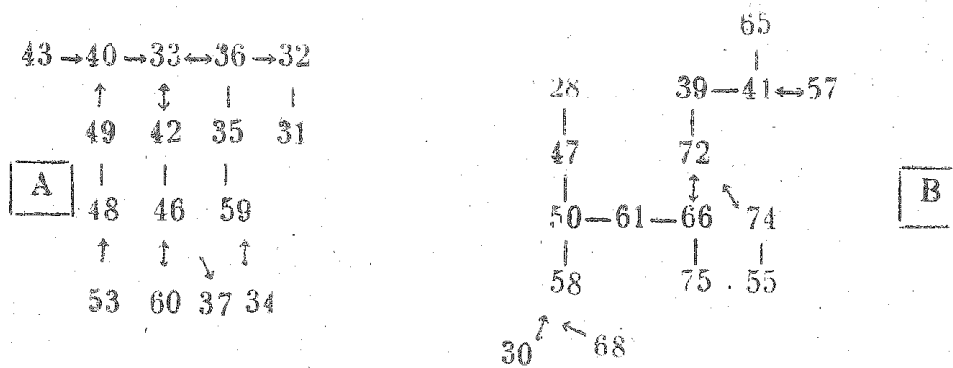


Hình 1: VỊ TRÍ ĐỊA LÝ CÁC ĐÀI
TRAM LỬA CHỌN
(Số này với tên ở bảng 1 !)

Bảng 3: Ma trận nhân tố đã xoay trục giao theo quy tắc Varimax

Tham biến	Nhân tố I	Nhân tố II	Nhân tố III	Nhân tố IV	Nhân tố V
X1	0,459	-0,848	-0,011	-0,204	-0,007
X2	0,109	-0,968	0,035	-0,161	-0,080
X3	0,87	-0,250	0,007	-0,268	0,055
X4	0,860	-0,321	-0,001	-0,265	0,045
X5	0,123	-0,966	0,029	-0,168	-0,031
X6	-0,673	-0,086	-0,185	-0,145	-0,115
X7	0,138	0,951	-0,043	0,141	0,025
X8	0,043	-0,093	0,771	0,593	0,024
X9	0,185	-0,060	0,883	-0,197	-0,067
X10	-0,256	0,213	-0,093	0,919	-0,025
X11	-0,205	0,253	-0,132	0,919	0,027
X12	0,153	-0,134	0,894	-0,150	-0,092
X13	-0,468	-0,318	-0,343	0,076	0,033
X14	0,101	0,428	0,408	0,118	0,458
X15	0,204	0,422	0,681	-0,218	0,272
X16	-0,589	0,087	-0,401	0,060	0,123
X17	0,055	-0,033	-0,073	-0,015	0,908
X18	-0,293	-0,713	0,017	-0,116	-0,568
X19	-0,637	-0,722	-0,107	0,112	-0,008
X20	0,700	-0,210	0,225	-0,458	-0,277
X21	0,789	0,388	0,146	-0,252	0,112
Tổng các giá trị riêng	6,848	5,697	2,548	1,580	1,321
Độ phân tán (%)	32,61	27,13	12,13	7,53	6,29

Như đã nói ở trên, bài toán phân vùng này được áp dụng kỹ thuật R của phân tích nhân tố. Bằng những phương pháp xác định và rút ra nhân tố, chúng ta thu được 5 nhân tố đặc trưng cho 21 tham biến ban đầu (xem bảng 3). Nhân tố thứ nhất chiếm 32,61% tổng phương sai, được đặc trưng chủ yếu thông qua các đặc trưng nhiệt độ, độ nắng và độ ẩm mùa hè. Nhân tố II chiếm 27,13% tổng phương sai và được xác định chủ yếu thông qua các đặc trưng nhiệt độ và độ nắng mùa đông. Các nhân tố còn lại: nhân tố III chiếm 12,13%, nhân tố IV-7,53% và nhân tố V-6,29% tổng phương sai. Như vậy nhân tố I và II chiếm gần 60% tổng phương sai và đóng vai trò quan trọng nhất khi phân tích kết quả thu được ở bảng 4. Các đặc trưng nhiệt độ thể hiện được là đặc trưng cơ bản nhất trong sự phân hóa khí hậu. Không gian các nhân tố trục giao $F(N,p)$ bây giờ được xác định nhờ ma trận các giá trị chuẩn hóa $Z(N,m)$ và ma trận nhân tố $A(p,m)$ bằng phương trình (1). Ma trận các chỉ số giống nhau giữa các đài trạm được xác định bằng phương trình (2). Kết quả thu được là một ma trận bậc (76,76), với tất cả các giá trị trên đường chéo chính đều bằng không. Ma trận các chỉ số giống nhau này là cơ sở chủ yếu để phân vùng khách quan khí hậu nước ta. Bằng phương pháp phân tích liên kết, thu được 8 vùng khí hậu khác nhau như sau:



(Ghi chú; Các chữ số là số thứ tự các đài trạm ở bảng 1 và hình 1).

Vấn đề đặt ra ở đây là phải vạch đường ranh giới giữa các vùng như thế nào cho hợp lý. Trên cơ sở các đài trạm đã được ghép thành vùng chúng ta căn cứ thêm vào địa hình của các đài trạm cũng như dựa thêm vào biến trình năm của nhiệt độ và lượng mưa để xác định đường ranh giới giữa các vùng. Ở đây yếu tố địa hình đóng vai trò quan trọng, (bởi vì với độ cao địa hình cũng như dạng địa hình giữa các khu vực có thể xuất hiện những hình thái khí hậu khác nhau. [7]. Do vậy trong điều kiện địa hình phức tạp trên lãnh thổ nước ta việc dựa thêm vào yếu tố địa hình khi vạch ranh giới giữa các vùng khí hậu là có cơ sở hợp lý. Hình 2 biểu diễn kết quả cuối cùng của bài toán. cụ thể gồm 8 vùng khí hậu khác nhau:

Vùng A: Vùng đồng bằng Bắc bộ

Vùng B: Vùng trung du và núi thấp Bắc bộ.

Vùng C: Vùng Nghệ Tĩnh — Bình trị thiên

Vùng D: Vùng Tây bắc

Vùng E: Vùng đồng bằng Nam bộ

Vùng F: Vùng ven biển từ Đà Nẵng đến Thuận hải

Vùng G: Vùng cao nguyên Trung bộ

Vùng H: Vùng núi cao Hoàng liên sơn, Việt bắc và Tam đảo. (xem hình 2)

Việc vạch ranh giới này tuy dựa vào kết quả của các phương pháp phân tích khách quan, tuy nhiên vẫn không hoàn toàn tránh khỏi sự can thiệp của các yếu tố chủ quan, bởi vì: thứ nhất, phương pháp phân tích liên kết được dùng ở chỉ số giống nhau nào đó trong việc xác lập một vùng, ít nhiều mang tính chất chủ quan của người lập. thứ hai, các ranh giới được vạch ra có nơi đi qua những vùng lãnh thổ rộng lớn mà không có đài trạm chọn vào phân tích.

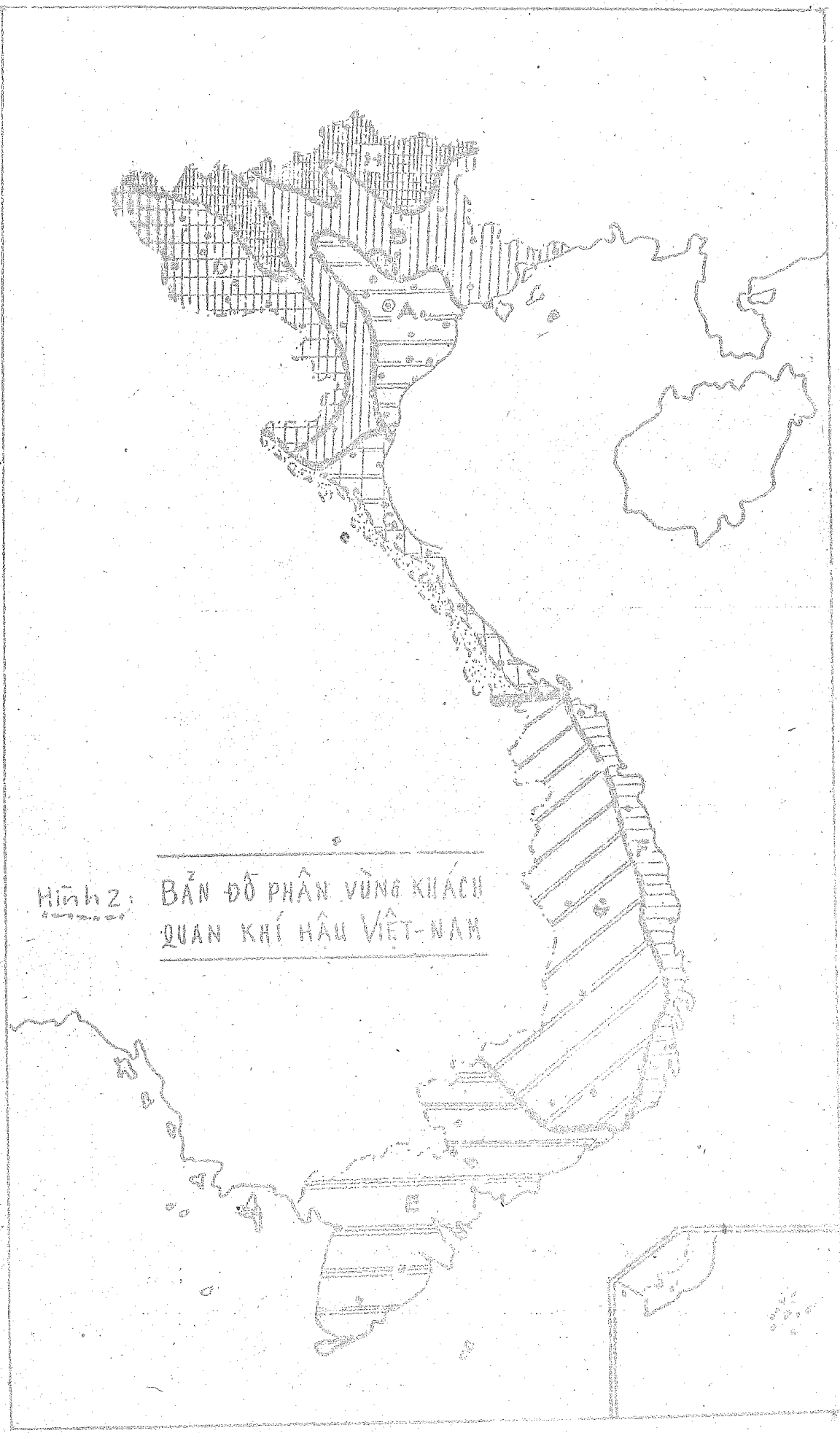
Bảng 4 biểu diễn các giá trị trung bình của các tham biến lựa chọn của 8 vùng khí hậu thu được ở trên. Căn cứ vào bảng 4 này chúng ta có thể rút ra được những nét đặc trưng cho từng vùng, đồng thời cũng tìm thấy sự khác nhau tương đối giữa các vùng đã tìm ra (căn cứ các tham biến quan trọng tìm được ở nhân tố I và nhân tố II).

Qua bảng 4 tác giả cũng thấy rằng trong phạm vi lãnh thổ Việt nam trên quy mô lớn có sự phân hóa khá rõ rệt, cụ thể là hình thành 2 miền khí hậu khác nhau (có tham khảo chỉ số nhiệt độ trung bình tháng I của Koppen và kết quả ở [3]):

Miền thứ nhất: gồm các vùng A, B, D, H và C – có mùa đông lạnh (đối với nhiệt đới) và ít mưa, mùa hè nóng nhiều mưa

Miền thứ hai: gồm các vùng F, G và E – hầu như ẩm và nóng quanh năm, mùa đông ít mưa nhiều nắng, mùa hè nhiều mưa.

Ranh giới giữa hai miền khí hậu nói trên là đèo Hải vân.



Bảng 4 **Trung bình các tham biến của các vùng khí hậu**

Miền	Vùng	Nhiệt độ (°C)						Lượng mưa (mm)						Độ ẩm (%)					Độ nắng (%)				
		X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈	X ₉	X ₁₀	X ₁₁	X ₁₂	X ₁₃	X ₁₄	X ₁₅	X ₁₆	X ₁₇	X ₁₈	X ₁₉	X ₂₀	X ₂₁	
Miền I	A	23,0	15,8	28,4	28,0	16,6	6,3	1,81	1740,6	32,0	267,7	284,5	26,5	13,53	84	82	83	1,01	38	27	50	1,95	
	B	22,4	15,0	27,6	27,3	16,0	7,8	1,85	1826,8	22,6	377,7	333,8	25,4	16,78	83	81	84	1,04	86	26	43	1,74	
	D	21,9	15,3	26,2	26,1	16,1	10,1	1,72	1782,4	19,9	342,3	339,5	33,9	18,38	83	81	87	1,07	42	42	34	0,85	
	H	16,6	9,6	21,4	21,0	10,5	6,6	2,24	2377,1	32,1	446,1	417,4	38,5	15,71	86	86	88	1,02	34	33	30	0,99	
	C	24,0	17,8	29,4	28,9	18,6	6,9	1,66	2175,5	76,1	133,9	155,6	85,3	3,50	84	88	74	0,84	39	25	55	2,21	
Miền II	P	26,2	22,6	29,0	29,0	22,8	7,0	1,29	1818,2	88,4	66,0	73,8	110,9	0,79	82	84	76	0,91	53	53	60	1,56	
	G	32,0	19,6	22,5	22,7	20,2	10,2	1,10	1881,4	22,5	300,6	279,3	26,9	41,51	84	78	89	1,14	50	66	34	0,51	
	K	27,0	25,4	27,3	27,4	25,7	7,7	1,08	1983,5	13,7	283,1	280,7	24,9	35,70	81	77	85	1,10	55	66	44	0,67	

GHI CHÚ: Trong cùng một tham biến của một vùng sự chênh lệch giữa các đài trạm không đáng kể. Điều phù hợp với ý nghĩa của việc xác lập vùng bằng chỉ số giống nhau lớn nhất.

5. Kết luận:

Từ việc đặt vấn đề và kết quả giải bài toán phân vùng khí hậu Việt Nam bằng phương pháp phân tích nhân tố và phân tích liên kết trên cơ sở các chỉ số giống nhau giữa các đài trạm trong không gian nhân tố trực giao chúng tôi rút ra những kết luận sau:

1. Bằng các phương pháp phân tích khách quan được áp dụng ở trên chúng ta có thể tiến hành phân vùng và phân loại khí hậu một cách khách quan và có kết quả bởi những cơ sở:

a) Phân tích nhân tố cho phép tìm ra các nhân tố đặc trưng đại diện cho tập hợp các tham biến của các yếu tố khí hậu lựa chọn. Không gian các nhân tố bé hơn không gian số liên ban đầu nhiều lần nhưng trong trong ý nghĩa toán thống kê lượng thông tin ban đầu không bị hao hụt. Đây là mục đích của phương pháp phân tích nhân tố.

b) Không gian các nhân tố trực giao cho phép tìm ra các chỉ số giống nhau giữa các đài trạm thông qua việc xác định khoảng cách giữa các đài trạm trong không gian nhân tố này, và không loại bỏ một nhân tố nào khi tính toán.

c) Ma trận các chỉ số giống nhau giữa các đài trạm bây giờ trở thành cơ sở chủ yếu cuối cùng để ta tập hợp các đài trạm khí hậu vào các vùng khác nhau một cách khách quan.

2. Tuy bài toán này là phân vùng khí hậu khách quan nhưng vẫn bị một vài yếu tố chủ quan khác chi phối, chẳng hạn việc lựa chọn các đài trạm và các tham biến cũng như việc căn cứ thêm vào yếu tố địa hình và biến trình năm của một vài yếu tố khí tượng quan trọng, v.v. trong khi vạch ranh giới giữa các vùng. Chính vì thế mà trong tương lai người ta còn tiến hành nghiên cứu thêm để hạn chế ngày càng nhiều những ảnh hưởng chủ quan, đồng thời nâng cao tính khách quan và mức chính xác của bài toán.

3. Các bài toán phân vùng và phân loại ở mục đích khác có thể áp dụng được phương pháp phân tích nhân tố và phân tích liên kết một cách thuận tiện và có kết quả, sau khi đã có sự cân nhắc thích đáng về phạm vi và sự phân bố các quan sát và sau khi đã kiểm tra mức đảm bảo của các tham biến lựa chọn.

4. Kết quả thu được ở trên tuy chưa hoàn toàn chính xác nhưng tương đối hợp lý. Kết quả đó cũng chỉ ra được sự tiện lợi và ít tốn kém về thời gian và kinh phí trong nghiên cứu phân vùng, đặc biệt khi bộ các chương trình tính đã được hoàn chỉnh. Để kiểm tra lại mức tin cậy của kết quả thu được phải sau 10, 15 năm nữa mới thực hiện, vì khi đó độ dài chuỗi số liệu mới đủ dài và mật độ đài trạm chọn vào phân tích mới tương đối hợp lý. Tất nhiên khi đó nên bổ sung thêm số liệu các đài trạm cần thiết của các nước láng giềng để tăng thêm tính chính xác và khách quan của bài toán.

TÀI LIỆU THAM KHẢO

1. W.BÖER: « Đề nghị về phân chia CHDC Đức thành các vùng khí hậu trên quy mô lớn thống nhất » (tiếng Đức) ZFM 17 (1975), 267-275.
2. W. JAHN và H. VAHLE: « Phân tích nhân tố và ứng dụng của nó » (tiếng Đức) Berlin 1975.
3. PHẠM NGỌC TOÀN và PHAN TẤT ĐẮC: « Phân vùng khí hậu Việt nam » (bản thảo đánh máy).
4. D. STEINER: « Phân tích nhân tố—Một phương pháp toán thống kê hiện đại cho các nhà địa lý trong việc phân vùng và phân loại khí hậu khách quan » (tiếng Đức) Geogr. helv. 20.
5. D. STEINER: « Phép xấp xỉ thống kê nhiều chiều cho việc phân vùng và phân loại khí hậu » (tiếng Anh). Tijdschrift van het Koninklijk nederlandsch aardrijkskundig genootschap Tweede Reeds. Diel LXXXII (1965)
6. R. Stelimacher: « Một vài phương pháp phân tích thống kê ứng dụng vào vấn đề phân loại khí hậu » (tiếng Đức),
Abh. d. Met.-Dienstes d. DDR Nr. 99, Band XIII (1971).
7. NGUYỄN TRỌNG HIỆU: « Một số nhận xét sơ lược về khí hậu biến đổi theo độ cao và thuộc loại khí hậu của vùng cao nước ta ». Thông báo về Hội nghị chuyên đề về Khí hậu tại Sapa VII/1976.