

MỘT PHƯƠNG PHÁP PHÂN VÙNG KHÁCH QUAN

PTS. Bùi Văn Đức

Vụ KHKT- Tổng cục Khí tượng Thủy văn

Sự biến đổi các điều kiện hình thành dòng chảy (ĐKHTDC) nói riêng và điều kiện địa lý tự nhiên nói chung không mang tính liên tục. Về tổng thể, trong một giới hạn không gian nào đó các điều kiện này biến đổi ít và có thể được coi như đồng nhất, song ra ngoài giới hạn này nó có sự khác biệt rất lớn. Nhiệm vụ của bài toán phân vùng là khám phá các quy luật phân bố theo không gian, biên giới của các vùng khác nhau và các quy luật chung trong từng không gian đồng nhất. Có nhiều phương pháp phân vùng, song đa số chúng thuộc loại phân vùng định tính. Kết quả của việc phân vùng theo các phương pháp này không tránh khỏi tính chủ quan và thiếu nhất quán.

Mục đích của bài báo là giới thiệu một số sơ đồ phân vùng khách quan, dựa trên cơ sở ứng dụng phương pháp xử lý thông tin trong phân tích thống kê. Sơ đồ phân vùng này đã được ứng dụng trong phân vùng theo điều kiện hình thành dòng chảy mùa hè thu của vùng Primorxki Krai thuộc Viễn Đông Liên Xô cũ [1,2] và phân vùng điều kiện hình thành lũ quét ở Việt Nam [3].

Chúng tôi hy vọng nó sẽ tìm thấy địa chỉ ứng dụng trong các bài toán tương tự trong nông nghiệp, lâm nghiệp nói riêng và phân vùng địa lý nói chung.

1. Cơ sở phương pháp

1.1- Sơ đồ tổng thể

Một yếu tố (Y) được quan tâm và là đối tượng để phân vùng được hình thành và phát triển trong sự ảnh hưởng của tập nhân tố (X) phức tạp. Bài toán phân vùng được bắt đầu từ việc phân tích các quy luật và nhân tố ảnh hưởng. Từ đó tiến hành sơ tuyển tập nhân tố ban đầu. Không gian lớn, được chia thành các tiểu vùng đơn vị (hoặc còn gọi là không gian đơn vị) mà trong đó yếu tố phân vùng và tập nhân tố ảnh hưởng (m chiều) tạo thành một vectơ hàng. Số liệu các không gian đơn vị (n) tạo thành ma trận $[X]_{n \times m+1}$. Ma trận số liệu ban đầu này thường được chuẩn hoá về dạng không thứ nguyên $[Z]_{n \times m+1}$.

$$Z_{ij} = \frac{X_{ij} - X_j}{\sigma_j} \quad (1)$$

trong đó: i-chỉ số không gian đơn vị; j-chỉ số nhân tố; σ_j -độ lệch chuẩn phương của nhân tố j; X_j -giá trị trung bình của nhân tố j.

Sơ đồ phân vùng bao gồm các bước sau:

a. Phân tích điều kiện hình thành yếu tố phân vùng

Bước 1. Phân tích và chọn các nhân tố điển hình (vectơ nhân tố)

Bước 2. Phân chia không gian cần nghiên cứu phân vùng thành các không gian đơn vị, mặc định đồng nhất theo các nhân tố và chỉ tiêu phân vùng.

b. Xử lý thông tin ban đầu

Bước 3. Nén thông tin và chuẩn hoá thông tin ban đầu bằng phương pháp phân tích nhân tố.

c. Phân vùng

Bước 4. Phân vùng bằng phương pháp phân tích "khoảng cách".

Bước 5. Lập bản đồ phân vùng từ các kết quả thống kê khách quan.

Các bước 3 và 4 là hai bài toán chuẩn trong thống kê [4], các bước còn lại phụ thuộc vào từng bài toán phân vùng cụ thể và nghệ thuật phân tích xử lý của chuyên gia. Do khối lượng bài viết bị khống chế, các phần sau chỉ trình bày tư tưởng chính của phương pháp, các vấn đề chi tiết sẽ được trình bày nếu có điều kiện.

1.2. Xác định nhân tố tổng hợp bằng phương pháp phân tích nhân tố

Phương pháp phân tích nhân tố là một trong những phương pháp nén và sàng lọc thông tin rất hiệu quả. Nó làm giảm đáng kể kích thước vector nhân tố mà vẫn giữ được các thông tin về mối liên hệ tồn tại giữa chúng và yếu tố. Phân tích nhân tố cho phép biến đổi tập thông tin ban đầu có mối liên hệ với nhau ($R_{ij} > 0$) thành tập thông tin độc lập tuyến tính có kích thước rút gọn và cho phép đánh giá mức độ tác động của các biến đến các nhân tố tổng hợp.

Mô hình phân tích nhân tố có dạng sau:

$$Z_{ij} = \sum_{l=1}^p a_{jl} F_{li} + U_j \quad (2)$$

trong đó: Z_{ij} -trị số đã được chuẩn hoá theo (1); a_{jl} -trọng số của nhân tố tổng hợp 1 của biến j ; F_{li} -nhân tố tổng hợp thứ l của không gian đơn vị thứ i ; U_j - sai số biến j ; p - số các nhân tố tổng hợp.

Việc biến đổi tập số liệu $[Z]$ thành tập nhân tố tổng hợp $[F]$ được thực hiện theo phương trình sau:

$$F_{li} = \frac{1}{\lambda_1} \sum_{j=1}^m a_{jl} Z_{ij} \quad (3)$$

trong đó: λ_1 -giá trị riêng của ma trận tương quan $[R]_{m \times m}$; a_{jl} -trọng số của biến j ban đầu.

$$R_{jk} = \frac{1}{(n-1)} \sum_{i=1}^n Z_{ij} Z_{ik} \quad (4)$$

Từ phương trình (2), sau một số biến đổi ta nhận được hệ phương trình dạng:

$$(R - \lambda E) S = 0 \quad (5)$$

Ở đây R -ma trận tương quan; λ -giá trị riêng ($l=1 \dots p$); E -ma trận đơn vị kích thước $(m \times m)$; S -ma trận các vector riêng kích thước $(m \times m)$.

Phương trình trên được giải theo thuật toán EIGEN [4], tính ra được các giá trị riêng λ_j . Ứng với mỗi giá trị riêng λ_j sẽ xác định được một vectơ riêng S_{jl} ($j=1..m$).

Trọng số của các nhân tố a_{jl} được tính bằng phương pháp lặp theo các phương trình sau:

$$\alpha_{jl}^{(1)} = S_{jl}$$

$$B_{jl}^{(1)} = \sum R_{ij}^1 \alpha_{jl}^{(1)} \quad (6)$$

$$\alpha_{jl}^{(2)} = B_{jl}^{(1)} / \max\{B_{jl}^{(1)}\} \quad (7)$$

$$B_{jl}^{(K)} = \sum R_{ij}^1 \alpha_{jl}^{(K)} \quad (8)$$

$$\alpha_{jl}^{(K+1)} = B_{jl}^{(K)} / \max\{B_{jl}^{(K)}\} \quad (9)$$

Trong đó: chỉ số trong dấu ngoặc ký hiệu bước lặp.

Giá trị ban đầu của α_{jl} được gán bằng S_{jl} và quá trình lặp tiến hành cho tới khi giá trị của trọng số giữa hai bước lặp liên tiếp nhỏ hơn sai số cho phép ε

$$|\alpha_{jl}^{(K)} - \alpha_{jl}^{(K+1)}| < \varepsilon_1 \quad (10)$$

$$a_{jl} = \alpha_{jl}^{(K+1)} \sqrt{\frac{\max\{B_{jl}^{(K)}\}}{\sum_{j=1}^p \alpha_{jl}^{(K)}}} \quad (11)$$

Ma trận tương quan dư được tính lại như sau:

$$R_{jk}^{1+1} = R_{jk}^1 - \sum_{i=1}^l a_{ij} a_{ik} \quad (12)$$

Việc xác định trọng số theo (11) sẽ dừng khi lũy tích phần trăm phương sai lớn hơn ngưỡng quy định $P_{K\%}$.

$$\frac{\sum_{i=1}^l \lambda_i}{m} \cdot 100 > P_K \quad (13)$$

Các trọng số a_{jl} xác định được theo phương pháp lặp trên cần được trực giao hoá thông qua phép xoay VARMAX [3]. Khi đó các trọng số của phép biến đổi (3) được coi là hoàn hảo, các nhân tố tổng hợp F_{jl} độc lập hoàn toàn với nhau (không có tương quan), có phân bố chuẩn và chứa đựng thông tin cô đọng của tập thông tin phân vùng khách quan.

1.3. Phân vùng bằng phương pháp phân tích khoảng cách

Phân tích khoảng cách là công cụ thống kê khách quan xác định mức độ đồng nhất (gần gũi) giữa các không gian đơn vị. Không gian đơn vị mang một ý nghĩa tương đối. Chúng có thể không bằng nhau, giá trị ở bước đầu là nhỏ nhất và được chia và gán theo phân tích định tính.

Phân tích khoảng cách có thể ứng dụng độc lập, bước biến đổi 1.2 có thể bỏ qua, nếu dám chắc rằng các nhân tố phân vùng là độc lập với nhau và độ chính xác cao. Tuy nhiên, cần xử lý vấn đề trọng số bằng phương pháp nào đấy. Trong quá trình nghiên cứu và tranh luận trong một số hội thảo, các chuyên gia đã đi đến sự công nhận sơ đồ ứng dụng tổng hợp 2 mô hình phân tích nhân tố và phân tích khoảng cách trong phân vùng khách quan mang một ý nghĩa khoa học và thực tiễn cao.

Công thức cơ bản của phân tích khoảng cách tương đối đơn giản, song thuật giải và kỹ thuật lập trình có phần hà đỗi chút.

$$DiK = \sqrt{\frac{1}{p} \sum_{l=1}^p (Fil - Fkl)^2} \quad (14)$$

Ở đây, D_{iK} -khoảng cách về tính đồng nhất giữa hai không gian đơn vị i và K ($i, K = 1..n$).

Sau mỗi loạt tính toán, ta có ma trận vuông $[D]_{n \times n}$ khoảng cách giữa các không gian đơn vị. Hai không gian có khoảng cách nhỏ nhất sẽ được nhập lại thành một không gian đơn vị mới. Các nhân tố phân vùng của hai không gian này sẽ được tính lại theo bình quân gia quyền về diện tích trước khi nhập. Mỗi bước lặp tổng số không gian đơn vị sẽ bị nhỏ đi một, và quá trình tính toán sẽ dừng khi khoảng cách nhỏ nhất giữa các không gian (vùng) lớn hơn ngưỡng cho phép [1].

2. Phân vùng tiềm năng và bài toán cảnh báo

Sau khi xác định được các nhân tố tổng hợp F , một ma trận số liệu mới được cấu thành bởi ma trận $[F]_{n \times p}$ và vector cột yếu tố phân vùng $(Y)_n$. Hướng ứng dụng này sẽ được minh họa trên ví dụ phân vùng nguy cơ xuất hiện lũ quét.

Thông tin phân vùng chứa trong ma trận số liệu ban đầu $[X]_{n \times m}$ dùng cho phân vùng lũ quét bao gồm: 1-tiềm năng mưa (lượng mưa ứng với tần suất 5%); 2-tiềm năng dòng chảy (lưu lượng ứng với tần suất 5%); 3-độ dốc lưu vực; 4-độ dốc lòng sông; 5-loại đất đá, xét theo khả năng thấm và phong hoá; 6-thảm phủ thực vật. Các thông tin này được phủ trên toàn bộ diện tích cần phân vùng (không gian tổng) ứng với 2670 ô (không gian đơn vị).

Các biến ban đầu được coi như chứa đựng đầy đủ các thông tin về khả năng hình thành lũ quét, song các thông tin chứa trong các biến còn mang tính trùng lặp (tương quan). Nếu sử dụng trực tiếp các biến này, kết quả phân vùng sẽ hạn chế về tính khách quan và độ chính xác. Vì vậy, đã tiến hành biến đổi ma trận $[X]_{n \times m}$ ($n=2670$ và $m=6$) thành ma trận nhân tố tổng hợp $[F]_{n \times p}$ ($p=3$). Lũy tích phần trăm phương sai đạt 81%.

Mỗi phần tử của vector cột yếu tố chứa số lần đã xuất hiện lũ quét trên ô - diện tích đơn vị. Chỉ số tổng hợp G_{ih} về khả năng xuất hiện nhóm h trong không gian i , được tính theo công thức tổng gia quyền.

$$G_{ih} = \sum_{l=1}^P b_l P_{il} \quad (15)$$

trong đó: b_l - trọng số của nhân tố tổng hợp F_l ; P_{il} - xác suất xuất hiện lũ quét theo một nhân tố tổng hợp F_l .

Để xác định xác suất P_{il} , nhân tố tổng hợp F_l được sắp xếp theo trình tự tăng dần hoặc giảm dần, mã số ô và giá trị các cột khác (bao gồm cả yếu tố) cần được sắp xếp theo. Tiếp theo cần thực hiện thống kê sau:

$$P_{il} = \frac{\sum_{i=nl}^{nl+1} K_i}{\sum_{i=1}^n K_i} \quad (16)$$

Ở đây tử số là số lần xuất hiện lũ quét trong nhóm vùng h .

Hệ số b_l chính là tỷ số giữa xác suất lớn nhất xuất hiện lũ quét của một nhóm nào đó theo nhân tố tổng hợp l với tổng xác suất này cho tất cả các nhân tố ($l=1..p$).

$$b_l = P_{\max l} / \sum_{l=1}^P P_{\max l} \quad (17)$$

Như vậy, nếu nhân tố F_l có ảnh hưởng lớn tới sự hình thành lũ quét, khi sắp xếp yếu tố này theo một trật tự, số lần xảy ra lũ quét có xu thế tập trung ở một nhóm h nào đấy và khi đó trọng số b_l sẽ lớn. Trong trường hợp phân bố tản mát của số lần xuất hiện lũ quét trong các nhóm, hệ số b_l sẽ nhỏ.

Số nhóm khả năng xuất hiện lũ quét (với chỉ số h) phụ thuộc vào tính chất và yêu cầu chi tiết của sự phân vùng. Trong phân vùng lũ quét, số nhóm được chia ra thành 4 tương ứng với 4 vùng [2].

- Vùng 1 nguy cơ xảy ra lũ quét cao, vùng này gồm 128 ô, chiếm 4,8%
- Vùng 2 nguy cơ xảy ra lũ quét khá, vùng này gồm 173 ô, chiếm 6,5%
- Vùng 3 nguy cơ xảy ra lũ quét trung bình, vùng này gồm 128 ô, chiếm 8,5%
- Vùng 4 nguy cơ xảy ra lũ quét nhỏ, vùng này gồm 2142 ô, chiếm 80,2%

NHẬN XÉT VÀ ĐỊNH HƯỚNG

Vấn đề phân vùng cần thiết trong nhiều lĩnh vực khác nhau, như kinh tế, nông nghiệp, lâm nghiệp, môi trường v.v. Phương pháp phân vùng này thực sự là một công cụ phân vùng mang tính khoa học và khách quan.

Bài toán phân vùng khách quan này đã được xây dựng thành công nghệ trên máy vi tính tại Trung tâm quốc gia dự báo KTTV và Viện KTTV, thuộc Tổng cục Khí tượng Thủy văn. Trong những năm qua công nghệ này đã được nghiên cứu và hoàn thiện. Nó đã được ứng dụng cho công trình phân vùng (1-theo điều kiện hình thành dòng chảy; 2- nguy cơ xuất hiện lũ quét).

Phương pháp phân vùng tiềm năng còn có thể ứng dụng trong bài toán dự báo KTTV, dự báo giá cả, nhiễm bẩn, suy thoái rừng, đất đai v.v... Những vấn đề chi tiết

tác giả bài viết xin được trao đổi trực tiếp (Tel: 824 4922 hoặc 851 4415, Bùi Văn Đức số 4 Đặng Thái Thân, Hoàn Kiếm, Hà Nội).

Tài liệu tham khảo

1. Bùi Văn Đức và Bephani N.F: Phân vùng lãnh thổ Primorxki Krai (Liên Xô cũ) theo điều kiện hình thành dòng chảy mùa hè thu. Tuyển tập các công trình nghiên cứu VINITI, No276 (Tiếng Nga).

2. Bùi Văn Đức: Nghiên cứu điều kiện hình thành và dự báo dòng chảy mùa hè thu vùng Primorxki Krai (Liên Xô cũ). Luận văn PTS. Ôdessa, Ukraine, 1985. (Tiếng Nga).

3. Bùi Văn Đức, Lương Tuấn Anh và một người khác: Phân vùng nguy cơ lũ quét bằng phương pháp phân tích nhân tố. Thuộc đề tài cấp Nhà nước "Nghiên cứu nguyên nhân hình thành và biện pháp phòng chống lũ quét", chủ nhiệm đề tài PTS. Cao Đăng Dư, thư ký đề tài PGS. Lê Bắc Huỳnh, cơ quan chủ trì Tổng cục Khí tượng Thủy văn, 1993-1995.

4. A.AFiFi và C. Eizen: Phân tích thống kê, hướng ứng dụng cho máy tính điện tử. NXB Mir, Maxcova, 1992, 480 tr (tiếng Nga).

Tin ngắn trong Ngành

Thông tư liên tịch về Methyl Bromide

Ngày 17-02-1998, GS.TS. Nguyễn Đức Ngữ, Tổng cục trưởng Tổng cục Khí tượng Thủy văn (KTTV), Chủ nhiệm Chương trình quốc gia về bảo vệ tầng ôzôn và GS. PTS. Ngô Thế Dân Thứ trưởng Bộ Nông nghiệp & Phát triển nông thôn (NN &PTNT) đã ký Thông tư liên tịch "Quy định về quản lý và kiểm soát sử dụng chất Methyl Bromide (MeBr)".

MeBr có công thức hoá học CH_3Br là chất bị kiểm soát bởi Nghị định thư Montreal về các chất làm suy giảm tầng ôzôn. MeBr là một trong những hoá chất có tiềm năng phá hủy ôzôn cao, được sử dụng rộng rãi trong nông nghiệp phục vụ xông đất, khử trùng, kiểm dịch thực vật. Theo số liệu của Chương trình Môi trường Liên hợp quốc (UNEP) hàng năm thế giới tiêu thụ và sử dụng từ 60.000-75.000 tấn MeBr.

Theo quy định tại Nghị định thư Montreal, MeBr phải hạn chế sản xuất từ năm 2001 và loại trừ hoàn toàn vào năm 2010 đối với các nước công nghiệp phát triển. Còn đối với những nước đang phát triển là một bên tham gia Nghị định thư Montreal như nước ta phải ngưng tiêu thụ và sử dụng MeBr vào năm 2002 trên cơ sở lượng tiêu thụ và sử dụng trung bình của những năm 1995-1998.

Ở nước ta, MeBr được sử dụng để kiểm dịch thực vật trong quá trình xử lý và bảo quản nông, lâm sản xuất khẩu (gạo, cà - phê, hạt điều, gỗ,...) với lượng tiêu thụ vào khoảng 300 tấn/năm.

Cùng với những qui định khác tại các văn bản " Pháp lệnh bảo vệ và kiểm dịch thực vật" (15- 02-1993), " Điều lệ về quản lý thuốc bảo vệ thực vật" ban hành kèm theo Nghị định 92-CP ngày 17-11-1993 của Chính phủ, Thông tư liên tịch đưa ra những qui định cụ thể cho các tổ chức, cá nhân có nhu cầu nhập khẩu và sử dụng MeBr; chế độ báo cáo hàng năm về tình hình sử dụng, tiêu thụ MeBr cho Bộ Nông nghiệp & Phát triển nông thôn và Chương trình quốc gia về bảo vệ tầng ôzôn.

Trong thông tư cũng nêu rõ thủ trưởng các cơ quan, đơn vị nhập khẩu và sử dụng MeBr có trách nhiệm quán triệt, hướng dẫn thực hiện thông tư này.

Văn phòng ôzôn