



## Original Article

# Integrating the Diversity Characteristics to Design a Quantitative Landscape Regionalization Framework: Case Study of Van Chan District, Yen Bai Province

Pham Minh Tam<sup>1,\*</sup>, Pham Hoang Hai<sup>2</sup>, Nguyen Cao Huan<sup>1</sup>, Pham Thu Thuy<sup>3</sup>

<sup>1</sup>*VNU University of Science, Hanoi, 34 Nguyen Trai, Hanoi, Vietnam*

<sup>2</sup>*Institute of Geography, Vietnam Academy of Science and Technology (VAST),  
18 Hoang Quoc Viet, Hanoi, Vietnam*

<sup>3</sup>*Science and Technology Department, Vietnam National University, Hanoi, 144 Xuan Thuy, Hanoi, Vietnam*

Received 08 August 2019

Revised 18 November 2019; Accepted 08 July 2020

**Abstract:** Landscape regionalization plays an important role in delineating the heterogeneous characteristics of territory, and provide the spatial fundamental data for natural resource planning and environmental protection activities. The integrating of the diversity indices (landscape metrics) is expressed the change of landscape structure by the richness and evenness of land-use objectives. In this study, a quantitative landscape regionalization framework is designed from 03 group (attribute factor, driving factor, and diversity factor) of basic landscape unit. By using k-means clustering, the study is classified into 06 sub-regions of 68 watersheds in the administration boundary of Van Chan district, Yen Bai province. With the comparison of region numbers in statistical and practical dimensions, the optimal results are edited and determined 15 sub-regions for uncertainty reduction of landscape regionalization.

**Keywords:** regionalization, quantitative modeling, landscape, diversity, cluster analysis, Van Chan.

\* Corresponding author.

E-mail address: [phamminhtam1989@gmail.com](mailto:phamminhtam1989@gmail.com)

<https://doi.org/10.25073/2588-1094/vnuees.4432>

# Tích hợp các đặc trưng đa dạng trong xây dựng mô hình định lượng về phân vùng cảnh quan: Trường hợp nghiên cứu tại huyện Văn Chấn, tỉnh Yên Bái

Phạm Minh Tâm<sup>1,\*</sup>, Phạm Hoàng Hải<sup>2</sup>, Nguyễn Cao Huân<sup>1</sup>, Phạm Thu Thủy<sup>3</sup>

<sup>1</sup>Trường Đại học Khoa học Tự nhiên, ĐHQGHN, 334 Nguyễn Trãi, Hà Nội, Việt Nam

<sup>2</sup>Viện Địa lý, Viện Hàn lâm Khoa học và Công nghệ Việt Nam (VAST),  
18 Hoàng Quốc Việt, Hà Nội, Việt Nam

<sup>3</sup>Ban Khoa học và Công nghệ, Đại học Quốc gia Hà Nội, 144 Xuân Thủy, Hà Nội, Việt Nam

Nhận ngày 08 tháng 8 năm 2019

Chỉnh sửa ngày 18 tháng 11 năm 2019; Chấp nhận đăng ngày 08 tháng 7 năm 2020

**Tóm tắt:** Phân vùng cảnh quan đóng vai trò quan trọng trong phân định đặc trưng bất đồng nhất của lãnh thổ, cung cấp nền tảng không gian cơ bản cho hoạt động quy hoạch sử dụng tài nguyên thiên nhiên và bảo vệ môi trường. Quá trình tích hợp các độ đo độ đa dạng giúp tái hiện những thay đổi về cấu trúc cảnh quan trên phương diện đặc trưng tính phong phú và đồng đều của lớp phủ sử dụng đất. Trong nghiên cứu này, mô hình định lượng trong phân vùng được xây dựng từ quá trình tích hợp ba nhóm thông tin (nhóm các yếu tố thuộc tính, nhóm các yếu tố động lực và nhóm các yếu tố đa dạng) của các đơn vị cảnh quan cơ sở. Trên cơ sở phân tích phân cụm k-mean, nghiên cứu tiến hành thực hiện phân thành 06 tiểu vùng từ dữ liệu định lượng của 68 tiểu lưu vực trong lãnh thổ hành chính huyện Văn Chấn, tỉnh Yên Bái. Từ quá trình so sánh số lượng tiểu vùng về mặt thống kê và thực tiễn, nghiên cứu đã tiến hành hiệu chỉnh và xác định 15 tiểu vùng tối ưu, giám đàng kê tính chủ quan trong quá trình phân vùng cảnh quan.

**Từ khóa:** phân vùng, mô hình định lượng, cảnh quan, đa dạng, phân tích cụm, Văn Chấn.

## 1. Mở đầu

Phân vùng là một tiến trình quan trọng trong phân tích dữ liệu địa lý đa thời gian [1], nhằm mô tả các đặc trưng bất đồng nhất về mặt không gian cũng như cơ chế phát sinh của chúng [2]. Ở quy mô cảnh quan, hầu hết các hiện tượng sinh thái và môi trường đều có thể phân định thành các khu vực đồng nhất tương đối về thuộc tính hay mối quan hệ không gian [3], trở thành nền tảng cơ bản cho quá trình giám sát, đánh giá, kiểm kê và quản lý tài nguyên môi trường [4]. Từ sự phát triển của phương thức và sự đa dạng

về mục tiêu phân vùng, hàng loạt ứng dụng tích hợp khả năng phân tích không gian đã được thực hiện như: bảo tồn đa dạng sinh học [5], kiểm soát chất lượng nước [6], phân tích biến đổi cảnh quan, hay đánh giá các giá trị chức năng phục vụ phát triển kinh tế xã hội [7]. Trong đó, các chỉ tiêu phân vùng cảnh quan được lựa chọn sao cho phù hợp với mục tiêu và đối tượng nghiên cứu [8], nhưng trong thực tiễn cần một số lượng dữ liệu không gian lớn và được giải đoán chi tiết [9]. Những dữ liệu thành phần này có thể gồm đặc trưng vùng sinh thái [10], thông tin môi trường [11], địa chất [12], khí hậu [13], hay một giả định

\* Tác giả liên hệ.

Địa chỉ email: phamminhtam1989@gmail.com

<https://doi.org/10.25073/2588-1094/vnuees.4432>

về mối liên kết giữa động lực thúc đẩy quá trình sinh thái và lớp phủ sử dụng đất [14]. Nhìn chung, các chỉ tiêu phân vùng này có thể phân loại thành hai nội dung chính là: thuộc tính (attributes) và yếu tố động lực (driving factors) [2]. Song, trong bối cảnh phát triển kinh tế - xã hội hiện nay, cảnh quan dưới góc độ tiếp cận từ lớp phủ sử dụng đất chứng kiến tốc độ thay đổi hết sức nhanh chóng, ảnh hưởng tới quá trình phân vùng phục vụ tổ chức lãnh thổ [15]. Điều này thúc đẩy quá trình tích hợp các đặc trưng phản ánh sự đa dạng cấu trúc cảnh quan vào tiến trình phân vùng; đặc biệt là trên phương diện phong phú và đồng đều [16]. Từ đó, cách tiếp cận hệ thống trong phân vùng cảnh quan nên được xây dựng từ tổng hòa ba nhóm chỉ tiêu trên (thuộc tính, động lực và tính đa dạng); trong khi, vẫn đảm bảo khả năng kết hợp nhiều lớp thông tin có độ tin cậy cao cho các mục đích ứng dụng khác nhau.

Về phương pháp, cách tiếp cận định lượng trong phân vùng cảnh quan trở nên phổ biến trong những năm gần đây, đem lại những lợi thế: khách quan về kết quả, dễ thực hiện và khả năng tinh chỉnh cao [17]. Một số phương pháp có thể kể tới như phân tích thành phần chính [18], phân loại và hồi quy [19], mô hình hóa chỉ số đồng nhất [20] hay lý thuyết tập mờ [21]. Ngoài ra, từ ý tưởng phân loại theo cấu trúc cảnh quan [22], các chỉ số cảnh quan bắt đầu tham gia vào bài toán phân vùng cảnh quan dưới dạng một biến thông tin quan trọng trong quá trình phân tích cụm đa biến (multivariate cluster analysis). Điều này mở ra khả năng “tương thích” với nhiều kiểu loại dữ liệu cũng như kiểm chứng tính chính xác ở nhiều quy mô không gian khác nhau [23]. Tuy nhiên, do mục đích và đối tượng hướng tới khác nhau, số lượng phân vùng trong cảnh quan vì vậy cũng biến thiên đa dạng. Do đó, sự thiếu sót trong quá trình kiểm chứng kết quả phân vùng cảnh quan cũng trở nên ít được chú ý.

Đối với khu vực miền núi, cách tiếp cận dựa vào lưu vực cho phép xác định các đơn vị không gian cơ bản trong tiến trình phân vùng cảnh quan hiệu quả hơn [2]. Mục tiêu của nghiên cứu hướng tới phát triển một mô hình đánh giá định lượng có khả năng phân vùng cảnh quan một cách hiệu

quả, toàn diện và khách quan hơn so với phương thức truyền thống. Nghiên cứu hướng tới: i) cải thiện khả năng hệ thống và toàn diện trên cơ sở tích hợp ba nhóm thông tin của cảnh quan; ii) tận dụng tính hiệu quả khi coi các tiểu lưu vực như các đơn vị không gian cơ bản; iii) tiến hành kiểm chứng kết quả nhằm nâng cao tính khách quan. Tuy nhiên, do giới hạn về khả năng thu thập dữ liệu, nghiên cứu chỉ tiến hành phân vùng cảnh quan thử nghiệm trên cơ sở ba nhóm chỉ tiêu cho lãnh thổ huyện Văn Chấn, tỉnh Yên Bái.

## 2. Đối tượng và phương pháp nghiên cứu

### 2.1. Mô hình lý thuyết

Khung lý thuyết về phân vùng cảnh quan được thực hiện theo các giai đoạn chính sau: i) Xác định các đối tượng và mục tiêu trong hoạt động phân vùng, từ đó hướng tới xác định tiềm năng ứng dụng vào cảnh quan; ii) Lựa chọn các biến đại diện (nhằm định hướng giải quyết trên quan điểm cảnh quan) cho ba nhóm yếu tố thuộc tính, động lực và tính đa dạng; iii) Xác định sự phân hóa không gian của cảnh quan trên cơ sở các khoanh vi đơn vị không gian cơ bản (ô lưới, lưu vực,...) và gán các giá trị cho các yếu tố tương ứng; iv) Kiểm tra mối quan hệ giữa các yếu tố thông qua phân tích thành phần chính PCA; v) Phân loại các đơn vị này thành các nhóm dựa trên quá trình phân tích cluster; vi) Đánh giá tính chính xác và điều chỉnh các kết quả sao cho tối ưu (Hình 1).

### 2.2. Đối tượng nghiên cứu

Khu vực nghiên cứu được lựa chọn là lãnh thổ huyện Văn Chấn, tỉnh Yên Bái. Đây là khu vực có diện tích tự nhiên lên tới 120.758,5 ha; trải dài trên các tọa độ từ 20°20'-21°45' vĩ độ Bắc, và 104°20'-104°53' kinh độ Đông. Nơi đây tiếp giáp huyện Mù Cang Chải ở phía bắc, giáp huyện Văn Yên và Trấn Yên ở phía Đông, phía Tây giáp huyện Trạm Tấu và phía Nam giáp tỉnh Sơn La. Ngoài ra, do nằm tại điểm kết thúc của dãy Hoàng Liên Sơn, cảnh quan tại khu vực có đặc trưng đa dạng khá cao. Lãnh thổ huyện được chia thành 68 tiểu lưu vực - đơn vị cơ bản (xem

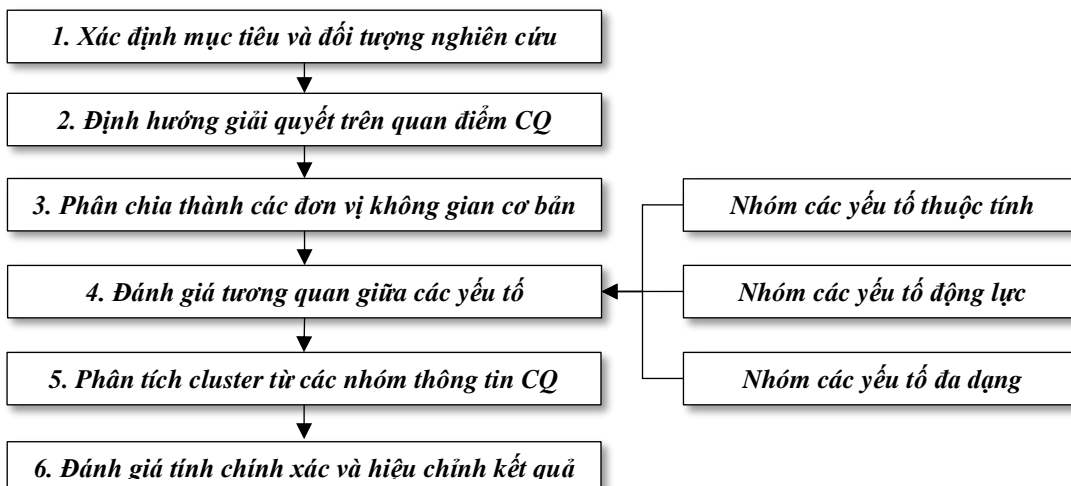
Hình 2), đóng vai trò là cơ sở khoa học quan trọng trong đánh giá tài nguyên và quá trình thủy văn diễn ra trong không gian [24], trở thành tiền đề cho nhiều nghiên cứu về phân vùng lãnh thổ khu vực miền núi [25]. Quá trình thống kê các thông tin cảnh quan từ thông số của các đơn vị này được thực hiện trên ba phương diện:

i) *Nhóm các yếu tố thuộc tính*: gồm các đặc trưng về *dung tích bề mặt (A1)* phản ánh khả năng chứa đựng lượng nước thường xuyên trong một lưu vực/đầu nguồn và được tính bằng tích của độ sâu thung lũng (*Valley Depth*) và diện tích bề mặt; cùng đặc trưng *phân cắt sâu (A2)* và *phân cắt ngang (A3)* của địa hình. Đây là nhóm các yếu tố thể hiện cấu trúc không gian của lãnh thổ.

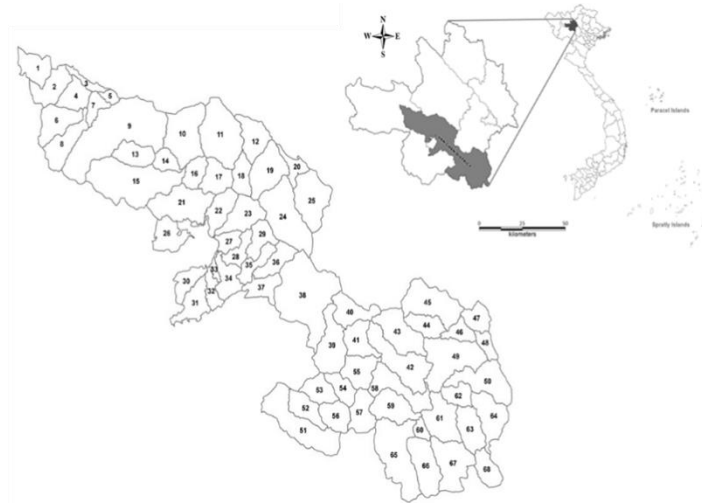
ii) *Nhóm các yếu tố động lực*: gồm 05 biến môi trường được sử dụng trong nghiên cứu này. Các biến này gồm nhiệt độ trung bình năm (F1), lượng mưa trung bình năm (F2), độ cao địa hình (F3) cùng các chỉ số viễn thám của ảnh viễn thám như chỉ số ẩm ướt của địa hình TWI (F4) và chỉ số thực vật có hiệu chỉnh ảnh hưởng của đất SAVI (F5); tương ứng với ý nghĩa mang tính “đại diện” mô tả khí hậu, điều kiện địa hình, đặc

trung địa chất, thổ nhưỡng và điều kiện thực vật trong lãnh thổ. Các dữ liệu không gian sau đó được chuyển sang dạng raster và sử dụng phương thức lấy lại mẫu (resampling) trong môi trường GIS để đưa về cùng một độ phân giải nhất quán là 1:100.000. Các giá trị này sau đó được thống kê dưới dạng giá trị trung bình cho từng biến. Đây là nhóm nhân tố thúc đẩy những biến động về năng lượng của cảnh quan.

iii) *Nhóm các yếu tố đa dạng*: Nghiên cứu sử dụng hai biến đặc trưng cho độ đa dạng là *chỉ số đồng đều Shannon-SHEI (D1)* và *chỉ số đa dạng Shannon-SHDI (D2)*. Độ đo SHEI thể hiện mức độ phân bố đều tối đa của các đối tượng bên trong cảnh quan. Ngoài ra, độ đo SHDI biểu thị mức độ phức tạp về cấu trúc và tỷ lệ thuận với mức độ đa dạng cảnh quan. Một bên thể hiện sự phân tán, trong khi, bên còn lại thể hiện sự chiếm ưu thế. Các chỉ số này được tính toán dựa trên công cụ Patch Analyst trong môi trường nền ArcGIS 10.4. Kết quả tính toán được tính từ dữ liệu hiện trạng lớp phủ/sử dụng đất năm 2017. Đây là nhóm yếu tố thể hiện đặc trưng và xu thế thay đổi về mặt cấu trúc của cảnh quan.



Hình 1. Mô hình lý thuyết về tiếp cận định lượng trong phân vùng cảnh quan.



Hình 2. Sơ đồ vị trí và phân bố không gian của 68 tiểu lưu vực khu vực huyện Văn Chấn, tỉnh Yên Bái (trong đó: 1, ... 68 là ký hiệu các tiểu lưu vực)

Các thông tin của chỉ tiêu A1, A2, A3, F3, F4 được tách chiết từ dữ liệu USGS DEM (độ phân giải 30m). Trong khi, thông tin của chỉ tiêu F1, F2 được thu thập từ WorldClim tại địa chỉ: <https://www.worldclim.org/>. Chỉ tiêu F5 được

tính toán từ dữ liệu ảnh Landsat 8 OLI (độ phân giải 30m) năm 2017. Dữ liệu sử dụng để tiến hành phân vùng cảnh quan được mô tả chi tiết trong bảng 1.

Bảng 1. Thống kê dữ liệu của ba nhóm yếu tố phân vùng theo các lưu vực sông huyện Văn Chấn, Yên Bái

| Lưu vực | Nhóm yếu tố thuộc tính |                         |                          | Nhóm yếu tố động lực |           |           |        |        | Nhóm yếu tố đa dạng |        | Số quan sát |
|---------|------------------------|-------------------------|--------------------------|----------------------|-----------|-----------|--------|--------|---------------------|--------|-------------|
|         | A1 (km <sup>3</sup> )  | A2 (m/km <sup>2</sup> ) | A3 (km/km <sup>2</sup> ) | F1 (°C)              | F2 (mm)   | F3 (m)    | F4     | F5     | D1                  | D2     |             |
| 1       | 0,693                  | 450,576                 | 1,4999                   | 2379,9696            | 1660,8767 | 787,2926  | 2,9105 | 5,0308 | 0,7787              | 1,2532 | 84          |
| 2       | 1,171                  | 634,280                 | 1,0346                   | 2181,2628            | 1763,8218 | 1107,3759 | 2,4827 | 5,5504 | 0,7869              | 1,4100 | 99          |
| 3       | 0,681                  | 649,799                 | 0,0147                   | 2386,2813            | 1646,3894 | 728,4066  | 2,1636 | 5,5205 | 0,3966              | 0,5498 | 30          |
| 4       | 0,536                  | 477,837                 | 1,3273                   | 2330,7936            | 1666,9605 | 861,3852  | 2,6256 | 5,3572 | 0,9442              | 1,5196 | 76          |
| 5       | 0,382                  | 874,877                 | 0,7112                   | 2332,2519            | 1671,0667 | 803,7518  | 2,0409 | 5,6440 | 0,8427              | 0,5841 | 19          |
| 6       | 1,469                  | 748,192                 | 1,0202                   | 2014,1343            | 1844,5214 | 1350,8548 | 2,2190 | 6,1443 | 0,3862              | 0,6216 | 111         |
| 7       | 0,539                  | 566,314                 | 1,2557                   | 2396,1195            | 1631,9478 | 742,9189  | 2,4377 | 5,5382 | 0,5786              | 0,9313 | 43          |
| 8       | 1,528                  | 764,392                 | 1,1025                   | 1986,9741            | 1845,1938 | 1424,0697 | 2,1500 | 5,9140 | 0,3466              | 0,5578 | 110         |
| 9       | 4,791                  | 580,706                 | 1,3471                   | 2295,0198            | 1691,1489 | 911,6673  | 2,5654 | 5,6616 | 0,8486              | 1,3658 | 444         |
| 10      | 1,685                  | 507,017                 | 1,3053                   | 2409,6502            | 1629,8485 | 703,7478  | 2,4715 | 5,7073 | 0,7700              | 1,2393 | 188         |
| ...     | ...                    | ...                     | ...                      | ...                  | ...       | ...       | ...    | ...    | ...                 | ...    | ...         |
| 60      | 0,208                  | 371,484                 | 1,7389                   | 2707,3686            | 1608,8714 | 200,6718  | 3,1082 | 5,4478 | 0,8051              | 1,2958 | 25          |
| 61      | 0,599                  | 270,524                 | 1,4825                   | 2705,9148            | 1620,4076 | 177,7600  | 2,6369 | 5,4068 | 0,9923              | 1,3757 | 132         |
| 62      | 0,305                  | 322,042                 | 1,4491                   | 2691,5647            | 1637,6320 | 184,6084  | 2,5317 | 5,7423 | 0,8924              | 1,2371 | 58          |
| 63      | 0,601                  | 352,957                 | 1,3356                   | 2673,2241            | 1604,1854 | 228,5204  | 2,6610 | 5,6395 | 0,8802              | 1,2202 | 100         |
| 64      | 1,027                  | 462,934                 | 1,1359                   | 2609,0845            | 1593,6740 | 334,0019  | 2,3431 | 6,1880 | 0,6825              | 1,2228 | 124         |
| 65      | 2,874                  | 560,095                 | 1,5112                   | 2520,0966            | 1550,8967 | 503,5501  | 2,6408 | 5,6778 | 0,7993              | 1,2864 | 241         |
| 66      | 1,009                  | 421,809                 | 1,1199                   | 2571,8571            | 1529,3003 | 430,9272  | 2,4189 | 5,8255 | 0,6858              | 1,1038 | 134         |
| 67      | 1,095                  | 398,506                 | 1,2996                   | 2623,1404            | 1554,4833 | 326,3816  | 2,4444 | 5,9453 | 0,9485              | 1,5266 | 158         |
| 68      | 0,437                  | 359,768                 | 1,2326                   | 2556,3064            | 1535,2950 | 427,6967  | 2,4708 | 5,7261 | 0,8247              | 1,3274 | 70          |

### 2.3. Phương pháp nghiên cứu

Nghiên cứu tiên hành đánh giá trên cơ sở phương pháp phân tích nhóm cụm k-mean (k-mean clustering) – một kỹ thuật tìm kiếm sử dụng phân tích thống kê không kiểm chứng nhằm khám phá và xác lập cấu trúc đặc trưng của dữ liệu [26]. Trong đó, tham số k (số lượng cụm) có tính thiếu chắc chắn cao nên được lựa chọn trong một khoảng giá trị nhất định (2-10). Và giá trị phân cụm tối ưu có thể được lựa chọn thông qua các tiêu chí đánh giá với giá trị k đạt được là lớn nhất. Các phương pháp nghiên cứu được thực hiện gồm:

i) *Chuẩn hóa dữ liệu*: Đây là một tiến trình quan trọng trong phân tích nhóm cụm cluster. Theo Kaufman & Rousseeuw (1990), thuật toán CLARA (Clustering for LARge Applications) được thiết kế cho phân vùng với bộ dữ liệu lớn:

$$z_{if} = \frac{x_{if} - m_f}{s_f} \quad (\text{Eq. 1.})$$

Trong đó:  $z_{if}$  là giá trị chuẩn hóa cho quan sát  $i$  của biến  $f$ ,  $x_{if}$  là giá trị ban đầu để quan sát  $i$  của biến  $f$ ,  $m_f$  là giá trị trung bình của biến  $f$  và  $s_f$  là giá trị đo độ phân tán cho biến  $f$ . Đồng thời, độ phân tán của biến  $f$  với số lượng quan sát  $n$  được xác định thông qua công thức:

$$s_f = \frac{1}{n} \left\{ |x_{1f} - m_f| + |x_{2f} - m_f| + \dots + |x_{nf} - m_f| \right\} \quad (\text{Eq. 2.})$$

Trong bối cảnh số lượng biến quan sát lớn, quá trình xác định sự phân tán này cho phép quá trình chuẩn hóa độ lệch chuẩn hiệu quả hơn các phương pháp khác.

ii) *Phân tích đa biến dựa trên phương pháp PCA*: Phân tích thành phần chính (*Principal Component Analysis*) được sử dụng như một phương thức để kiểm tra mối quan hệ giữa các yếu tố tham gia phân vùng cảnh quan ngay sau khi dữ liệu từng biến số được chuẩn hóa. Khi đó, tác động biên và nhận diện mức độ ảnh hưởng tới kết quả nghiên cứu của chúng được thống kê dựa trên phần mềm Xlstats.

iii) *Phương pháp phân cụm k-means và lựa chọn giá trị tối ưu “Elbow”*: Phân loại dữ liệu không phân cấp k-mean được sử dụng khá phổ biến trong các nghiên cứu tài nguyên và môi trường. Với tiến trình phân cụm, cần phải tính toán nhằm xác định sự tách biệt giữa các đối tượng. Nếu như trước đây, quá trình này được định lượng thông qua các phép tính khoảng cách Euclide với các biến xác định theo tiêu chuẩn. Tuy nhiên, trong nghiên cứu này, số lượng cụm tối ưu được xác định thông qua phương pháp Elbow và chỉ số phù hợp (*Goodness of fit*). Trong đó, phương pháp Elbow sử dụng đánh giá tỷ lệ phần trăm phương sai cho mục đích giải thích giá trị tối ưu số cụm. Giá trị ấy được lựa chọn sao cho gia tăng giá trị mà không làm quá trình mô hình hóa dữ liệu trở nên tốt hơn. Chính xác hơn, nếu một giá trị tỷ lệ phương sai với số lượng cụm tương ứng gia tăng tới một giá trị tới hạn, mức độ gia tăng biên sẽ giảm, tạo thành một góc trong đồ thị - gọi là “khủy tay”. Thời điểm này, giá trị tối ưu k được lựa chọn. Ngoài ra, chỉ số phù hợp cũng được sử dụng như một phương thức để lựa chọn giá trị số lượng cụm [27]:

$$\text{Good of fit} = (B/(k-1))/(W/(n-k)) \quad (\text{Eq. 3.})$$

Trong đó:  $k$ : số lượng nhóm;  $n$ : số lượng mẫu;  $B$ : phân tán giữa các nhóm;  $W$ : phân tán trong một nhóm.

*Quá trình đối sánh và biên chỉnh kết quả theo mục tiêu*: được thực hiện trên cơ sở phân tích lặp đi lặp lại nhiều lần với số lượng nhóm tăng dần từ 2 đến 10. Số lượng nhóm  $k$  có chỉ số kiểm định cao nhất có thể được coi là phù hợp một cách thuần túy ở khía cạnh thống kê. Trên thực tế, không hề có cơ sở lý luận nào đủ để xác định một khoảng tối ưu cho số lượng nhóm. Điều này chỉ có thể xác định dựa trên nhu cầu quản lý cảnh quan của lãnh thổ, nên số lượng nhóm trong trường hợp này thường không quá lớn (nhỏ hơn 1/3 tổng số). Tuy nhiên, đặc thù của quy hoạch không gian còn cần xem xét đến chức năng của lãnh thổ nên số lượng nhóm không thể quá nhỏ; cũng như xem xét tới nhu cầu và mục tiêu phân vùng của lãnh thổ [2].

### 3. Kết quả nghiên cứu

Về mặt quan điểm phân vùng, phân vùng là một tiến trình quan trọng trong phân tích dữ liệu địa lý đa thời gian [1], nhằm mô tả các đặc trưng bất đồng nhất về mặt không gian cũng như cơ chế phát sinh của chúng [2]. Ở quy mô cảnh quan, hầu hết các hiện tượng sinh thái và môi trường đều có thể phân định thành các khu vực đồng nhất tương đối về thuộc tính hay mối quan hệ không gian [3], trở thành nền tảng cơ bản cho quá trình giám sát, đánh giá, kiểm kê và quản lý tài nguyên môi trường [4]. Như vậy, có thể hiểu: “phân vùng cảnh quan đóng vai trò như một tiến trình phân chia các đơn vị lãnh thổ cơ bản có đặc trưng đồng nhất tương đối đồng nhất về thuộc tính hay mối quan hệ không gian từ những yếu tố bất đồng nhất về mặt không gian cũng như cơ chế phát sinh của chúng (trong thực tiễn).

Nguyên tắc phân vùng dựa trên phân tích định lượng phải đảm bảo: i) giá trị định lượng phải mang tính đại diện cho các đơn vị/khoanh vi cơ sở; ii) phân vùng phải đảm bảo tính tổng hợp của kết quả từ các thông tin độc lập nhằm

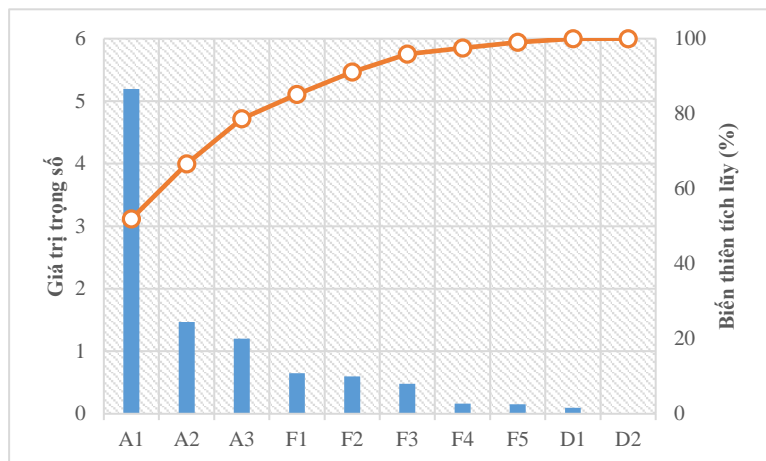
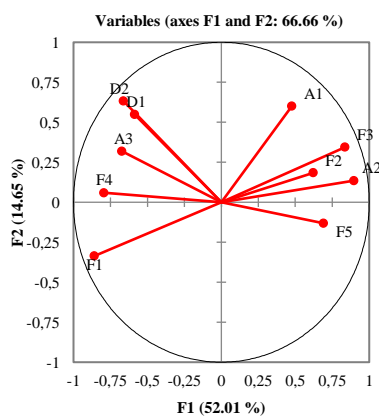
đối sánh với thực tiễn; iii) đảm bảo tính thống nhất về ý nghĩa kết quả; iv) kết quả cuối cùng phải được “quyết định” nhằm hài hòa các yêu cầu của nghiên cứu với kiểm định mô hình thống kê.

Từ kết quả chuẩn hóa dữ liệu đầu vào theo công thức của Kaufman & Rousseeuw (1990), thử nghiệm KMO (Kaiser-Meyer-Olkin test) được sử dụng để đánh giá mức độ đầy đủ của dữ liệu của từng biến và toàn bộ mô hình. Theo đó, kết quả thống kê tỷ lệ phương sai giữa các biến tạo nên phương sai chung, nếu giá trị càng thấp thì dữ liệu phải được tiến hành phân tích nhân tố nhằm giảm số lượng biến trong mô hình. Tuy nhiên, với giá trị KMO = 0.739, quá trình lấy mẫu cho phân vùng cảnh quan được coi là tương đối đầy đủ (Bảng 2).

Quá trình phân tích thành phần chính PCA đã cho thấy các yếu tố thuộc tính đã giải thích 76,68% tổng phương sai của các biến tham gia phân vùng cảnh quan (p-value < 0.0001). Tương quan so sánh cặp giữa các nhóm yếu tố được giải thích cụ thể thông qua biểu đồ vector tại Hình 3.

Bảng 2. Kết quả thống kê thử nghiệm KMO đối với dữ liệu sau chuẩn hóa

|            | A1    | A2    | A3    | F1    | F2    | F3    | F4    | F5    | D1    | D2    | Final |
|------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| <b>KMO</b> | 0.763 | 0.927 | 0.857 | 0.678 | 0.863 | 0.671 | 0.769 | 0.698 | 0.589 | 0.679 | 0.739 |

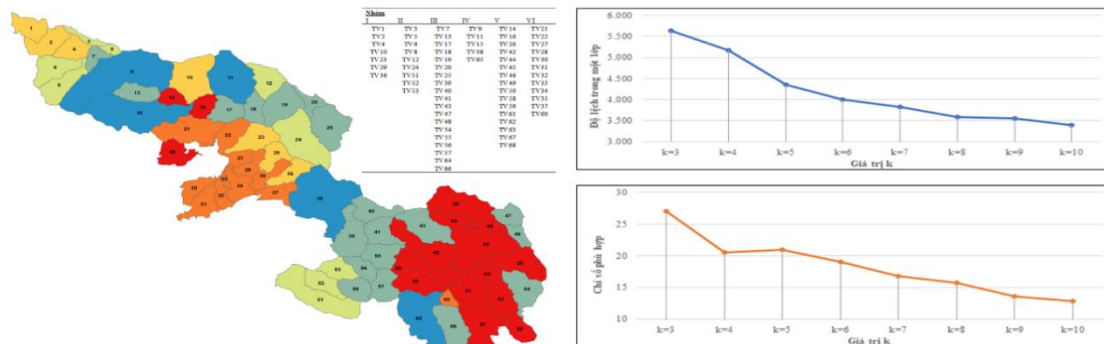


Hình 3. Tương quan cặp, giá trị trọng số và biến thiên tích lũy (%) của các biến trong phân vùng cảnh quan.

Kết quả phân tích PCA cho thấy: i) chỉ tiêu *dung tích bề mặt (A1)*, *phân cắt sâu (A2)*, *lượng mưa trung bình năm (F2)* và *độ cao địa hình (F3)* có tương quan tích cực với nhau; ii) các chỉ tiêu này có tương quan nghịch đối với *hiệu chỉnh ảnh hưởng của đất SAVI (F5)*. Dung tích bề mặt là yếu tố quan trọng nhất, giải thích tới 66.03% phương sai của nhóm chỉ tiêu thuộc tính. Điều này cho thấy mức độ tương đồng về dung tích sẽ ảnh hưởng rất lớn tới phạm vi và khả năng hình thành của các tiểu vùng cảnh quan. Trên cơ sở sử dụng phương pháp phân cụm k-means, quá trình lựa chọn giá trị nhóm tối ưu được thực hiện qua các phương diện sau: i) Về mặt thống kê, giá trị k có thể xác định thông qua phương pháp Elbow và giá trị chỉ số phù hợp của Dale & Fortin (2014). Các giá trị này được biểu diễn chi tiết trong Hình 4. Theo đó, nếu dựa trên chỉ số phù hợp, k=4 là giá trị tối ưu thì phương pháp Elbow lựa chọn giá trị k=6 là tối ưu.

ii) Về mặt không gian và quan điểm phân vùng, số lượng các tiểu vùng cảnh quan sẽ tồn tại những khác biệt mà kết quả thống kê không thể đáp ứng được. Để các dữ liệu này phân tích trở nên có ý nghĩa, lãnh thổ thường được phân chia thành các vùng có đặc trưng đồng nhất hay có khả năng bị chi phối bởi một quá trình sinh thái

chiếm ưu thế. Thông thường, quá trình này được xác định qua hai cách: i) nhóm các đối tượng có vị trí liền kề có giá trị tương tự nhau thông qua quá trình phân cụm không gian (spatial clustering), ii) phân chia các khu vực dựa trên mức độ khác nhau làm căn cứ để phân định ranh giới (boundary detection) [27]. Nếu về mặt lý thuyết, hai cách tiếp cận này đem lại kết quả phân vùng giống nhau. Nhưng phần lớn các kết quả không đạt được hiệu quả cao khi giá trị của các biến trong các đơn vị cơ sở không phân bố đồng đều trong không gian. Quá trình sử dụng phương pháp phân cụm k-means trong phân vùng cảnh quan hướng tới ưu tiên xác định giá trị số lượng cụm cần thiết để đạt được sự giống nhau về đặc trưng với giá trị sai lệch thấp nhất; cũng như giảm thiểu sự chủ quan trong hoạt động ra quyết định. Nhưng trên thực tiễn, cách tiếp cận này gặp phải một số khó khăn: i) các vị trí lấy mẫu ngẫu nhiên đôi khi không thể mang tính đại diện cho đặc trưng của một đơn vị cảnh quan cơ sở; ii) đòi hỏi kiến thức hoặc thông tin độc lập về dữ liệu sinh thái trên nhiều quy mô không gian và thời gian khác nhau nhằm tiên đoán được kết quả, làm tiền đề để đối sánh với thực tiễn; iii) giá trị kiểm định chỉ có hiệu quả về mặt thống kê, tính thống nhất về ý nghĩa của các kết quả kiểm định thường thấp; iv) mục tiêu phân vùng càng đạt được hiệu quả cao khi mức độ phân hóa càng chi tiết trong khi vẫn đảm bảo sai số thấp nhất; v) các tiểu vùng thường phải nằm lân cận nhau. Điều này dẫn tới nghiên cứu lựa chọn k=6 là số lượng nhóm tối ưu.



Hình 4. Kết quả phân vùng cảnh quan huyện Văn Chấn trên cơ sở phân cụm k-means với giá trị k=6 và biểu đồ chỉ số kiểm định giá trị tối ưu.

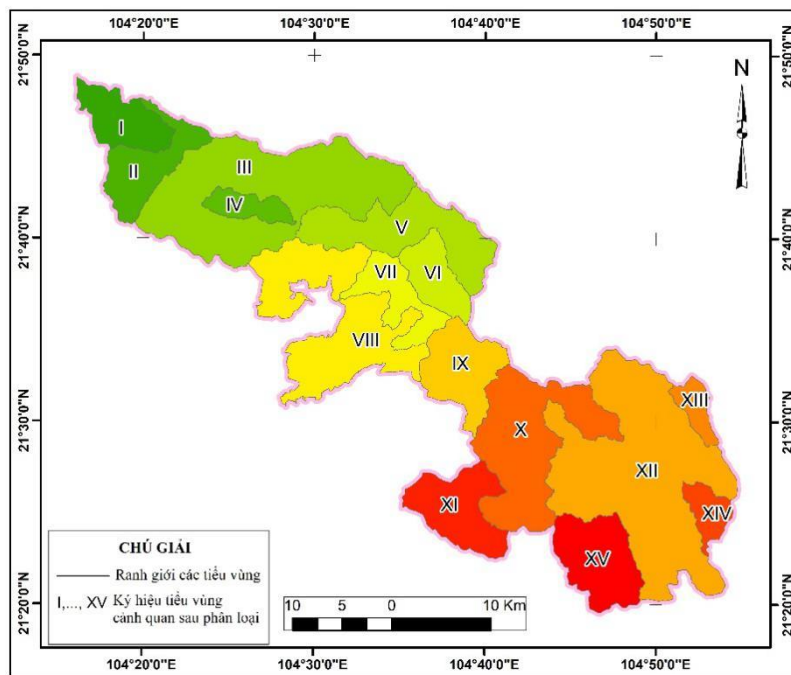


Ngoài ra, trên cơ sở phương pháp định lượng trong phân vùng cảnh quan, nghiên cứu tiến hành biên chỉnh kết quả về mặt thống kê sao cho phù hợp với yêu cầu phân vùng đã đề cập phía trên. Đồng thời, sự kết hợp của các thông tin cơ bản về đặc trưng thành tạo cảnh quan của khu vực nghiên cứu góp phần “chính xác hóa” kết quả nghiên cứu. Theo đó, kết quả phân vùng cảnh quan huyện Văn Chấn phục vụ mục đích khai

thác và sử dụng hợp lý tài nguyên được điều chỉnh thành 15 tiểu vùng như sau: *i) Lưu vực sông Ngòi Hút*: chỉ gồm tiểu vùng cảnh quan I và II nằm trên địa bàn xã Tú Lệ và Nậm Búng, phía Tây Bắc của huyện; *ii) Lưu vực sông Ngòi Thia*: gồm các tiểu vùng cảnh quan từ III – VIII nằm trong lòng thung lũng Nghĩa Lộ; *iii) Lưu vực sông Ngòi Lao*: gồm các tiểu vùng IX-XV nằm ở phía Nam của huyện (Hình 5 và Bảng 3).

Bảng 3. Thống kê các đơn vị tiểu vùng cảnh quan khu vực huyện Văn Chấn, tỉnh Yên Bái

|                     | Lưu vực sông     | Tiểu vùng (TV) cảnh quan và đơn vị cơ sở  |
|---------------------|------------------|---|
| Phân vùng cảnh quan | a. LVS Ngòi Hút  | TV I. (1, 2, 4), TV II. (3, 5, 6, 7, 8).  |
|                     | b. LVS Ngòi Thia | TV III. (9, 10, 11, 12, 15), TV IV. (13, 14), TV V. (17, 18, 19, 20, 25), TV VI. (24), TV VII. (23, 29, 36), TV VIII. (21, 22, 26, 27, 28, 30, 31, 32, 33, 34, 37).                           |
|                     | c. LVS Ngòi Lao  | TV IX. (38), TV X. (39, 40, 41, 43, 54, 55, 56, 57), TV XI. (51, 52, 53), TV XII. (42, 44, 45, 46, 49, 50, 58, 59, 61, 62, 63, 67, 68), TV XIII. (47, 48), TV XIV. (64), TV XV. (60, 65, 66). |



Hình 5. Kết quả phân vùng CQ sau hiệu chỉnh khu vực huyện Văn Chấn, tỉnh Yên Bái.

#### 4. Kết luận và kiến nghị

Phân vùng cảnh quan trên cơ sở tiếp cận định lượng là một phương thức phân loại đối tượng không gian rất hiệu quả trong nghiên cứu sinh thái cảnh quan và môi trường. Trên cơ sở tích hợp thêm nhóm yếu tố tính đa dạng, mô hình lý thuyết trong phân vùng cảnh quan đã mở rộng khả năng khai thác các độ đo định lượng và tối ưu hóa đặc trưng đồng nhất trong từng lớp thông tin/cấu trúc chuyên biệt của cảnh quan nhằm phân loại các nhóm lãnh thổ địa lý tự nhiên có tính đồng nhất. Căn cứ vào phương pháp phân cụm k-means và kiểm chứng về mặt thống kê, mô hình phân vùng cảnh quan đã lựa chọn giá trị tối ưu  $k=6$  cho địa bàn huyện Văn Chấn, tỉnh Yên Bái. Đây là cơ sở khoa học quan trọng để tiến hành kết hợp với các quan điểm phân vùng khác nhằm khu biệt 15 tiểu vùng cảnh quan.

Nhìn chung, cách tiếp cận này cho thấy khả năng tích hợp thông tin cấu trúc không gian trong sắp xếp và quy hoạch không gian; cũng như cho phép tinh chỉnh kết quả phân vùng theo mục tiêu và đối tượng nghiên cứu khác nhau. Tuy nhiên, hướng tiếp cận này cũng cần lưu ý một số điểm sau: i) Quá trình lựa chọn các biến phân vùng cảnh quan phụ thuộc chặt chẽ ý kiến chuyên gia – đây là thuận lợi cũng là hạn chế của nghiên cứu; ii) Quá trình phân cụm đã giảm đáng kể tính chủ quan trong xác định giá trị số lượng phân vùng tối ưu nhưng ranh giới trong không gian phụ thuộc “chặt chẽ” vào phương thức xác định các đơn vị không gian cơ sở; iii) Khả năng “khái quát hóa” thông tin trên một phạm vi lớn đem lại lợi thế nhất định, nhưng giảm khả năng phân định ranh giới một cách chính xác nếu dữ liệu trong từng đơn vị cơ sở thay đổi không theo quy luật không gian; iv) Phương pháp phân cụm k-means trong không gian có thể sử dụng dữ liệu từ nhiều nguồn (định tính hoặc định lượng, đơn biến hoặc đa biến) với số lượng mẫu bất kỳ tuy không tạo ra các cụm bị trùng lặp nhưng đòi hỏi sự tương đồng về giá trị khi tạo ra các cụm ngẫu nhiên. Do đó, tiếp cận định lượng sẽ đem lại hiệu quả cao khi coi đây là một bước “tiền xử lý” trước khi tiến hành phân vùng cảnh quan.

#### Tài liệu tham khảo

- [1] T. Ott, F. Swiaczny, Time-Integrative Geographic Information Systems: Management and Analysis of Spatio-Temporal Data, Springer-Verlag Berlin Heidelberg Publisher, 2001, 234 pages.
- [2] C. Xu, Sh. Sheng, T. Chi, X. Yang, S. An, M. Liu, Developing a quantitative landscape regionalization framework integrating driving factors and response attributes of landscapes, *Landscape and Ecological Engineering*, Volume 10, Issue 2, 2014, pp. 295–307. <https://doi.org/10.1007/s11355-013-0225-8>.
- [3] M.G. Turner, Spatial and temporal analysis of landscape patterns, *Landscape Ecology*, Volume 4, Issue 1, 1990, pp. 21-30. <https://doi.org/10.1007/BF02573948>.
- [4] G.P. Robertson, L.W. Burger, C.L. Kling, R.R. Lowrance, D.J. Mulla, Methods for Environmental Management Research at Landscape and Watershed Scales. Managing Agricultural Landscapes for Environmental Quality, *Journal of Soil and Water Conservation Society*, Ankeny, IA., 2007, 196 pages.
- [5] R.P. Powers, N.C. Coops, J.L. Morgan, M.A. Wulder, T.A. Nelson, C.R. Drever, S.G. Cumming, A remote sensing approach to biodiversity assessment and regionalization of the Canadian boreal forest, *Progress in Physical Geography: Earth and Environment*, Volume 37, Issues 1, 2013, pp. 36–62. <https://doi.org/10.1177/0309133312457405>.
- [6] M.A. Solans, A. Mellado-Díaz, A Landscape-Based Regionalization of Natural Flow Regimes in the Ebro River Basin and Its Biological Validation, *River Research and Applications* 31, 2015, pp. 457- 469. <https://doi.org/10.1002/rra.2860>.
- [7] B. Martín-López, I. Palomo, M. García-Llorente, I. Iniesta-Arandia, A.J. Castro, D.G.D. Amo, E. Gómez-Baggethun, C. Montes, Delineating boundaries of social ecological systems for landscape planning: A comprehensive spatial approach, *Land Use Policy*, Volume 66, 2017, pp. 90-104. <https://doi.org/10.1016/j.landusepol.2017.04.040>.
- [8] G. McMahon, E. Wiken, D. Gauthier, Toward a Scientifically Rigorous Basis for Developing Mapped Ecological Regions, *Environmental Management*, Volume 34, Supplement 1, 2004, pp. S111–S124. <https://doi.org/10.1007/s00267-004-0170-2>.
- [9] T.F. Stepinski, J. Niesterowicz, J. Jasiewicz, Pattern-based Regionalization of Large Geospatial

- Datasets Using Complex Object-based Image Analysis, *Procedia Computer Science*, Volume 51, 2015, pp. 2168-2177. <https://doi.org/10.1016/j.procs.2015.05.491>.
- [10] Y. Liu, B. Fu, Sh. Wang, W. Zhao, Global ecological regionalization: from biogeography to ecosystem services, *Current Opinion in Environmental Sustainability*, Volume 33, 2018, pp. 1-8. <https://doi.org/10.1016/j.cosust.2018.02.002>.
- [11] B.J. Amiri, G. Junfeng, N. Fohrer, F. Mueller, J. Adamowski, Regionalizing Flood Magnitudes using Landscape Structural Patterns of Catchments, *Water Resources Management*, Volume 32, Issue 7, 2018, pp. 2385-2403. <https://doi.org/10.1007/s11269-018-1935-3>.
- [12] J.J. Starn, K. Belitz, Regionalization of groundwater residence time using metamodeling, *Water Resources Research* 54, 2018, pp. 6357-6373. <https://doi.org/10.1029/2017WR021531>.
- [13] M. Gao, X. Chen, J. Liu, Z. Zhang, Regionalization of annual runoff characteristics and its indication of co-dependence among hydro-climate-landscape factors in Jinghe River Basin, China, *Stochastic Environmental Research and Risk Assessment*, Volume 32, Issue 6, 2018, pp. 1613-1630. <https://doi.org/10.1007/s00477-017-1494-9>.
- [14] W. Cao, S. Zhou, S. Wu, Land-use regionalization based on landscape pattern indices using rough set theory and catastrophe progression method, *Environmental Earth Sciences*, Volume 73, Issue 4, 2015, pp. 1611-1620. <https://doi.org/10.1007/s12665-014-3511-3>.
- [15] R. Wang, H. Yang, Landscape Regionalization for Highway Corridor Planning from Landscape Ecology Perspective: A Case Study of Shandong, China, 2018 International Conference on Sensing, Diagnostics, Prognostics, and Control (SDPC), China, 2018, pp. 503-508.
- [16] K. McGarigal, S. A. Cushman, E. Ene, FRAGSTATS v4: Spatial Pattern Analysis Program for Categorical and Continuous Maps. Computer software program produced by the authors at the University of Massachusetts, Amherst, 2012.
- [17] W.W. Hargrove, F.M. Hoffman, Potential of multivariate quantitative methods for delineation and visualization of ecoregions, *Environmental Management* 34 (Suppl. 1), 2004, pp. S39-S60. <https://doi.org/10.1007/s00267-003-1084-0>.
- [18] S. Ragetti, J. Zhou, H. Wang, Assessment of parameter regionalization methods for modeling flash floods in China, *Geophysical Research Abstracts*, Vol. 19, EGU2017-8018, 2017.
- [19] G.H. Shah, A.N.S. Badana, C. Robb, W.C. Livingood, Cross-Jurisdictional Resource Sharing in Changing Public Health Landscape: Contributory Factors and Theoretical Explanations, *Journal of Public Health Management and Practice*, Volume 22, Number 2, 2016, pp. 110-119. <https://doi.org/10.1097/PHH.0000000000000368>.
- [20] J. Niesterowicz, T.F. Stepinski, On using landscape metrics for landscape similarity search, *Ecological Indicators*, Volume 64, 2016, pp. 20-30. <https://doi.org/10.1016/j.ecolind.2015.12.027>.
- [21] O. Hall, W. Arnberg, A method for landscape regionalization based on fuzzy membership signatures, *Landscape and Urban Planning* 59, 2002, pp. 227-240. [https://doi.org/10.1016/S0169-2046\(02\)00050-6](https://doi.org/10.1016/S0169-2046(02)00050-6).
- [22] K.H. Riitters, J.D. Wickham, T.G. Wade, Evaluating anthropogenic risk of grassland and forest habitat degradation using landcover data. *Landscape Online* 13, 2009, pp. 1-14. <https://doi.org/10.3097/LO.200913>.
- [23] D.R. Grafius, R. Corstanje, J.A. Harris, Linking ecosystem services, urban form and green space configuration using multivariate landscape metric analysis, *Landscape Ecology*, 2018, Volume 33, Issue 4, pp. 557-573. <https://doi.org/10.1007/s10980-018-0618-z>.
- [24] M.J. Todd, P.J. Wigington, E.A. Sproles, Hydrologic Landscape Classification to Estimate Bristol Bay, Alaska Watershed Hydrology. *Journal of the American Water Resources Association (JAWRA)* 53 (5), 2017, pp. 1008- 1031. <https://doi.org/10.1111/1752-1688.12544>.
- [25] B. Choubin, K. Solaimani, M. Habibnejad Roshan, A. Malekian, Watershed classification by remote sensing indices: A fuzzy c-means clustering approach, *Journal of Mountain Science*, Volume 14, Issue 10, 2017, pp. 2053-2063. <https://doi.org/10.1007/s11629-017-4357-4>.
- [26] L. Kaufman, P.J. Rousseeuw, *Finding groups in data*, Wiley, New York, 1990, 342 pages.
- [27] M. Dale, M. Fortin, *Spatial Analysis: A Guide for Ecologists*, Cambridge University Press. 438 pages.