



TUYỂN TẬP BÁO CÁO HỘI NGHỊ TOÀN QUỐC

KHOA HỌC TRÁI ĐẤT VÀ TÀI NGUYÊN VỚI PHÁT TRIỂN BỀN VỮNG (ERSD 2024)

HÀ NỘI 14 - 11 - 2024

ERSD 2024



NHÀ XUẤT BẢN GIAO THÔNG VẬN TẢI

Phân vùng địa chất theo phương pháp học máy không giám sát (k-means) khu vực trung tâm Việt Nam theo tài liệu trọng lực

Phan Thị Hồng^{1,*}, Đỗ Minh Phương²

¹ Trường Đại học Mỏ - Địa chất

² Liên đoàn Vật lý Địa chất, Cục Địa chất Việt Nam

TÓM TẮT

Bài báo trình bày kết quả áp dụng phương pháp học máy không giám sát (k-means) trên tài liệu trọng lực để phân chia các vùng đồng nhất về thuộc tính vật lý của trường khu vực trung tâm Việt Nam. Tham số đầu vào của phương pháp gồm 04 thuộc tính của trường (đi thường trọng lực ở độ sâu 5km, gradient ngang, thuộc tính phương sai và hệ số truy tìm trục đi thường), kết hợp áp dụng phương pháp điểm khuỷu tay (Elbow method) chúng tôi phân thành 09 cụm tối ưu có đặc điểm gần nhau về thuộc tính của trường. Kết quả phân cụm cho thấy phía Tây Nam, trung tâm và phía Nam khu vực có tính chất vật lý biến đổi phức tạp, phản ánh hoạt động kiến tạo phức tạp với sự xáo trộn của các lớp đất đá khác nhau và đây là vị trí tiềm năng hình thành các mỏ khoáng sản nội sinh có nguồn gốc magma. Phía Bắc, Tây Bắc và phía Đông của khu vực có sự thay đổi không đáng kể về thuộc tính vật lý của trường, phản ánh sự bình ổn của các lớp đất đá với phương cấu trúc kéo dài theo phương Tây Bắc - Đông Nam. Như vậy, đối với những khu vực có ít thông tin địa vật lý, việc sử dụng các thuộc tính vật lý tương quan của cùng một trường địa vật lý là bộ dữ liệu đầu vào đáng tin cậy của phương pháp học máy không giám sát (k-means), kết quả này là cách tiếp cận ban đầu hiệu quả để xác định các cấu trúc địa chất mới và đánh giá nguồn gốc thành tạo các mỏ khoáng sản ẩn sâu trong khu vực.

Từ khóa: k-means; trọng lực; COSCAD 3D; khu vực trung tâm Việt Nam

1. Đặt vấn đề

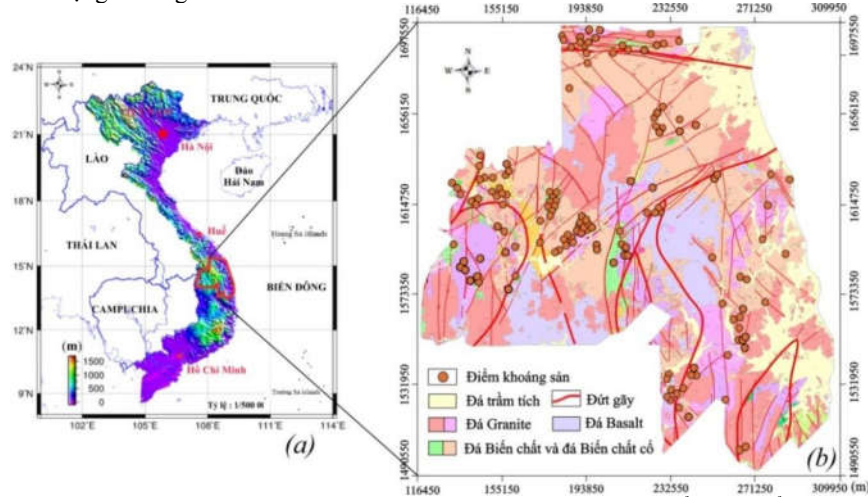
Thuật toán phân cụm k-means đã được sử dụng hiệu quả trong nhiều ngành khác nhau như công nghệ phần mềm, khai thác dữ liệu, phân tích hình ảnh, phân loại thạch học, phân tích dữ liệu địa vật lý [Bosch, 1999; Bosch và nnk, 2002; MacQueen, 1967; MaxMignotte, 2008], mục đích của thuật toán là tách các nhóm đối tượng trong một nhóm giống nhau về thuộc tính và khác với các đối tượng trong các nhóm khác. Cụ thể trong lĩnh vực địa vật lý, Bauer và cộng sự (2003) đã sử dụng phương pháp k-means để phân tách thành các nhóm thạch học khác nhau với các thuộc tính đầu vào gồm: vận tốc P nghịch đảo, tỷ lệ Poisson, kết hợp với dữ liệu vật lý thạch học; Delfiner và cộng sự (1987) đã được sử dụng phương pháp k - means để phân loại thạch học trong các lỗ khoan, Di Giuseppe (2017) cũng đã áp dụng phương pháp k - means để xác định đặc điểm cấu trúc địa chất của hệ thống địa nhiệt núi lửa, Campi Flegrei (Naples, Ý) với bộ dữ liệu đầu vào gồm các thông số điện trở suất, vận tốc sóng P và mật độ; Bernardinetti và cộng sự (2017) đã ứng dụng phân tích cụm k-means với bộ dữ liệu bao gồm điện trở suất, sóng địa chấn khúc xạ (sóng P) để xác định các đặc tính vật lý khác nhau và đặc tính cơ học của tầng chứa nước. Nhìn chung, phần lớn các công trình sử dụng thuật toán phân cụm k-means thường được áp dụng cho các tập dữ liệu đầu vào từ nhiều phương pháp địa vật lý khác nhau và ít nghiên cứu tập trung vào việc áp dụng thuật toán k-means với các tham số đầu vào từ các thuộc tính của một trường địa vật lý. Hơn nữa, trong trường hợp khu vực nghiên cứu ở một vùng mới có rất ít thông tin địa vật lý, việc sử dụng các đặc tính của trường địa vật lý làm tham số đầu vào cho phương pháp k-means có khả năng phát hiện ra các thông tin mới phục vụ cho các hoạt động thăm dò địa chất tiếp theo.

Khu vực Trung tâm Việt Nam là nơi phát triển và giao thoa của các khối lớn như đai uốn nếp Trường Sơn, đới khâu Tam Kỳ - Phước Sơn, khối nâng Kon Tum. Trong đó, các đứt gãy sâu theo phương Tây Bắc - Đông Nam (đứt gãy Po Kô), phương Đông Bắc - Tây Nam và á kinh tuyến là ranh giới giữa các cấu trúc nêu trên (Hai Thanh Tran và nnk, 2014; Quyen Minh Nguyen, 2019; Trần Văn Trị và Nguyễn Xuân Bao, 2008). Khu vực nghiên cứu đặc trưng bởi các thung lũng hẹp, chia cắt hiểm trở, địa hình phức tạp,

* Tác giả liên hệ

Email: phanthihong@humg.edu.vn

những núi cao kéo dài theo phương Đông Bắc, Tây Bắc và kinh tuyến. Những đặc điểm cấu trúc này là điều kiện tiên quyết thuận lợi cho việc hình thành các mỏ khoáng sản có nguồn gốc magma và nhiều nơi trong khu vực xuất hiện các điểm khoáng sản xuất lộ trên bề mặt (Hình 1b) (Nguyễn Trường Lưu và nnk, 2020; 2000; 2014; Nguyễn Xuân Sơn và nnk, 2000). Để xác định sự khác biệt về cấu trúc địa chất liên quan đến vị trí tiềm năng khoáng sản ở những khu vực có ít thông tin địa vật lý như khu vực trung tâm Việt Nam là một thách thức đối với các nhà xử lý, để giải quyết thách thức này, chúng tôi thấy rằng việc tích hợp nhiều thuộc tính liên quan của cùng một trường bằng cách sử dụng phương pháp học máy không giám sát (k-means) là một cách tiếp cận ban đầu hiệu quả để phát hiện các cấu trúc địa chất tiềm năng liên quan đến triển vọng khoáng sản ẩn sâu.

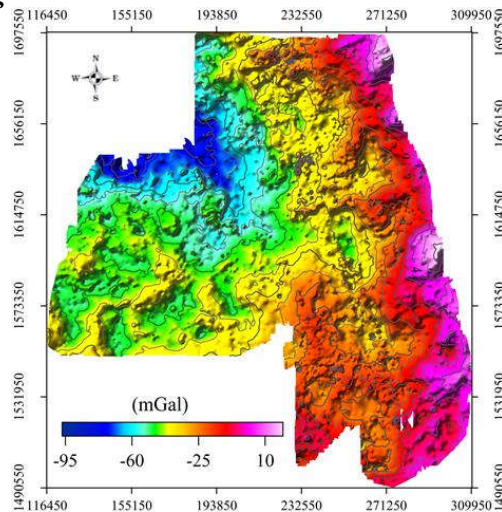


Hình 1. (a)- Vị trí khu vực nghiên cứu, độ cao địa hình so với mực nước biển thay đổi từ 10÷1600 m; (b)- Bản đồ địa chất và vị trí xuất lộ điểm quặng khu vực Trung tâm Việt Nam [22-25]

Trong bài báo này, chúng tôi áp dụng phương pháp học máy không giám sát (k-means) với 04 thuộc tính đầu vào của trường trọng lực gồm: dị thường trọng lực khu vực ở độ sâu 5km, gradient ngang trọng lực, đặc trưng phương sai và hệ số truy tìm trực dị thường trọng lực khu vực để phân phân thành các cụm có đặc trưng gần nhau về thuộc tính vật lý của trường. Các kết quả được phân chia thành các vùng đồng nhất về các thuộc tính vật lý của trường liên quan đến cấu trúc địa chất và phát hiện các vị trí tiềm năng liên quan đến mỏ khoáng sản ẩn sâu. Kết quả này có khả năng sẽ phát hiện thêm các thông tin mới phục vụ cho việc nghiên cứu, thăm dò trữ lượng khoáng sản của khu vực nghiên cứu trong tương lai.

2. Nguồn số liệu sử dụng và phương pháp nghiên cứu

2.1. Nguồn số liệu sử dụng



Hình 2. Dị thường trọng lực Bughe tại mặt quan sát tỷ lệ 1:100 000 khu vực trung tâm Việt Nam gồm cả dị thường trọng lực khu vực và dị thường trọng lực địa phương, giá trị trọng lực thay đổi từ -100 mGal đến 20 mGal

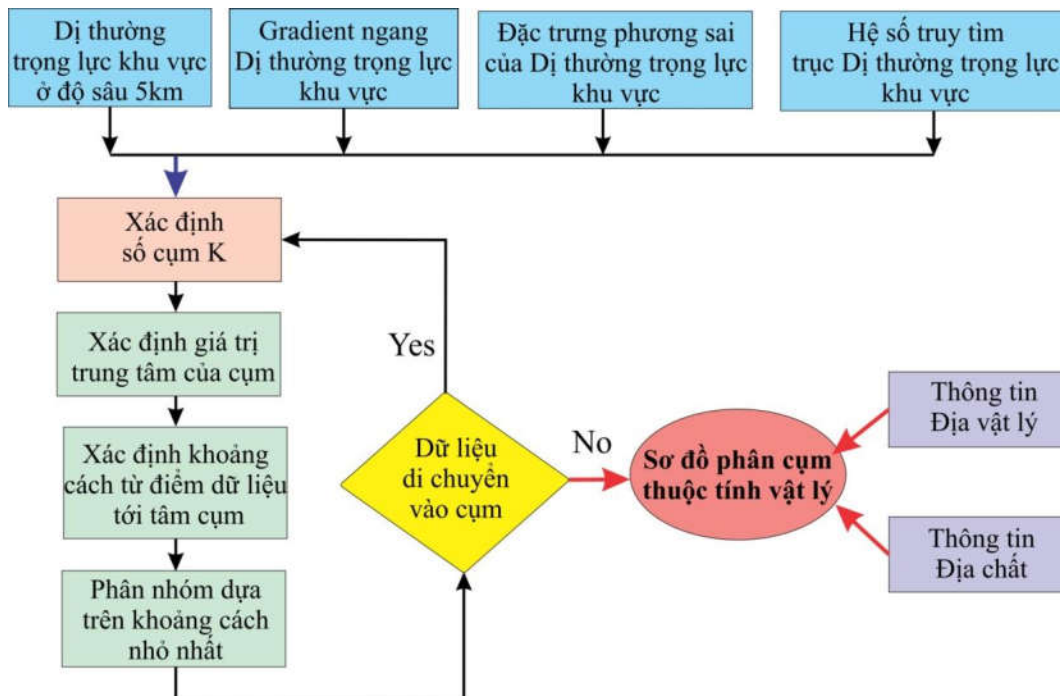
Nguồn số liệu sử dụng trong bài báo được thu thập từ nguồn số liệu dị thường trọng lực Bughe đầy đủ đo trên mặt đất tỷ lệ 1:100 000 với độ chính xác đo ghi từ $0,1 \pm 0,25$ mGal [22-25], sau khi hiệu chỉnh độ cao, hiệu chỉnh lớp giữa với mật độ $\delta_{lg} = 2,67$ g/cm³ và hiệu chỉnh địa hình theo phương pháp Prisivanco L. N. (Blakely, 1996) và được thể hiện trên hình 2.

Quan sát hình 2, chúng ta thấy dị thường trọng lực Bughe thay đổi từ $-100 \div 20$ mGal theo hướng Tây Bắc – Đông Nam. Phía Tây Bắc khu vực nghiên cứu được đặc trưng bởi giá trị dị thường thấp nhất trong khu vực với biên độ dao động từ $-100 \div -50$ mGal, phía Đông và Đông Nam được đặc trưng bởi giá trị dị thường cao nhất trong khu vực với biên độ thay đổi từ $-30 \div 15$ mGal. Phía Tây Nam của khu vực nghiên cứu quan sát thấy dải dị thường kéo dài theo phương Đông Bắc - Tây Nam với biên độ dao động từ $-40 \div -30$ mGal. Bên cạnh đó trong khu vực còn xuất hiện một số khối bất đồng nhất địa phương có giá trị trường thay đổi từ $-10 \div 5$ mGal, tập trung ở phía Bắc, trung tâm, và phía Nam khu vực nghiên cứu. Phía Đông Nam xuất hiện các đới dị thường địa phương biên độ dao động từ $-20 \div 5$ mGal.

2.2. Phương pháp nghiên cứu

Ngày nay có rất nhiều thuật toán học máy không giám sát được ứng dụng trong xử lý và phân tích tài liệu địa vật lý (Aline Melo và Yaoguo Li, 2011; Bosch, 1999; Bosch và nnk, 2002; Mac Queen, 1967; Max Mignotte, 2008). Tuy nhiên, trong nghiên cứu này chúng tôi chỉ tập trung vào thuật toán phổ biến nhất trong các kỹ thuật học không giám sát, đó là thuật toán phân cụm k-means. Thuật toán k-means là một trong những thuật toán phân cụm đơn giản và phổ biến nhất, được xuất bản lần đầu tiên vào năm 1955. Thuật toán có lịch sử phong phú và đa dạng vì nó được phát hiện độc lập trong các lĩnh vực khác nhau (Ball và Hall, 1965; Mac Queen, 1967). Mặc dù thuật toán k-means được đề xuất lần đầu tiên cách đây hơn 50 năm nhưng nó vẫn là một trong những thuật toán được sử dụng rộng rãi nhất để phân cụm, dễ sử dụng, dễ thực hiện, hiệu quả, đơn giản. Trong phân cụm k-means, khoảng cách Mahalanobis trong không gian Hilbert được sử dụng làm khoảng cách giữa các tham số riêng lẻ và tính đến các mối quan hệ trong không gian đặc trưng của trường. Chúng tôi sử dụng khoảng cách Mahalanobis tổng quát vì thuật toán mang lại hiệu suất phân cụm tốt nhất, cả về giá trị trung bình và độ lệch chuẩn của số lượng đường cong được phân loại sai (Andrea Martino và nnk, 2017; Nikitin và Petrov, 2008).

Thuật toán k-means (Andrea và nnk, 2017; Nikitin và Petrov, 2008) chia thành các cụm khác nhau, với các bộ dữ liệu đầu vào bao gồm: dị thường trọng lực, gradient ngang trọng lực (Nikitin và Petrov, 2008), thuộc tính phương sai (Petrov, 1997; Petrov và nnk, 2010) và hệ số truy tìm trực dị thường trọng lực (Hong và nnk, 2020; Hong, 2022; Nikitin và Petrov, 2008) để phân tách thành các cụm có tính chất tương tự khu vực trung tâm Việt Nam.



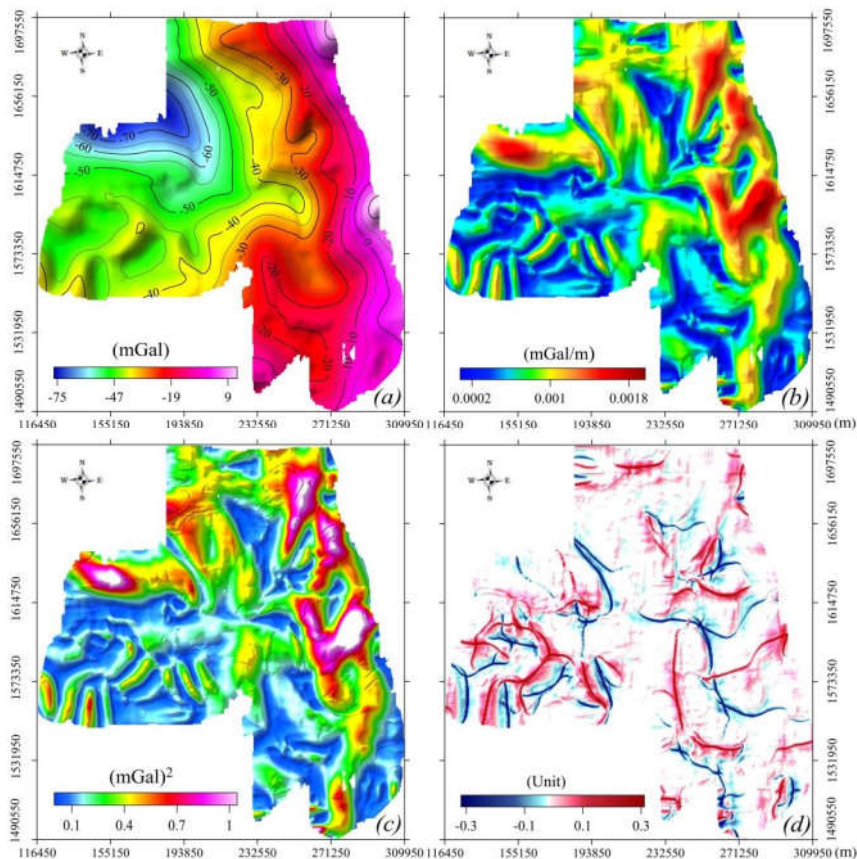
Hình 3. Sơ đồ minh họa việc tích hợp dị thường trọng lực và thuộc tính trường bằng thuật toán phân cụm k-means áp dụng ở khu vực trung tâm Việt Nam

3. Kết quả và thảo luận

3.1. Kết quả xử lý tham số đầu vào của phương pháp học không giám sát

Như đã phân tích ở trên, trong nghiên cứu này chúng tôi sử dụng tập dữ liệu gồm 04 thuộc tính đầu vào để đưa vào thuật toán phân cụm k-means gồm: Độ thường trọng lực khu vực ở độ sâu 5km (hình 4a), gradient ngang độ thường trọng lực khu vực (hình 4b), thuộc tính phương sai độ thường trọng lực khu vực (hình 4c) và kết quả truy tìm trực độ thường trọng lực khu vực (hình 4d).

Các dải cực trị liên tục của giá trị gradient ngang (hình 4b) và thuộc tính phương sai của trường (hình 4c) phản ánh ranh giới của các khối không đồng nhất, ranh giới của các hệ thống đứt gãy. Các dải kéo dài của trục độ thường dương (hiển thị bằng màu đỏ), của trục độ thường âm (hiển thị bằng màu xanh đậm) (hình 4d) phản ánh xu hướng của các khối độ thường dương và độ thường âm trong khu vực nghiên cứu. Tổng hợp các kết quả xử lý (hình 4a, 4b, 4c, 4d) cho chúng ta thông tin về đặc điểm địa chất trong khu vực, cụ thể là các khoảng cực trị mở rộng của giá trị gradient ngang, của phương sai (hoặc trục độ thường) phát triển theo phương Tây Bắc - Đông Nam; Đông Bắc - Tây Nam, á kinh tuyến.



Hình 4. (a)- Độ thường trọng lực khu vực sau khi loại bỏ độ thường địa phương ở độ sâu 5km; (b)- Gradient ngang độ thường trọng lực khu vực; (c)- Thuộc tính phương sai trường độ thường trọng lực khu vực; (d)- Kết quả truy tìm trực độ thường trọng lực khu vực

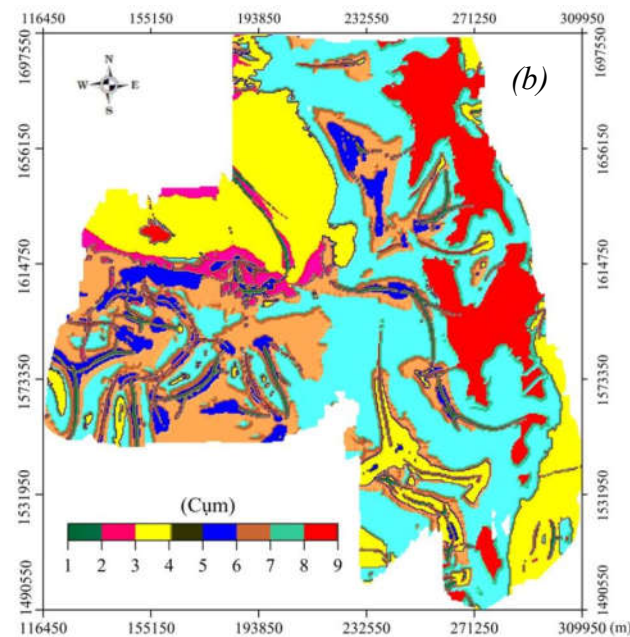
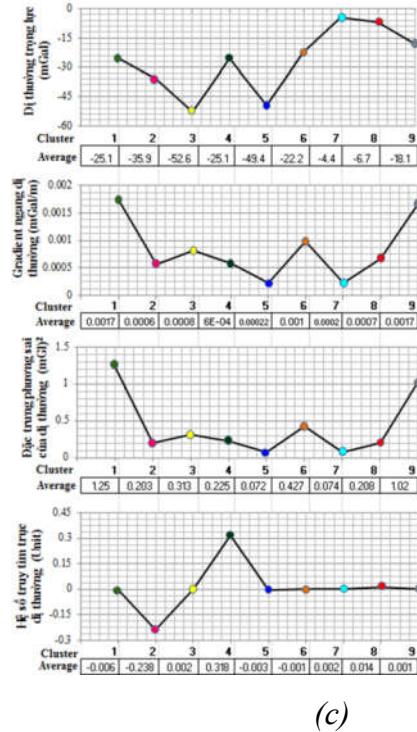
3.2. Kết quả áp dụng phương pháp học máy không giám sát (k-means)

Áp dụng thuật toán học máy không giám sát k-means với 04 thuộc tính đầu vào (hình 4a, 4b, 4c, 4d) kết hợp phương pháp điểm khuỷu tay (Elbow method) chúng tôi phân thành 09 cụm tối ưu có đặc điểm gần nhau về thuộc tính trường độ thường trọng lực, giá trị gradient ngang, thuộc tính phương sai và hệ số truy tìm trực độ thường (hình 4).

Kết quả phân chia thành 09 cụm (hình 5) cho thấy vùng nghiên cứu tồn tại tính chất vật lý phân cắt phức tạp ở phía Tây Nam, phía Nam và trung tâm khu vực nghiên cứu, so với bản đồ cấu trúc địa chất, khu vực này thuộc hệ rift nội lục địa sau va chạm Mesozoic (Hai Thanh Tran và nnk, 2014; Quyên Minh Nguyen, 2019). Như vậy, kết quả áp dụng thuật toán phân cụm k-means rất phù hợp với dữ liệu địa chất. Ở phía Bắc, Tây Bắc và Đông của khu vực nghiên cứu (hình 5), các đặc tính vật lý của trường ít thay đổi,

chứng tỏ các vùng này khá ổn định theo phương Tây Bắc - Đông Nam. Do vùng phía Đông khu vực nghiên cứu giáp biển Đông nên kết quả này hoàn toàn phù hợp với các quy luật địa chất. Quan sát hình 5 cho thấy hoạt động kiến tạo phức tạp ở phía Tây Nam và phía Nam khu vực nghiên cứu, chứng tỏ hoạt động kiến tạo này mang tính chất khu vực và chúng đóng vai trò là kênh dẫn đưa magma từ sâu lên bề mặt, xuyên qua các trầm tích gần bề mặt và hình thành các mỏ khoáng sản ẩn sâu.

Số cụm	Giá trị trọng lực khu vực ở độ sâu 5km (mGal)	Gradient ngang giá trọng trọng lực khu vực (mGal/m)	Đặc trưng phương sai giá trọng lực khu vực (mGal ²)	Hệ số truy tìm trực giá trọng lực khu vực
1	-25.1	0.001746	1.247	-0.006
2	-35.93	0.000569	0.203	-0.238
3	-52.6	0.000815	0.313	0.002
4	-25.08	0.000588	0.225	0.318
5	-49.4	0.000221	0.072	-0.003
6	-22.2	0.000981	0.427	-0.001
7	-4.4	0.000221	0.074	0.002
8	-6.7	0.000667	0.208	0.014
9	-18.02	0.001672	1.02	0.001



Hình 5. Kết quả phân thành 09 cụm tối ưu theo phương pháp học máy không giám sát (k-means), các chấm tròn màu sắc ở hình (a) tương ứng với màu sắc của các cụm ở hình (b). Trong đó (a, b)- Giá trị trung tâm của 09 cụm theo trường giá trọng trọng lực, gradient ngang, thuộc tính phương sai và hệ số truy tìm trực giá trọng; (c)- Kết quả phân chia thành 09 cụm có đặc điểm gần giống nhau thuộc tính của trường giá trọng trọng lực khu vực trung tâm Việt Nam

Kết luận

Kết quả xử lý và luận giải tài liệu giá trọng trọng lực Bughe khu vực trung tâm Việt Nam đưa ra một số kết luận:

- Thuật toán học máy không giám sát (k-means) kết hợp phương pháp điểm khuỷu tay (Elbow method) chúng tôi phân thành 09 cụm tối ưu có đặc điểm gần nhau về thuộc tính của trường được áp dụng hiệu quả trên tài liệu giá trọng trọng lực khu vực trung tâm Việt Nam; kết quả phân cụm này hoàn toàn trùng khớp với địa chất kiến tạo trong khu vực.

- Phía Tây Nam, trung tâm, và phía Nam khu vực nghiên cứu cho thấy thuộc tính vật lý của trường

khá phức tạp cho thấy tại khu vực này các lớp đất đá bị vỡ nhàu và đứt đoạn theo nhiều hướng khác nhau và đây là vị trí tiềm năng hình thành các mỏ khoáng sản nội sinh có nguồn gốc magma.

- Phía Bắc, Tây Bắc, và phía Đông khu vực nghiên cứu cho thấy các thuộc tính vật lý của trường ít thay đổi phản ánh các lớp đất đá khu vực này khá bình ổn với phương cấu trúc kéo dài theo phương Tây Bắc - Đông Nam.

- Đối với những khu vực có ít thông tin địa vật lý, việc sử dụng các thuộc tính vật lý tương quan của cùng một trường địa vật lý là bộ dữ liệu đầu vào đáng tin cậy của thuật toán học máy không giám sát (k-means) với khoảng cách Mahalanobis tổng quát, kết hợp phương pháp điểm khuỷu tay (Elbow method), kết quả này là một cách tiếp cận ban đầu hiệu quả để phát hiện các cấu trúc tiềm năng liên quan đến khoáng sản ẩn sâu.

Tài liệu tham khảo

Aline Melo, Yaoguo Li, 2011. Geology differentiation by applying unsupervised machine learning to multiple independent geophysical inversions. *Geophysical Journal International*, 2058-2078.

Andrea Martino, Andrea Ghiglietti, Francesca Ieva, Anna Maria Paganoni, 2017. A k-means procedure based on a Mahalanobis type distance for clustering multivariate functional data. *Methodology*, 22 pages.

Blakely, R.J, 1996. *Potential theory in gravity and magnetic application*. Cambridge University Press, 461 pages.

Ball, G., Hall, D. Isodata, 1965. A novel method of data analysis and pattern classification. *Technical report NTIS AD 699616*. Stanford Research Institute, Stanford, CA.

Bauer, K., Schulze, A., Ryberg, T., Sobolev, S.V. & Weber, M.H, 2003. Classification of lithology from seismic tomography: a case study from the Messum igneous complex, Namibia, *J. geophys. Res.*, 108(B3), 2156–2202.

Bosch M, 1999. Lithologic tomography: from plural geophysical data to lithology estimation. *J. Geophys. Res.* 104:749–766.

Bosch, M., Zamora, M., Utama, W, 2002. Lithology discrimination from physical rock properties. *Geophysics* 67:573–581.

M.G. Di Giuseppe, A. Troiano, D. Patella, M. Piochi, S. Carlino, 2017. A geophysical k-means cluster analysis of the Solfatara-Pisciarelli volcano geothermal system, Campi Flegrei (Naples, Italy). *Journal of Applied Geophysics*, 1-11.

Delfiner, P., Peyret, O. & Serra, O, 1987. Automatic determination of lithology from well logs, *SPE Format. Eval.*, 2(03), 303–310.

Hai Thanh Tran, Khin Zaw, Jacqueline A. Halpin, Takayuki Manaka, Sebastien Meffre, Chun-Kit Lai, Youjin Lee, Hai Van Le, Sang Dinh, 2014. The Tam Ky- Phuoc Son shear zone in central Vietnam: Tectonic and metallogenic implications. *Gondwana Research* 26, 144-164.

Hong T. Phan, Aleksey V. Petrov, Phuong M. Do, Giau M.lai, Luu T. Nguyen, 2020. Geological structure of central Vietnam by interpretation processing of gravitational survey data using the “COSCAD 3D” computer technology. *Geology and Exploration*, 77-90.

Hong P. T., 2022. *Computer technology for interpretational processing of gravity and magnetic processing data using probabilistic methods statistical approach (by the example of the territory of central vietnam)*. Doctoral thesis, Russia, 126 pages.

Mac Queen, J., 1967. Some methods for classification and analysis of multivariate observations. In: *Fifth Berkeley Symposium on Mathematics, Statistics and Probability*. University of California Press, pp. 281–297b.

Max Mignotte, 2008. Segmentation by Fusion of Histogram-Based K-Means Clusters in Different Color Spaces. *Ieee transactions on image processing*, vol. 17, no. 5. 780-787.

Nikitin A.A., Petrov A.V., 2008. *Theoretical foundations of geophysical information processing*. Study guide, Moscow 113 pages.

Nguyen Truong Luu, Hoang Van Long, Phan Duc Le, Pham Van Hung, Nguyen Thien Cong, Tran Duan, 2020. Applying a combination of geophysical-geological methods in investigating and detecting copper-uran mineral ores in Kon Ra-Kon Tum area. *Geological Journal*, Series A, №. 371-372, 1-18.

Nguyen Truong Luu et al, 2000. Flight measured from 1:50 000 scale gamma spectrum and 1: 100 000 scale gravity measurement in the central region of Vietnam. Information Center, *Geological Archives*, Hanoi, 218 pages.

Nguyen Truong Luu et al, 2014. Flight measured from the gamma spectrum at the scale of 1:50 000 and measured the gravity at the scale of 1:100 000 in the southern Pleiku region. Information Center, *Geological Archives*, Hanoi, 159 pages.

Nguyen Xuan Son et al, 2000. Flight results measured from the 1: 50 000 scale gamma spectrum and

measured in the Kon Tum area. Information Center, *Geological Archives*, Hanoi, 259 pages.

Petrov A.V., 1997. *Methods of adaptive filtering and multivariate analysis of variance in the computer system for processing geophysical information "COSCAD 3D"*. Doctoral thesis, Russia 192 pages.

Petrov A.V., Yudin D.B., Soeli Hou, 2010. Processing and interpretation of geophysical data by methods of a probabilistic-statistical approach using computer technology "COSCAD 3D". *Earth sciences*, №2, UDC 551-214, 126-132.

Quyen Minh Nguyen, Quinglai Feng, Jian-Wei Zi, Tianyu Zhao, Hai Thanh Tran, Thanh Xuan Ngo, Dung My Tran, Hung Quoc Nguyen, 2019. Cambrian intra-oceanic arc trondhjemite and tonalite in the Tam Ky – Phuoc Son Suture zone, central Vietnam: Implications for the early Paleozoic assembly of the Indochina block. *Gondwana Research* 70, 151-170.

S. Bernardinetti, S. Maraio, F. N. A. Bruno, V. Cicala, S. Minucci, M. Giannuzzi, M. Trotta, F. Cureda, S. Febo, M. Vacca, E. Guastaldi, T. Colonna, F. Bonciani, E. Tufarolo, F. Brogna, A. Zirulia, O. Milighetti, 2017. Potential shallow aquifers characterization through an integrated geophysical method: multivariate approach by means of k-means algorithms. *Italian Journal of Groundwater* - AS21-278: 07 – 14.

Tran Van Tri, Nguyen Xuan Bao, 2008. Map of geological structure at 1:500 000 scale in the central region of Vietnam. Information Center, *Geological Archives*, Hanoi.

ABSTRACT

Geology differentiation by applying unsupervised machine learning (k-means) of gravity field in central area of Vietnam

Phan Thi Hong¹, Do Minh Phuong²

¹ Hanoi University of Mining and Geology

²Geophysical Division, The Vietnam Geological Department

This paper presents the results of applying the unsupervised machine learning method (k-means) on gravity data to divide homogeneous regions in terms of physical attributes of the field in central area of Vietnam. The input parameters of the method include 04 field attributes (gravity anomaly, horizontal gradient, variance attribute and the tracing coefficient of the gravity anomalous axis), the results are separated into 09 clusters with characteristics close to each other in terms of physical properties of the field. Clustering results show that the Southwest, Center and South of the area have complex physical changes, reflecting complex tectonic activities with the disturbance of different soil and rock layers, and this is potential locations for forming endogenous mineral deposits of magmatic origin. The North, Northwest and East of the area have insignificant changes in the field's physical properties, reflecting the stability of the soil and rock layers with the structural direction extending in the Northwest - Southeast direction. Thus, for areas with little geophysical information, using correlated physical attributes of the same geophysical field is a reliable input data set of unsupervised machine learning methods. (k-means), this result is an effective initial approach to identify new geological structures and evaluate the origin of mineral deposits hidden deep in the area.

Keywords: k-means; gravity field; COSCAD 3D; central area of Vietnam