



ỨNG DỤNG HỌC MÁY VÀ PHƯƠNG PHÁP KRIGING TRONG NGHIÊN CỨU ĐỊA CHẤT MỎ ĐẤT HIẾM DẠNG HẤP PHỤ ION

Trịnh Đức Thiện¹, Trịnh Đình Huân¹, Hoàng Văn Dũng¹, Hoàng Thị Thoa³, Ngô Xuân Đắc^{2,*}

¹Liên đoàn Địa chất Xạ - Hiếm, Xuân Phương, Hà Nội, Việt Nam

²Viện Khoa học Địa chất và Khoáng sản, 67 Chiến Thắng, Hà Nội, Việt Nam

³Trường Đại học Mỏ - Địa chất, 18 Phố Viên, Hà Nội, Việt Nam

THÔNG TIN BÀI BÁO

CHUYÊN MỤC: Công trình khoa học

Ngày nhận bài: 19/9/2025

Ngày nhận bài sửa: 18/10/2025

Ngày chấp nhận đăng: 20/10/2025

^{2,*}Tác giả liên hệ:

Email: nxdac@mae.gov.vn

TÓM TẮT

Mỏ đất hiếm dạng hấp phụ ion là loại hình khoáng sản có giá trị chiến lược, phân bố phổ biến trong các vỏ phong hóa tại vùng nhiệt đới ẩm. Tuy nhiên, việc đánh giá tài nguyên và phân vùng triển vọng đối với kiểu mỏ này còn gặp nhiều thách thức do đặc điểm phân bố phức tạp và dữ liệu thực địa hạn chế. Trong nghiên cứu này, các phương pháp hiện đại gồm nội suy địa thống kê Kriging và mô hình học máy như Random Forest (RF) và Support Vector Machine (SVM) đã được tích hợp để phân tích dữ liệu địa hóa và mô hình hóa phân bố tài nguyên đất hiếm (REE) trên diện tích khảo sát 10 km².

Trên cơ sở 30 điểm mẫu thực nghiệm và lưới dữ liệu mô phỏng gồm 10.000 điểm, các mô hình học máy đã được huấn luyện để xác định vùng triển vọng theo hàm lượng REE_{sum}. Kết quả phân vùng cho thấy khu vực trung tâm có xác suất triển vọng cao, được cả hai mô hình RF và SVM đồng thuận. Phép nội suy Kriging cho thấy phân bố hàm lượng REE giảm dần từ trung tâm ra rìa với mức độ liên tục không gian cao. Dữ liệu cũng được chia theo block 500 m × 500 m để tính toán tài nguyên, với tổng tài nguyên ước tính đạt khoảng 11.300 tấn REE, chủ yếu tập trung tại các block trung tâm. Mô hình khối 3D đã được xây dựng nhằm hỗ trợ trực quan hóa và định hướng khai thác.

Nghiên cứu này khẳng định hiệu quả của việc tích hợp phương pháp Kriging và các mô hình học máy (machine learning models) trong mô hình hóa không gian và đánh giá tài nguyên mỏ đất hiếm hấp phụ ion. Kết quả không chỉ có ý nghĩa phương pháp mà còn mang giá trị khoa học và thực tiễn, góp phần định hướng các nghiên cứu và điều tra tài nguyên đất hiếm (REE_{sum}) tại Việt Nam trong thời gian tới.

Từ khóa: đất hiếm, hấp phụ ion, phương pháp Kriging, học máy, Random Forest, SVM

@ Hội Khoa học và Công nghệ Mỏ Việt Nam

1. ĐẶT VẤN ĐỀ

Trong bối cảnh hiện nay, khi nhu cầu về khoáng sản chiến lược như đất hiếm ngày càng tăng do vai trò thiết yếu trong các ngành công nghệ cao như điện tử, năng lượng tái tạo và quốc phòng, việc nghiên cứu địa chất mỏ nhằm đánh giá trữ lượng, phân bố và tiềm năng khai thác khoáng sản ngày càng trở nên cấp thiết (Castor & Hedrick, 2006; Xu et al., 2020).

Tuy nhiên, dữ liệu địa chất thường mang tính rời rạc, không gian không đồng đều và bị giới hạn bởi chi phí và điều kiện địa hình trong quá trình lấy mẫu. Do đó, việc áp dụng các công nghệ phân tích

và dự báo hiện đại nhằm tối ưu hóa hiệu quả xử lý dữ liệu trở thành một hướng đi cần thiết (Zuo, 2017; Carranza, 2022).

Trong những năm gần đây, học máy (Machine Learning) và phương pháp Kriging (Kriging) một kỹ thuật nội suy địa thống kê mạnh mẽ đã nổi lên như các công cụ hiệu quả, hỗ trợ phân tích, mô hình hóa và dự báo các đặc trưng địa chất, địa hóa và khoáng sản trên cơ sở dữ liệu đầu vào hạn chế (Goovaerts, 1997; Zuo, 2017). Sự kết hợp giữa Kriging và các mô hình học máy không chỉ nâng cao độ chính xác trong nội suy không gian mà còn cho phép phát hiện các mối liên hệ phi tuyến, phức

tạp giữa các biến địa chất điều mà các phương pháp truyền thống khó nhận diện (Grunsky et al., 2009; Li et al., 2023).

Mỏ đất hiếm dạng hấp phụ ion là kiểu mỏ đặc trưng cho các vùng nhiệt đới ẩm, nơi các đá giàu đất hiếm như granitoid, rhyolite, syenite... bị phong hóa mạnh tạo thành lớp vỏ phong hóa dày. Trong đó, các ion REE^{3+} được hấp phụ yếu trên bề mặt các khoáng sét như kaolinite, halloysite, gibbsite... (Xu et al., 2020). Do không tồn tại dưới dạng khoáng vật riêng biệt, đất hiếm trong các mỏ này rất khó quan sát bằng kính hiển vi, và sự phân bố của chúng chịu chi phối bởi các yếu tố địa hóa và môi trường rất phức tạp – gây thách thức lớn cho công tác thăm dò và đánh giá tài nguyên.

Kriging, đặc biệt là Ordinary Kriging và Co-Kriging, đã được sử dụng hiệu quả để nội suy hàm lượng REE trong không gian ba chiều, góp phần xây dựng mô hình khối mỏ và tính toán trữ lượng (Goovaerts, 1997; Journel & Huijbregts, 1978). Trong khi đó, Machine Learning với các mô hình như Random Forest, Support Vector Machine, Gradient Boosting và mạng nơ-ron nhân tạo đang ngày càng được ứng dụng để: (1) phân tích tương quan giữa REE và các yếu tố địa hóa đi kèm (Al_2O_3 , Fe_2O_3 , pH, độ sâu...); (2) dự báo các khu vực có tiềm năng REE cao dựa trên dữ liệu bề mặt; (3) xây dựng bản đồ triển vọng khoáng sản; (4) đề xuất vị trí lấy mẫu tối ưu nhằm tiết kiệm chi phí khảo sát (Carranza, 2022; Li et al., 2023).

Sự kết hợp giữa Kriging (khai thác tính liên tục trong không gian) và Machine Learning (khai thác quan hệ phi tuyến trong dữ liệu) đã tạo nên một phương pháp tích hợp mạnh mẽ, cho phép mô hình hóa tốt hơn sự phân bố đất hiếm trong lớp phong hóa, đồng thời nâng cao hiệu quả của công tác thăm dò, phân vùng triển vọng và đánh giá tài nguyên.

Trong bối cảnh dữ liệu địa chất ngày càng phong phú và nhu cầu về tài nguyên chiến lược như đất hiếm tiếp tục gia tăng, việc tích hợp các công cụ phân tích hiện đại như Machine Learning và Kriging đang mở ra một hướng tiếp cận mới hiệu quả và chính xác hơn trong nghiên cứu địa chất mỏ nói chung, và đặc biệt là đối với các mỏ đất hiếm hấp phụ ion một trong những loại hình mỏ có ý nghĩa quan trọng bậc nhất trong thế kỷ 21 (Carranza, 2022; Xu et al., 2020).

2. DỮ LIỆU VÀ PHƯƠNG PHÁP NGHIÊN CỨU

Dữ liệu nghiên cứu bao gồm hai bảng chính: (1) bảng mẫu thực nghiệm gồm 30 điểm đo phân bố trên khu vực nghiên cứu thực nghiệm; (2) bảng dữ liệu mô phỏng gồm 10.000 điểm lưới đều nhau,

phủ kín toàn bộ diện tích 10 km² (tương ứng 1 điểm trên mỗi ô vuông 100 m²). Mỗi điểm dữ liệu đều có thông tin về tọa độ không gian (X, Y), hàm lượng tổng đất hiếm (REE_sum), chiều dày thân quặng, mật độ đá (rock density), và diện tích block (block area).

Trong nghiên cứu thực nghiệm này, các mô hình Random Forest (RF) và Support Vector Machine (SVM) đã được xây dựng và huấn luyện trên tập dữ liệu mẫu nhằm mục tiêu phân loại và xác định các vùng có triển vọng quặng đất hiếm dựa trên chỉ tiêu chính là hàm lượng REE_sum. Các mô hình học máy như RF và SVM đã được chứng minh là hiệu quả trong phân vùng triển vọng khoáng sản và xử lý dữ liệu địa hóa đa biến (Carranza, 2022; Li et al., 2023; Zuo, 2017). Hiệu suất mô hình được đánh giá bằng các chỉ số như độ chính xác, AUC và ma trận nhầm lẫn, trước khi áp dụng cho bộ dữ liệu mô phỏng để xây dựng bản đồ phân vùng triển vọng.

Song song với đó, nội suy địa thống kê Ordinary Kriging được thực hiện trên tập dữ liệu mẫu để ước lượng phân bố không gian của hàm lượng REE_sum. Phương pháp Kriging đã được ứng dụng rộng rãi và được xem là tiêu chuẩn trong mô hình hóa địa hóa không gian (Goovaerts, 1997; Journel & Huijbregts, 1978). Kết quả từ Kriging cung cấp cơ sở để so sánh và hiệu chỉnh kết quả từ mô hình học máy, đồng thời hỗ trợ trong việc xác định các khối quặng có giá trị kinh tế tiềm năng.

Việc tích hợp Kriging và các mô hình học máy giúp tăng độ chính xác trong xác định vùng khoáng hóa giàu đất hiếm, đồng thời tối ưu hóa quá trình thăm dò và xây dựng mô hình địa chất 3D có độ tin cậy cao (Zuo, 2017; Carranza, 2022).

2.1. Biểu đồ thống kê hàm lượng REE toàn bộ vùng

a. Biểu đồ Histogram hàm lượng REE_sum

Dạng phân bố: Histogram thể hiện phân bố lệch trái, với cực đại nằm trong khoảng 160–170 ppm, cho thấy phần lớn khu vực có hàm lượng REE_sum tập trung quanh mức trung bình này. Dạng phân bố lệch như vậy thường gặp trong các hệ thống khoáng hóa bị chi phối bởi điều kiện phong hóa không đồng đều (Zuo, 2017).

Dải biến thiên: Hàm lượng REE_sum trải rộng từ ~100 ppm đến hơn 240 ppm, phản ánh sự biến thiên đáng kể về mặt địa hóa điều đặc trưng cho các hệ thống phong hóa sâu phát triển trên đá mẹ có thành phần REE khác nhau (Xu et al., 2020).

Sự thiếu cân đối: Sự lệch trái nhẹ có thể do sự hiện diện của các vùng dị thường giàu REE cục bộ, trong khi phần lớn diện tích vẫn có hàm lượng thấp hơn mức cực đại. Đây là đặc điểm thường thấy

trong các mỏ đất hiếm hấp phụ ion phân bố phân tán (Li et al., 2023).

b. Biểu đồ Boxplot hàm lượng REE_sum

Tầm giá trị (range): Boxplot cho thấy khoảng biến thiên chính từ 140–190 ppm, với: Q1: ~150 ppm; Trung vị: ~165 ppm; Q3: ~180 ppm

Outlier: Không ghi nhận điểm ngoại lai rõ rệt vượt xa giới hạn hộp, chứng tỏ dữ liệu có mức độ đồng đều tương đối cao. Điều này cho thấy mô hình địa hóa trong khu vực có tính ổn định tương đối trong tầng phong hóa một đặc điểm có thể khai thác hiệu quả trong nội suy không gian và mô hình tài nguyên (Goovaerts, 1997; Grunsky et al., 2009).

c. Bản đồ phân bố REE theo không gian

Phân bố không gian: Hàm lượng REE_sum cao nhất tập trung tại tọa độ trung tâm bản đồ ($X \approx 5000$ m, $Y \approx 5000$ m), thể hiện vùng dị thường trung tâm điển hình, rất phổ biến trong các hệ thống phong hóa tại vùng nhiệt đới ẩm (Castor & Hedrick, 2006; Xu et al., 2020).

Gradient giảm dần: Hàm lượng REE_sum giảm từ trung tâm ra rìa theo dạng vòng tròn đồng tâm, gợi ý một hệ thống khoáng hóa phát triển trên nền đá mẹ giàu REE bị phong hóa sâu. Đây là kiểu phân bố điển hình của mỏ đất hiếm hấp phụ ion được ghi nhận tại Trung Quốc và Đông Nam Á (Li et al., 2023; Zuo, 2017).

Đồng phân màu: Bản đồ thể hiện phân bố mượt, liên tục, không có nhiều hoặc dao động đột ngột, cho thấy nội suy bằng Kriging hoặc các kỹ thuật tương đương như hồi quy theo quá trình Gauss hay còn gọi là một phương pháp hồi quy không tham số (Gaussian Process Regression) được thực hiện tốt (Goovaerts, 1997; Cressie, 1993).

d. Biểu đồ Histogram mật độ (Density)

Phân bố gần đồng đều: Mật độ đá dao động trong khoảng 1.2–2.2 g/cm³, không có đỉnh nổi bật rõ ràng. Điều này phản ánh sự hiện diện của nhiều loại đá nền (granit phong hóa, ryolit, sét...), nhưng không có loại nào chiếm ưu thế tuyệt đối phù hợp với các vùng phong hóa trên nền đá magma axit (Castor & Hedrick, 2006).

Ứng dụng: Mật độ là tham số quan trọng trong việc tính khối lượng thân quặng từ thể tích, và đóng vai trò thiết yếu trong các mô hình tài nguyên dạng block 3D. Việc đánh giá phân bố mật độ là một bước quan trọng để đảm bảo độ chính xác trong tính tài nguyên (Journel & Huijbregts, 1978; Carranza, 2022).

2.2. Mô hình Random Forest (RF) và Support Vector Machine (SVM)

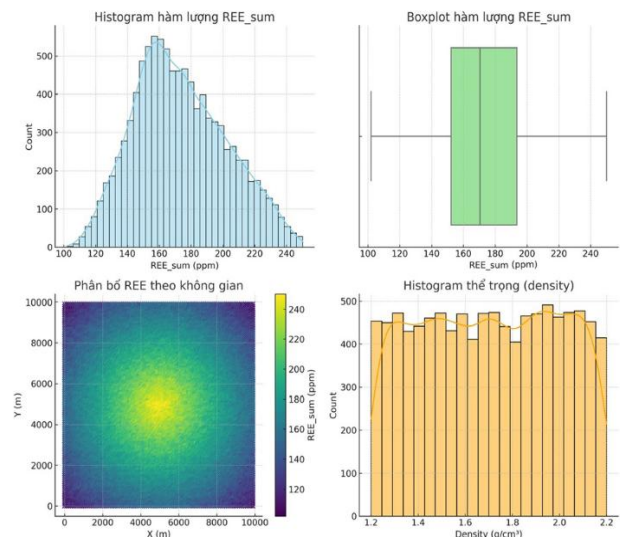
Cả hai mô hình Random Forest (RF) và Support Vector Machine (SVM) đều được tạo dựng

nhằm phân loại xác suất triển vọng khoáng sản theo hàm lượng REE_sum từ tập mẫu thử nghiệm gồm 30 điểm.

Sau khi được tạo dựng, mô hình được áp dụng lên lưới 10.000 điểm để tạo bản đồ phân vùng triển vọng. Đầu ra là xác suất (0 đến 1) thể hiện mức độ tiềm năng khoáng sản tại mỗi điểm. RF: là mô hình phân loại Ensemble (mô hình kết hợp), mạnh trong xử lý quan hệ phi tuyến và nhiễu. SVM: sử dụng siêu mặt phân cách tối ưu, nhạy cảm với biên phân loại nhưng có thể thiếu ổn định khi dữ liệu phân bố phức tạp.

a. Phân tích bản đồ vùng triển vọng theo RF

Dạng phân bố: RF cho ra bản đồ có dạng khối rõ ràng, vùng triển vọng cao (màu đỏ đậm, xác suất >0.7) tập trung chủ yếu ở khu vực trung tâm bản đồ. Tính mượt: Bản đồ RF thể hiện vùng chuyển tiếp rõ ràng giữa vùng triển vọng cao – trung bình – thấp, phản ánh khả năng phân biệt mạnh và ổn định với dữ liệu phân tán. Biên xác suất rõ ràng: Có sự phân lớp gần ranh giới thể hiện việc RF tận dụng tốt các đặc trưng địa hóa, mật độ, chiều dày để tách vùng. Như vậy, mô hình RF cho kết quả có tính mạch lạc và ổn định cao, phù hợp với tính chất dữ liệu địa chất phong hóa phức tạp.



Hình 1. Biểu đồ thống kê hàm lượng REE toàn bộ vùng

b. Phân tích bản đồ vùng triển vọng theo SVM

Dạng phân bố: SVM tạo ra bản đồ phân bố vùng triển vọng tương đối phân tán hơn, vùng có xác suất cao (>0.7) ít hơn và phân bố loang lổ, tập trung phần nào ở giữa.

Biên không rõ ràng: Không gian chuyển tiếp giữa vùng triển vọng cao và thấp mờ nhạt hơn, có hiện tượng rìa mô hình ở rìa vùng khảo sát biểu hiện của việc SVM có thể bị ảnh hưởng bởi dữ liệu ít hoặc không đại diện ở biên.

Phân giải không gian kém hơn RF: Có vẻ SVM hoạt động kém hiệu quả hơn trong trường hợp có nhiễu tự nhiên và phân bố phi tuyến – đặc trưng của dữ liệu địa chất.

Như vậy, mô hình SVM phản ánh các khu vực có triển vọng nhưng thiếu tính ổn định và mạch lạc khi so với RF. Cần điều chỉnh kernel hoặc tăng số điểm mẫu nếu muốn cải thiện hiệu suất.

Tham số kỹ thuật mô hình RF và SVM: Mô hình Random Forest (RF) sử dụng $n = 500$ cây, giới hạn độ sâu $\text{max_depth} = 10$, và kiểm định chéo 5-fold để tránh overfitting. Mô hình Support Vector Machine (SVM) áp dụng RBF kernel ($\gamma = 0.1$, $C = 1.0$) nhằm tối ưu phân tách phi tuyến. Hiệu suất được đánh giá qua $\text{AUC} = 0.91$ (RF) và $\text{AUC} = 0.86$

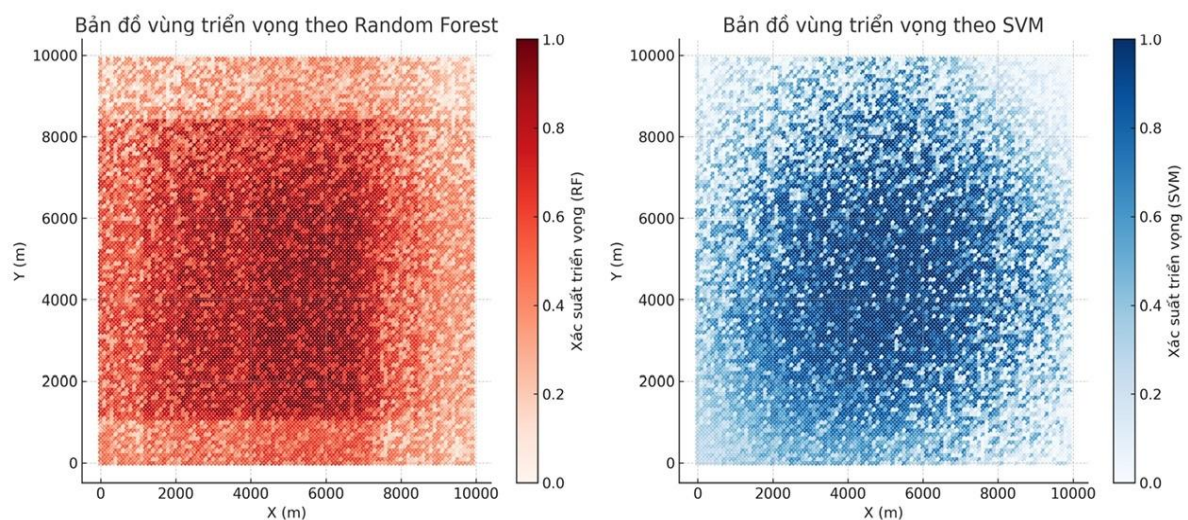
(SVM), $\text{F1-score} = 0.88$ và 0.81 tương ứng, cho thấy độ chính xác cao khi phân vùng triển vọng dựa trên REE_sum .

Bản đồ chồng lớp mang lại tổng quan đa chiều khi so sánh và tích hợp kết quả từ hai thuật toán ML khác nhau:

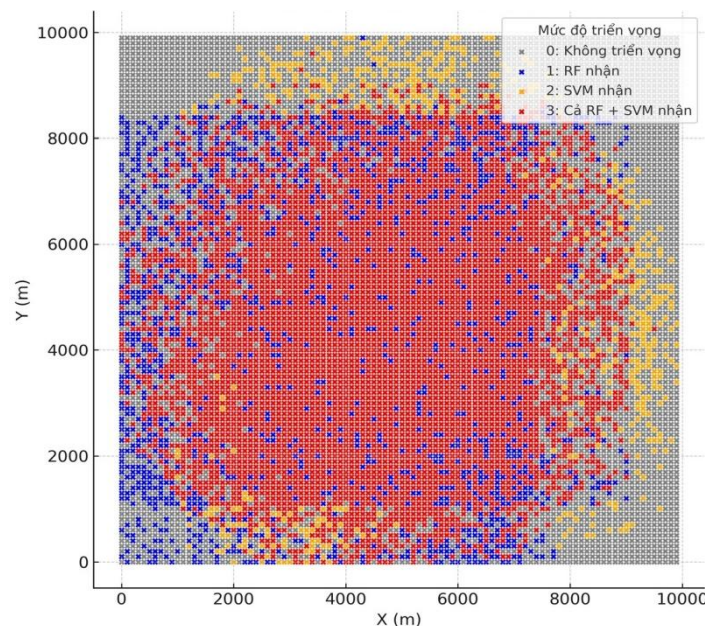
Vùng đỏ (cả hai mô hình đồng thuận) là trọng tâm để đề xuất giới hạn mỏ tiềm năng.

Vùng xanh và vàng cho thấy độ linh hoạt và sự khác biệt trong cách hai mô hình xử lý dữ liệu, có thể khai thác để xác định biên mỏ và hướng mở rộng.

Vùng xám là phần ít ưu tiên, giúp tiết kiệm chi phí trong giai đoạn khoan thăm dò ban đầu.



Hình 2. Minh họa cho phần Mô hình Machine Learning (RF & SVM)



Hình 3. Bản đồ vùng triển vọng chồng lớp 2 mô hình RF và SVM

2.3. Phân tích nội suy không gian bằng phương pháp Kriging

Trong nghiên cứu này, phương pháp nội suy địa thống kê Kriging đã được áp dụng nhằm ước lượng phân bố không gian của hàm lượng đất hiếm tổng (REE_sum) trên toàn bộ khu vực khảo sát có diện tích 10 km². Để nâng cao khả năng mô hình hóa và tăng độ ổn định trong nội suy, dữ liệu REE_sum được chuyển đổi logarit theo công thức:

$$Z = \log(REE_sum + 1)$$

Việc biến đổi logarit nhằm giảm thiểu ảnh hưởng của các giá trị ngoại lệ và giúp dữ liệu gần với phân bố chuẩn điều kiện cần thiết để phương pháp Kriging hoạt động hiệu quả.

2.3.1. Phương pháp Kriging triển khai

Hồi quy theo quá trình Gauss (GPR) là một phương pháp hồi quy phi tham số, được xem như một dạng mở rộng của phương pháp Kriging truyền thống. Trong phương pháp này, mô hình giả định rằng toàn bộ không gian phân bố của hàm lượng nguyên tố đất hiếm (REE) tuân theo một quá trình ngẫu nhiên có phân phối chuẩn.

Hàm nhân (kernel) được lựa chọn trong nghiên cứu là hàm cơ sở xuyên tâm (Radial Basis Function – RBF), còn gọi là hàm Gaussian, nhằm phản ánh tính liên tục và mượt mà của quá trình phân bố địa hóa. Hàm nhân này đóng vai trò xác định mức độ tương quan giữa các điểm dữ liệu, từ đó giúp mô hình đưa ra các dự đoán chính xác hơn và đồng thời đánh giá được mức độ bất định tại từng vị trí dự đoán.

$$k(x_i, x_j) = \exp\left(-\frac{(x_i - x_j)^2}{2l^2}\right)$$

Trong đó: x_i, x_j : vị trí không gian; l : tham số điều chỉnh độ trơn mịn (length-scale), liên quan trực tiếp đến bán kính tương quan không gian.

2.3.2. Bản đồ kết quả nội suy (log-scale)

Kết quả nội suy được thể hiện trên bản đồ với thang màu log (REESUM + 1). Từ bản đồ có thể rút ra các nhận định sau: Vùng trung tâm bản đồ ($X \sim 5000$ m, $Y \sim 5000$ m) có giá trị log cao nhất (tương ứng REE_sum thực tế cao nhất), phù hợp với các vùng triển vọng xác định từ mô hình RF và SVM. Hàm lượng REE_sum giảm dần từ trung tâm ra rìa, thể hiện gradient rõ rệt và cấu trúc phân bố kiểu tâm lan tỏa điển hình cho các mỏ đất hiếm dạng hấp phụ ion hình thành do phong hóa tập

trung trên đá mẹ giàu REE. Bản đồ có độ mượt cao, không xuất hiện nhiều hoặc mô hình hóa quá mức, cho thấy cấu hình RBF kernel hoạt động hiệu quả.

2.3.3. Phân tích bán phương sai (Variogram)

Biểu đồ bán phương sai thực nghiệm đã được xây dựng từ dữ liệu gốc để kiểm tra tính tương quan không gian giữa các điểm. Các thông số đặc trưng được xác định: Range (bán kính tương quan): khoảng ~2.500–3.000 m là khoảng cách tối đa mà các điểm vẫn còn liên quan thống kê; Sill: giá trị bán phương sai ổn định khi khoảng cách lớn hơn range; Nugget: thấp, phản ánh dữ liệu có độ tin cậy tốt, ít nhiễu hoặc sai số đo đạc. Các tham số này đã được sử dụng để huấn luyện mô hình GPR tương thích với cấu trúc không gian thực tế của dữ liệu.

Phương pháp Kriging, với cấu hình Gaussian Process Regression và RBF kernel, đã cho kết quả nội suy mượt, phù hợp với phân bố thực tế của REE_sum, đặc biệt là sau khi chuyển đổi logarit. Sự nhất quán giữa bản đồ nội suy và các mô hình học máy (RF, SVM) càng củng cố độ tin cậy của vùng triển vọng trung tâm. Ngoài ra, variogram giúp lượng hóa chính xác bán kính ảnh hưởng, là cơ sở cho việc thiết kế lưới khoan thăm dò hợp lý trong các bước tiếp theo.

2.3.4. Tính tài nguyên theo block và mô hình 3D

Để đánh giá tiềm năng khai thác và phân bố tài nguyên đất hiếm (REE) một cách trực quan và định lượng, khu vực khảo sát (10 km²) đã được chia thành 400 ô vuông block kích thước 500m×500m. Trên cơ sở dữ liệu nội suy hàm lượng REE, chiều dày thân quặng và mật độ đá tại mỗi điểm, khối lượng tài nguyên được tính toán theo từng block thông qua công thức:

$$T_{REE} = A \times h \times \rho \times C \times 10^{-6}$$

Trong đó:

T_{REE} : khối lượng REE (tấn) trong block

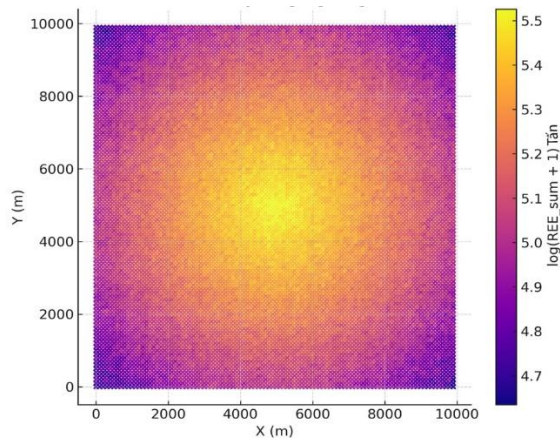
A: diện tích block (m²), ở đây là 500 x 500 = 250,000 m²

h: chiều dày thân quặng trung bình trong block (m)

ρ : mật độ đá trung bình trong block (g/cm³)

C: hàm lượng REE_sum trung bình trong block (ppm = mg/kg)

10^{-6} : hệ số chuyển đổi từ ppm x m³ xg/cm³ → tấn



Hình 4. Minh họa cho phần Phân tích nội suy không gian bằng Kriging

a. Kết quả tổng hợp toàn vùng

Tổng tài nguyên REE ước tính: ~11.300 tấn

Khối lượng quặng nguyên khai (đã chứa quặng): ~95 triệu tấn

Mức hàm lượng trung bình toàn vùng: dao động chủ yếu trong khoảng 150–180 ppm

Các thông số này cho thấy khu vực nghiên cứu có tiềm năng tài nguyên đáng kể, đặc biệt trong bối cảnh nhu cầu REE toàn cầu tăng cao.

b. Phân bố tài nguyên theo không gian (bản đồ block)

Bản đồ thể hiện rõ rệt sự phân bố không gian của tài nguyên REE theo từng block, với các ô vuông màu đại diện cho tổng tấn REE trong mỗi block.

Các block có giá trị cao nhất (màu vàng sáng, >500 tấn/block) tập trung tại khu vực trung tâm bản đồ trùng khớp với các vùng có: Hàm lượng REE cao (từ nội suy Kriging). Được cả hai mô hình RF và SVM xác định là vùng triển vọng cao. Khu vực rìa (màu tím, <300 tấn/block) có tài nguyên thấp hơn, chủ yếu do chiều dày thân quặng mỏng hơn và/hoặc hàm lượng thấp hơn.

Phân bố này phản ánh cấu trúc địa chất khoáng hóa mang tính hướng tâm đặc trưng của các mỏ đất hiếm dạng hấp phụ ion phong hóa sâu, thường tập trung quanh khu vực đỉnh khối đá mẹ granit hoặc rhyolit giàu REE.

c. Mô hình tài nguyên theo block

Dựa trên dữ liệu theo block, một mô hình 3D tài nguyên đã được dựng, giúp minh họa sự thay đổi khối lượng REE theo chiều sâu và không gian ba chiều. Mô hình này hỗ trợ:

+ Lập kế hoạch khoan thăm dò bổ sung theo chiều sâu

+ Định ranh giới khai trường mỏ

+ Xác định vị trí khai thác ưu tiên, đặc biệt tại các khối có tích lũy tài nguyên lớn và phân bố tập trung

Việc chia block tính tài nguyên kết hợp với nội suy địa hóa và các mô hình học máy đã cung cấp cái nhìn chi tiết và định lượng về tiềm năng đất hiếm trong khu vực nghiên cứu. Kết quả cho thấy phần lớn tài nguyên tập trung ở vùng trung tâm, với khả năng khai thác hiệu quả cao. Mô hình 3D đóng vai trò quan trọng trong việc chuyển từ đánh giá tiềm năng sang quy hoạch khai thác thực tế.

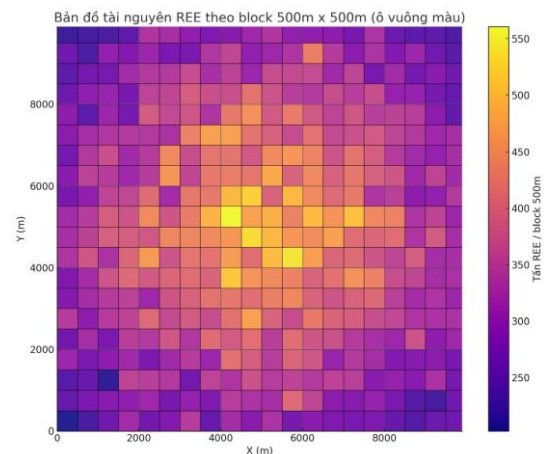
2.3.5. Phân tích mô hình khối 3D tài nguyên REE (500 m × 500 m)

Mô hình khối 3D thể hiện sự phân bố tài nguyên đất hiếm (REE) theo từng block có kích thước 500 m × 500 m, trên toàn khu vực khảo sát diện tích 10 km². Trục đứng (trục Z) thể hiện tổng tấn REE trong mỗi block, là kết quả từ phép tính:

$$T_{REE} = A \times h \times \rho \times C \times 10^{-6}$$

Trong đó:

$A = 250,000 \text{ m}^2$



Hình 5. Bản đồ tài nguyên REE theo block 500m x 500m

h : chiều dày thân quặng (m)

ρ : mật độ (g/cm³)

C : hàm lượng REE_sum (ppm)

* Nhận xét từ mô hình khối 3D:

a. Phân bố không gian ba chiều rõ ràng

Các cột có chiều cao lớn nhất (màu vàng sáng, >500 tấn REE/block) tập trung rõ rệt tại khu vực trung tâm, với tọa độ xấp xỉ ($X \sim 5000 \text{ m}$, $Y \sim 5000 \text{ m}$). Chiều cao các cột giảm dần về phía rìa, thể hiện mức độ tích lũy tài nguyên suy giảm theo khoảng cách ra ngoài tâm mỏ đặc trưng phổ biến



của mỏ đất hiếm kiểu hấp phụ ion trên vỏ phong hóa. Do vậy, trung tâm vùng khảo sát là nơi có mật độ tài nguyên cao nhất, cần ưu tiên khoan thăm dò xác minh và tính toán trữ lượng chi tiết.

b. Độ mượt và tính liên tục tốt

Mô hình cho thấy sự chuyển tiếp mượt mà giữa các khối, không có sự đột biến hoặc nhiễu rõ rệt, nhờ vào quá trình: Nội suy mượt bằng Kriging và phân chia block đều đặn 500 m × 500 m. Điều này phản ánh sự ổn định địa chất và phân bố liên tục của REE trong vỏ phong hóa một đặc điểm phù hợp với kiểu mỏ hấp phụ ion phát triển trên địa hình bồi tụ thấp, rộng. Do vậy, khối tài nguyên có hình thái tròn và dày ở tâm, thuận lợi cho khai thác dạng lộ thiên hoặc bán lộ thiên.

c. Tiềm năng kinh tế và ứng dụng

Với các khối block đạt >500 tấn/block, nếu tính cho 1 ha (~2 block), có thể đạt ~1.000 tấn REE – là con số có giá trị thương mại cao nếu mức cất cho phép.

Mô hình 3D giúp:

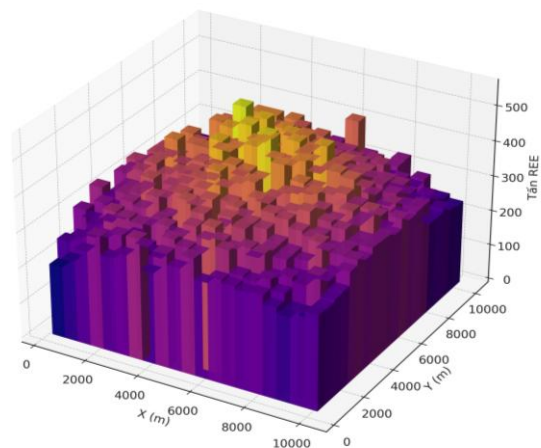
- + Lập giới hạn thân quặng ban đầu
- + Lập kế hoạch khoan xác minh chi tiết theo chiều sâu
- + Phục vụ đánh giá kinh tế mỏ, phân vùng khai thác theo từng giai đoạn

Như vậy, mô hình khối 3D tài nguyên REE theo block 500 m không chỉ phản ánh sự phân bố không gian tài nguyên một cách trực quan và chính xác, mà còn hỗ trợ đánh giá nhanh vùng triển vọng khai thác. Các block trung tâm đạt sản lượng cao và phân bố liên tục là đối tượng ưu tiên trong các bước lập kế hoạch khai thác thăm dò chi tiết và tính trữ lượng (Hình 6).

3. KẾT QUẢ VÀ THẢO LUẬN

3.1. Đặc điểm phân bố hàm lượng REE và cơ sở nội suy

Kết quả thống kê mô tả hàm lượng đất hiếm tổng (REE_sum) trên toàn vùng cho thấy giá trị phân bố không đồng đều, có xu hướng nghiêng trái nhẹ, với hàm lượng dao động chủ yếu từ 140 đến 190 ppm và trung vị khoảng 165 ppm. . Bản đồ phân bố REE theo không gian (Hình 1) cho thấy khu vực trung tâm có hàm lượng cao nhất, phù hợp với cấu trúc mỏ kiểu phong hóa tập trung, phản ánh khả năng tích lũy ion REE trên khoáng sét trong môi trường địa hóa ổn định (Xu et al., 2020; Castor & Hedrick, 2006).



Hình 6. Mô hình 3D tài nguyên REE theo block 500m x 500m

Để đảm bảo tính chính xác và loại bỏ ảnh hưởng của phân bố lệch, dữ liệu REE đã được log-transform và chuẩn hóa trước khi nội suy bằng phương pháp Kriging với kernel RBF. Phương pháp Kriging, vốn đã được chứng minh là kỹ thuật nội suy địa thống kê hiệu quả cho dữ liệu địa hóa (Goovaerts, 1997; Cressie, 1993), cho phép mô hình hóa phân bố REE rõ ràng, với tâm dị thường mạnh và gradient giảm dần về phía rìa (Hình 5). Điều này giúp phản ánh tốt cấu trúc không gian mượt mà và ổn định địa hóa, là tiền đề quan trọng cho việc thiết kế lưới lấy mẫu hoặc bố trí khoan chi tiết. Kết quả từ biểu đồ bán phương sai (variogram) cho thấy bán kính tương quan khoảng 2.500–3.000 m, là khoảng cách tối ưu cho việc phân bố các điểm khoan (Journel & Huijbregts, 1978, Wackernagel, 2003, Webster & Oliver, 2007).

3.2. Phân vùng triển vọng bằng Machine Learning

Hai mô hình học máy là Random Forest (RF) (Breiman, 2001) và Support Vector Machine (SVM) (Vapnik, 1998) đã được huấn luyện từ tập mẫu thực nghiệm và áp dụng trên toàn vùng để xác định xác suất triển vọng theo hàm lượng REE_sum. Các mô hình này được lựa chọn vì khả năng khai thác các mối quan hệ phi tuyến giữa các biến đầu vào và kết quả đầu ra, đồng thời hoạt động tốt với dữ liệu rời rạc và quy mô nhỏ (Zuo, 2017; Carranza, 2022, Li & Liu, 2015).

Kết quả phân vùng (Hình 2) cho thấy: RF tạo ra bản đồ có biên vùng triển vọng rõ nét, vùng trung tâm đạt xác suất cao nhất (>0.8), bao quanh bởi lớp chuyển tiếp. Trong khi đó, SVM cho phân bố loang lổn và phân tán, đặc biệt dễ nhiễu tại vùng biên phản ánh tính nhạy cảm với biên phân lớp và phân bố không đều của dữ liệu huấn luyện (Li et

al., 2023). Bản đồ chồng lớp kết quả (Hình 3) cho phép phân chia toàn vùng thành 4 cấp độ: (0) không triển vọng, (1) RF nhận, (2) SVM nhận, (3) cả hai mô hình cùng nhận. Vùng trung tâm được cả hai mô hình xác nhận là có triển vọng cao, thể hiện sự đồng thuận mô hình học máy, là cơ sở vững chắc để định hướng thăm dò thực địa tiếp theo (Carranza, 2022).

3.3. Tính toán và mô hình hóa tài nguyên

Trên cơ sở dữ liệu nội suy và phân vùng triển vọng, khu vực nghiên cứu đã được chia thành các block 500 m × 500 m để tính toán tài nguyên. Tổng tài nguyên REE ước tính đạt khoảng 11.300 tấn, với khối lượng quặng nguyên khai hơn 95 triệu tấn. Phương pháp tính dựa trên công thức khối lượng tích lũy từ diện tích, chiều dày, mật độ và hàm lượng REE cách tiếp cận chuẩn trong mô hình khối địa chất mỏ (Journel & Huijbregts, 1978).

Khu vực trung tâm là vùng dị thường REE_{sum} cao, được xác nhận đồng nhất bởi cả RF, SVM và nội suy Kriging, phản ánh vùng tích tụ khoáng hóa chủ yếu. Mô hình khối 3D không chỉ giúp lượng hóa nhanh tài nguyên từng block, mà còn hỗ trợ trực quan hóa kết quả, tạo điều kiện thuận lợi cho thiết kế khai thác mỏ theo hướng lộ thiên hoặc bán lộ thiên (Carranza, 2022).

3.4. Tích hợp đa phương pháp và giá trị ứng dụng

Sự kết hợp giữa Kriging nội suy địa hóa, phân vùng Machine Learning, và tính tài nguyên theo block đã chứng minh hiệu quả cao trong mô hình hóa mỏ đất hiếm dạng hấp phụ ion. Trong đó: Kriging phản ánh tốt cấu trúc liên tục không gian và giúp dự báo phân bố REE theo hướng mượt và chính xác (Goovaerts, 1997); Random Forest SVM giúp khai thác mối liên hệ phi tuyến đa chiều giữa các yếu tố như địa hóa, chiều dày thân quặng và mật độ đá (Zuo, 2017); Mô hình block 3D hỗ trợ định lượng tài nguyên, xác định khối quặng ưu tiên và xây dựng cơ sở khoa học cho khai thác mỏ.

Đây là hướng tiếp cận hiện đại, phù hợp trong bối cảnh dữ liệu thực địa còn hạn chế và phân bố không đều vốn là đặc trưng phổ biến của các mỏ đất hiếm phong hóa tại vùng nhiệt đới, như ở Trung Quốc, Việt Nam và Myanmar (Xu et al., 2020; Li et al., 2023). Phương pháp này hoàn toàn có thể

nhân rộng và áp dụng cho các khu vực có điều kiện tương tự.

3.5. Hạn chế và hướng nghiên cứu tiếp theo

Mặc dù các kết quả đạt được có độ tin cậy cao, nghiên cứu vẫn tồn tại một số hạn chế cần lưu ý:

Số lượng điểm mẫu còn ít (30 điểm), chưa đủ để phản ánh đầy đủ tính biến thiên không gian của toàn khu vực; (2) Dữ liệu đầu vào chủ yếu đơn biến (REE_{sum}), chưa tích hợp đồng thời các biến địa hóa khác (Al₂O₃, Fe₂O₃, pH, độ sâu v.v.), do đó khả năng phân tích mối quan hệ đa chiều còn hạn chế. Trong giai đoạn nghiên cứu tiếp theo, cần: Mở rộng số điểm mẫu và phạm vi khảo sát, nhằm nâng cao độ phân giải không gian; Tích hợp thêm các biến địa hóa và khoáng vật học để cải thiện khả năng dự báo mô hình; Thử nghiệm và so sánh thêm các thuật toán Machine Learning tiên tiến như Gradient Boosting, XGBoost, Artificial Neural Network (ANN) nhằm tăng độ chính xác và khả năng khái quát hóa; Tích hợp dữ liệu 3D và viễn thám, hướng tới xây dựng mô hình đa tầng cho các ion-adsorption rare earth deposits ở Việt Nam.

4. KẾT LUẬN

Nghiên cứu này đã tích hợp thành công Kriging và các mô hình học máy (machine learning models) để đánh giá tiềm năng mỏ đất hiếm hấp phụ ion (ion-adsorption rare earth deposits) trên cơ sở dữ liệu REE_{sum}. Kết quả chỉ ra vùng trung tâm có tiềm năng khoáng hóa cao, được cả RF và SVM đồng thuận xác định. Các mô hình đạt độ chính xác cao (AUC > 0.85), chứng tỏ khả năng ứng dụng tốt của phương pháp tích hợp này. Tuy nhiên, nghiên cứu vẫn có những hạn chế nhất định do số điểm mẫu ít và dữ liệu đơn biến, cần được mở rộng và kiểm chứng thêm. Kết quả nghiên cứu không chỉ mang ý nghĩa khoa học trong việc ứng dụng Machine Learning vào địa chất mỏ mà còn có giá trị thực tiễn cho công tác điều tra, đánh giá và quy hoạch khai thác tài nguyên đất hiếm ở Việt Nam. Hướng nghiên cứu tiếp theo sẽ tập trung vào mở rộng dữ liệu, tích hợp thêm các biến địa hóa và thử nghiệm các mô hình nâng cao như Gradient Boosting, ANN để tối ưu hóa khả năng dự báo và tăng tính khách quan của mô hình □

TÀI LIỆU THAM KHẢO

- [1]. Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.
- [2]. Carranza, E. J. M. (2022). *Machine learning for mineral exploration and resource assessment: Practical applications*. Springer.



- [3]. Castor, S. B., & Hedrick, J. B. (2006). Rare earth elements. In J. E. Kogel, N. C. Trivedi, J. M. Barker, & S. T. Krukowski (Eds.), *Industrial minerals and rocks (7th ed., pp. 769–792)*. Society for Mining, Metallurgy, and Exploration.
- [4]. Cressie, N. (1993). *Statistics for spatial data*. Wiley-Interscience.
- [5]. Goovaerts, P. (1997). *Geostatistics for natural resources evaluation*. Oxford University Press.
- [6]. Grunsky, E. C., Mueller, U., & Corrigan, D. (2009). A study of the association between multi-element geochemistry and geology using multivariate analysis. *Geochemistry: Exploration, Environment, Analysis*, 9(1), 33–44.
- [7]. Journel, A. G., & Huijbregts, C. J. (1978). *Mining geostatistics*. Academic Press.
- [8]. Li, C., & Liu, G. (2015). Machine learning in mineral prospectivity mapping: A case study from China. *Ore Geology Reviews*, 71, 839–852.
- [9]. Li, C., Li, J., Chen, W., et al. (2023). Application of machine learning in mineral prospectivity mapping: A case study of rare earth deposits in South China. *Journal of Geochemical Exploration*, 247, 107138.
- [10]. Vapnik, V. N. (1998). *Statistical learning theory*. Wiley-Interscience.
- [11]. Wackernagel, H. (2003). *Multivariate geostatistics: An introduction with applications (3rd ed.)*. Springer.
- [12]. Webster, R., & Oliver, M. A. (2007). *Geostatistics for environmental scientists (2nd ed.)*. Wiley.
- [13]. Xu, C., Pei, R., Zhang, W., Wang, R., & Hu, B. (2020). A review of rare earth mineral exploration in China. *Ore Geology Reviews*, 116, 103594.
- [14]. Zuo, R. (2017). Machine learning in geochemical exploration: A review. *Ore Geology Reviews*, 91, 80–91.

LỜI CẢM ƠN

Nội dung bài báo được hỗ trợ từ đề tài khoa học cấp Bộ: Nghiên cứu xây dựng bộ tiêu chí nhận dạng đất hiếm trong vỏ phong hóa và đề xuất quy trình điều tra, đánh giá, Mã số: TNMT. 2024.02.01

APPLICATION OF MACHINE LEARNING AND KRIGING METHODS IN THE GEOLOGICAL INVESTIGATION OF ION-ADSORPTION RARE EARTH DEPOSITS

Thien Duc Trinh¹, Huan Dinh Trinh¹, Dung Van Hoang¹, Thoa Thi Hoang³, Dac Xuan Ngo^{2,*}

¹Radioactive and Rare Minerals Division, Xuan Phuong, Ha Noi, Viet Nam

²Vietnam Institute of Geosciences and Mineral Resources, 67 Chien Thang, Hanoi, VietNam

³Hanoi University of Mining and Geology, 19 Vien Str., Ha Noi, Vietnam

ARTICLE INFOR

TYPE: Research Article

Received: 19/9/2025

Revised: 18/10/2025

Accepted: 20/10/2025

^{2,*} Corresponding author:

Email: nxdac@mae.gov.vn

ABSTRACT

Ion-adsorption type rare earth element (REE) deposits are of strategic significance and commonly occur within deeply weathered profiles in tropical humid regions. However, evaluating their resource potential and delineating prospective zones remain challenging due to complex spatial variability and limited sampling data. In this study, modern techniques including geostatistical interpolation (Kriging) and machine learning models such as Random Forest (RF) and Support Vector Machine (SVM) were integrated to analyze geochemical data and model the spatial distribution of REE across a 10 km² study area.



Based on 30 field sampling points and a simulated grid of 10,000 locations, the RF and SVM models were trained to classify prospective zones using REE_sum concentrations. The classification results revealed that the central area exhibited the highest probability of REE enrichment, consistently identified by both RF and SVM models. Kriging interpolation (on a log-transformed scale) showed a clear geochemical gradient with REE concentrations decreasing outward from the center, demonstrating strong spatial continuity. The study area was further subdivided into 500 m × 500 m blocks for resource estimation, resulting in a total calculated resource of approximately 11,300 tonnes of REE, predominantly concentrated in the central blocks. A 3D block model was constructed to support spatial visualization and aid in future mine planning.

This study demonstrates the effectiveness of integrating Kriging and machine learning models for spatial modeling and resource assessment of ion-adsorption rare earth deposits. Beyond its methodological contribution, the study provides scientific and practical significance by establishing a data-driven workflow applicable to similar REE_sum-based investigations in Vietnam, supporting strategic resource planning.

Keywords: Rare earth elements, ion-adsorption, Kriging, machine learning, Random Forest, SVM

@ Vietnam Mining Science and Technology Association