



# GEOLOGY DIFFERENTIATION BY APPLYING UNSUPERVISED MACHINE LEARNING OF GRAVITY FIELD PROPERTIES IN CENTRAL AREA OF VIETNAM

Hong Phan THI<sup>1\*</sup>, Phuong Do MINH<sup>2</sup>, Huu Tran VAN<sup>1</sup>

## Abstract

This paper presents an application of the unsupervised learning method (k-means clustering) analysis technique to a four-variate gravity anomaly attribute parameter set distributed over a 24240km<sup>2</sup> square in the central region of Vietnam.

The dataset consists of the gravity anomalous field attribute, the horizontal gradient attribute, the variance attribute, and the tracing coefficient of the gravity anomalous axis. Seven, nine, eleven and thirteen compact and separated clusters have been identified.

The results of similarities clustering of the physical properties of the gravity field coincide with the regional geological map.

The research results show that the Southwest, the Center, and the South of the study area have complex changing

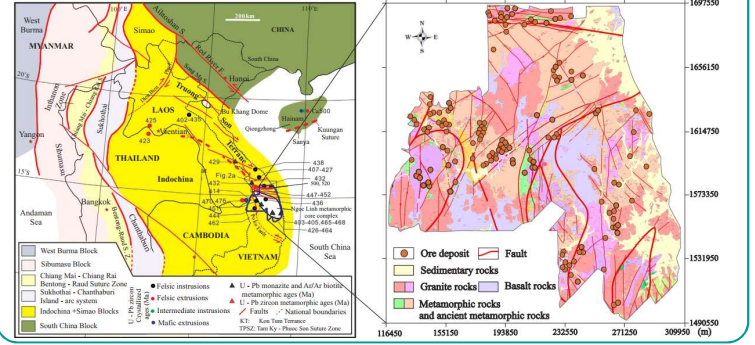
physical properties, this result reflects the complicated tectonic activities in these areas with the presence of crumpled and fractured rock layers in different directions and these locations are the potential places to form endogenous mineral deposits of magma origin.

The Northwest, the North, and the East parts of the research area witness negligible changes in the field's physical properties, reflecting the stability of the soil and rock layers in this area, with the direction of extending structure from the Northwest to the Southeast.

For areas with little geophysical information, the use of correlated physical attributes of the same geophysical field is a reliable input data set of the k-means clustering algorithm with the generalized Mahalanobis distance, this result is a good initial approach to explore the most obvious structures in the data.

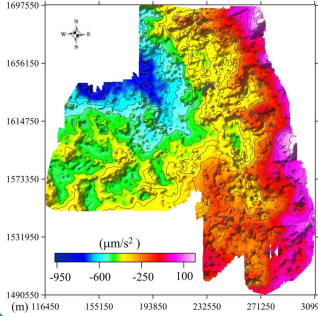
## Local geology

Simplified tectonic map of Southeast Asia, Geological map and distribution location of regional ore points in central region of Vietnam have shown this place of development and interference of large blocks such as Trung Son terrace, Tam Ky - Phuoc Son suture zone, and Ngoc Linh metamorphic core complex lifting block. In which, deep faults along the Northwest-Southeast (Po Ko fault), Northeast-Southwest and sub-meridian directions are the boundary between the abovementioned structures.



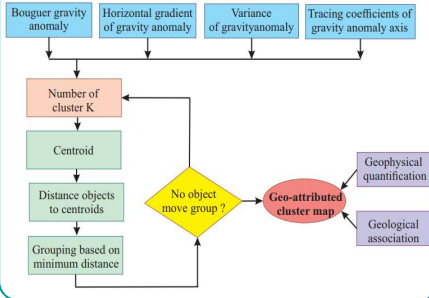
## Gravitational anomaly datasets

The data is used is the ground-measured gravitational anomaly data of full Bouguer correction at the scale of 1:100.000 and the accuracy of 1±25 μm/s<sup>2</sup>. Bouguer correction density was selected as 2,67g/cm<sup>3</sup> and Picivanco's method was used to correct the terrain. The value gravitational anomaly varies from -1000±200μm/s<sup>2</sup>.

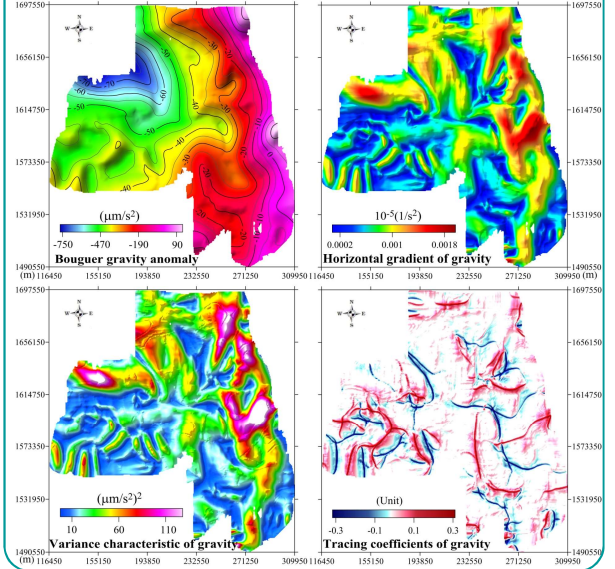


## Methods

In this study, we only focus on the most popular algorithm in unsupervised learning techniques, which is the k-means clustering algorithm. In the k-means clustering, the Mahalanobis distance in Hilbert spaces is used as the distance between the individual parameters and takes into account the relationships in the feature space of the field. Input data sets including: gravity anomaly, horizontal gradient, variance characteristic, tracing coefficients of gravitational anomaly axis.

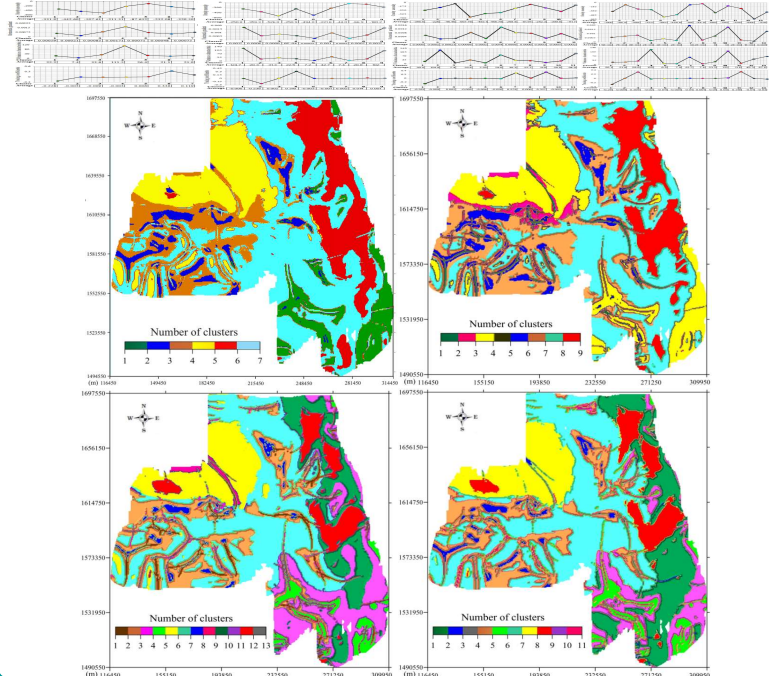


## Input data of k-means method



## The Results of applying the k-means clustering algorithm

Centroidal coordinates of the four-variate geophysical classes have been deduced by the k-means analysis. K-means clustered four-variate geophysical classes model section across the common survey area obtained by combining the univariate regional gravity anomaly field, horizontal gradient, variance and tracing coefficients, results division into 07, 09, 11, and 13 clusters with close characteristics of the properties of the gravity anomaly field in the central region of Vietnam.



## Discussion

The results division into 07, 09, 11, and 13 clusters shown that the study area exists the complex cleavage physical properties in the Southwest, the South and the Center of the area, compared with the geological structure map, this area belongs to the intracontinental rift system after the Mesozoic collision, that is the result of post-extension and collision between the Indochinese plate and Sibumasu. Thus, the results of applying the k-means clustering algorithm are well-matched with the geological data.

In the North, Northwest, and East of the area, the physical attributes of the field shown little or no change, indicating that these regions are quite stable with the Western structural direction extending to the Northwest-Southeast. Because the East of the study area is adjacent to the East Sea, this result is completely consistent with geological laws.

Results shown that the terracotta tectonic activity still exists in the Southwest and the South of the study area, which proves that this tectonic activity is regional and they act as channels to pump magma blocks from deep to the surface, penetrate the near-surface sediments and form deposits of deep-hidden mineral mines.

## Conclusion

The application of the k-means clustering algorithm to the central area of Vietnam has shown that the clustering approach can be used as an effective method to describe geological differences.

Correlation-based clustering successfully identified geological units using input data including four physical attributes of the same gravity anomaly. The results were divided into 07, 09, 11, and 13 clusters, having close characteristics of the physical attributes of gravity completely coinciding with the tectonic geology in the area.

For areas with little geophysical information, the use of correlated physical attributes of the same geophysical field is a reliable input data set of the k-means clustering algorithm with the generalized Mahalanobis distance, this result is a good initial approach to explore the most obvious structures in the data. This work shows that the application of unsupervised machine learning is feasible in new areas with little geophysical information to identify geological differences.