



Segmentation of Homogeneous Regions of Gravity Field Properties by Machine Learning Method in Central Area of Vietnam

Hong Phan THI^{1)*}, Phuong Do MINH²⁾, Huu Tran VAN¹⁾

¹⁾ HUMG Hanoi University of Mining and Geology, Faculty of Petroleum and Energy, Hanoi, Vietnam

²⁾ Geophysical Division, General Department of Geology and Minerals, Hanoi, Vietnam

* Corresponding author: phanthihong@humg.edu.vn

<http://doi.org/10.29227/IM-2023-02-17>

Submission date: 16-08-2023 | Review date: 11-09-2023

Abstract

This paper presents the results of applying the unsupervised learning method (K-means clustering) on the gravity anomaly field in the central region of Vietnam to separate the research area into different clusters, which are homologous in physical properties. In order to achieve the optimal results, the input parameter plays an important role. In this paper, we chose 04 input attributes including the gravity anomalous field attribute, the horizontal gradient attribute, the variance attribute, and the tracing coefficient of the gravity anomalous axis. The obtained results have shown that the research area could be divided into 7 clusters, 9 clusters, 11 clusters, and 13 clusters with close characteristics of the physical properties of the gravity field. 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. The clustering results according to the K-means unsupervised learning algorithm in central Vietnam initially increase the reliability of the decisions of geologists and geophysicists in interpreting the geological structure and evaluating the origin of deep-hidden mineral deposits in the area.

Keywords: K-means, unsupervised learning method, gravity field, central area of Vietnam, COSCAD 3D

1. Introduction

Central Vietnam is the place of development and interference of large blocks such as Truong Son sticky belt, Tam Ky - Phuoc Son hatch, and Kon Tum lifting block. In which, deep faults along the Northwest-Southeast, Northeast-Southwest and minor meridians are the boundary between the above-mentioned structures [4, 20, 21]. The research area is characterized by narrow valleys and treacherous divisions with complex topography including many mountain ranges >700m high [7] (Fig.1a) which extend to the Northeast, Northwest, and the Meridians. These are favorable prerequisites to forming endogenous mineral deposits of magma origin, and the identification of geological structures related to the location of endogenous mineral deposits is an important mission and top priority in the coming years. Furthermore, all of this information is characterized by different residual density values and can be displayed on the gravity anomalous field values measured from ground-based gravity survey (Fig.1b), and these field values are the consequence of overlapping of gravity anomalous fields from many geological objects in the subsurface, in combination with deep geological structures, which are needed to characterize, so they are random. These factors make the gravity anomalous field caused by objects with high density often complicated, which leads to difficulty in the process of reasoning.

Along with the development of technology 4.0, there have been many applications of unsupervised learning algorithms in the processing of potential field data [1, 3]. Computerization of the exploration process allows the application of new algorithms, modern algorithms, and statistical algorithms in

the processing and interpretation of geological and geophysical data. Paying special attention to characteristics of geological and geophysical information, selecting the optimal set of features to cluster with uniform physical properties, and evaluating the correlation between different physical attributes within the same cluster [5, 9, 14].

In this study, we apply an unsupervised learning algorithm (K-means clustering) on the gravity anomalous field data in the central region of Vietnam to divide the area into homogeneous regions of the field's physical attributes (attributes: gravity field, horizontal gradient value, statistics, anomaly axis tracing). This result will help to quickly identify homogenous layers (saving time, and cost...) to provide information for geologists to have an intuitive view of the study area.

2. Data and methods

2.1. Data used

The data used in this paper are the ground-measured gravitational anomaly data of full Bouguer correction at the scale of 1:100.000 and the accuracy of 0,1-0,25 mGal [10, 11, 12, 13]. Bouguer correction density was selected as 2,67g/cm³ and Picivanco's method was used to correct the terrain [2].

Looking at Figure 1(b), it can be seen that the Bouguer gravity anomaly changes from -100 to 20 mGal, and the gravity anomaly tends to increase gradually from Northwest - Southeast toward the sea. The Northwest of the study area is characterized by a strong negative gravity anomaly with amplitude ranging from -100 to -55 mGal, and the East and Southeast by a range of gravity anomaly values varying from -30 to 20 mGal. There are local anomalies with amplitude from -30 to

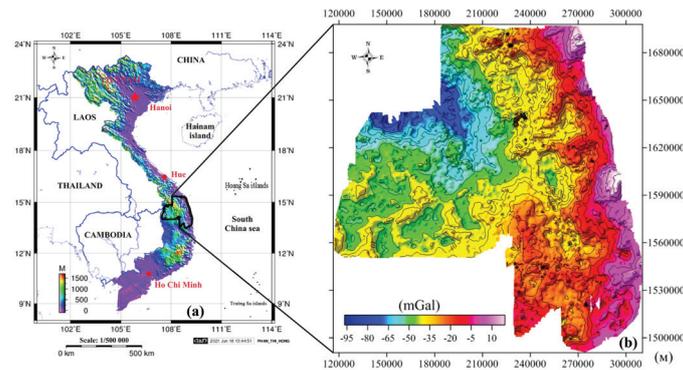


Fig. 1. Location and topography elevation from 10-1600 m of the study area (a) and Map of the full Bouguer gravitational anomaly field of the study area at a scale of 1:100.000, where the value varies from -100-20 mGal (b)

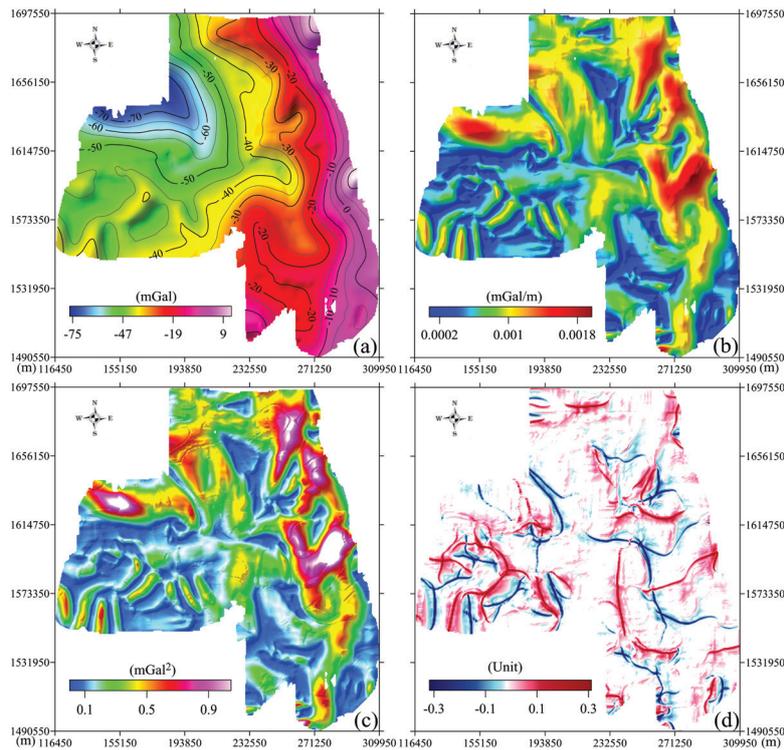


Fig. 2. The regional anomaly field after removing the shallow local anomaly field (a); Horizontal Gradient of gravity anomaly in the research area (b); regional anomaly variance (c); tracing coefficients of anomaly axis in the research area (d)

-15 mGal in the North, Southwest, Center, and South of the study area.

2.2. Methods

Nowadays, there are many unsupervised learning algorithms applied in processing and analyzing of geophysical data [1, 3]. However, in this study, we only focus on the most popular algorithm in unsupervised learning techniques, which is the K-means clustering algorithm [1, 3, 8, 9]. In the K-means algorithm, the Mahalanobis distance is used as the distance between the individual parameters and takes into account the relationships in the feature space of the field. The K-means algorithm automatically divides the data of the central region of Vietnam into different clusters, hence the attribute data (attributes: gravity field, horizontal gradient value, statistics, anomaly tracing) are in the same cluster with similar properties.

2.2.1. K-means method

The k-means clustering algorithm is an unsupervised learning method with a set of points with close features. The advan-

tage of this algorithm is simple and time-efficient computation. The block diagram of the algorithm is as follows [5, 6, 9, 17]:

- For all p signals, determining the standard deviation σ_i , the minimum f_{\min}^i and the maximum f_{\max}^i at $i = 1, 2, 3, \dots, p$ value of the signal;
- Each signal f_i ($i=1, 2, \dots, p$) is normalized to correspond to the standard deviation: $\alpha_i = \frac{f_i}{\sigma_i}$;
- A vector k of size p is randomly selected for the initial center of the class $C_j = \{c_{1j}, c_{2j}, \dots, c_{pj}\}$; $j=1, 2, \dots, k$ and the individual randomly selected components of each vector, satisfying the inequality: $\frac{y_{\min}^i}{\sigma_i} < c_{ij} < \frac{y_{\max}^i}{\sigma_j}$;
- Classifying the normalized signs of the original network into classes, and at each point of the network $x = \{x_1, x_2, \dots, x_p\}$ related to layer m, if the distance from the center of this layer to the point is minimal $r_{\min} = \text{Min} \{x_1, x_2, \dots, x_p\}$;
- According to the classification results, new vectors are determined $C^n = \{c_1^n, c_2^n, \dots, c_p^n\}$; $j=1, 2, \dots, k$. Each vector component m is the average value $c_m^n = \frac{1}{n_m} \sum_{j=1}^{n_m} x_j^m$, $i = 1, 2, \dots, p$

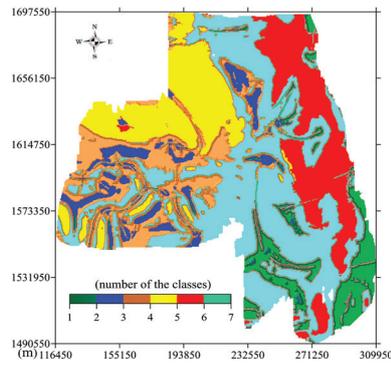


Table 1: Physical property average values for 07 homogenous clusters according to unsupervised machine learning (K-means clustering)

Number of the homogenous clusters	Gravitational anomaly field Δg (mGal)	Horizontal gradient of gravity anomaly field (mGal/m)	Variance characteristic of the gravity anomaly field (mGal ²)	Gravity anomaly tracing coefficient
1	-33.190	$0.594 \cdot 10^{-3}$	0.222	-0.228
2	-44.269	$0.21 \cdot 10^{-3}$	0.074	-0.003
3	-50.761	$0.849 \cdot 10^{-3}$	0.336	-0.009
4	-21.131	$0.171 \cdot 10^{-2}$	1.115	-0.003
5	-8.7632	$0.713 \cdot 10^{-3}$	0.264	0.006
6	-23.208	$0.477 \cdot 10^{-3}$	0.211	0.336
7	-32.959	$0.751 \cdot 10^{-3}$	0.259	0.119

Fig. 3. The results of automatic division into 07 clusters with close characteristics of the properties of the gravity anomaly field in the central region of Vietnam

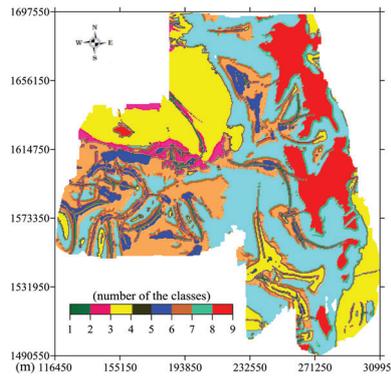


Table 2: Physical property average values for 09 homogenous clusters according to unsupervised machine learning (K-means clustering)

Number of the homogenous clusters	Gravitational anomaly field Δg (mGal)	Horizontal gradient of gravity anomaly field (mGal/m)	Variance characteristic of the gravity anomaly field (mGal ²)	Gravity anomaly tracing coefficient
1	-25.082	$1.746 \cdot 10^{-3}$	1.247	-0.006
2	-35.927	$0.569 \cdot 10^{-3}$	0.203	-0.238
3	-52.596	$0.815 \cdot 10^{-3}$	0.313	0.002
4	-25.094	$0.588 \cdot 10^{-3}$	0.225	0.318
5	-49.396	$0.221 \cdot 10^{-3}$	0.074	-0.003
6	-22.158	$0.981 \cdot 10^{-3}$	0.427	-0.001
7	-4.400	$0.221 \cdot 10^{-3}$	0.074	0.002
8	-6.686	$0.667 \cdot 10^{-3}$	0.208	0.014
9	-18.015	$1.672 \cdot 10^{-3}$	1.02	0.001

Fig. 4. The results of automatic division into 09 clusters with close characteristics of the properties of the gravity anomaly field in the central region of Vietnam

calculated in n_m points, falling into class m after classification, performed in the previous step of the algorithm;

- In the selected index, the distance between the old center and the new center of the layer is estimated $\vec{R}_j(C, \vec{C}^n)$, $j=1,2,\dots,k$;
- If at least one of the k classes, the distance is greater than the pre-selected value $\vec{R}_j > \varepsilon$, then the center of the old vector is assigned the new value $C_j = \vec{C}_j^n$, $j=1,2,\dots,k$, and the procedure is repeated from the step where the observations are classified. In the opposite case, the inequality $\vec{R}_j(\vec{C}, \vec{C}^n) < \varepsilon$ is satisfied for all classes, the result of the last iteration is the last step of the iteration.

This algorithm converges after a finite number of iterations (converging fairly quickly), the main disadvantage of the algorithm is the need to know the number of clusters in advance, and not fully consider the associations of the field attribute space. To overcome the limitations of the algorithm, in this study, we determined the number of clusters according to the geological data on the surface (Tran and Nguyen, 2008), and used the Mahalanobis distance as the spatial distance (Nikitin and Petrov, 2008; Hong, 2020).

2.2.2. The input parameter of the K-means clustering algorithm

The input parameter plays a very important role in the unsupervised learning algorithm, in this paper we chose 4 attributes: gravity anomalous field attribute, variance, horizontal gradient, and anomaly axis tracing coefficient. In the study, we have selected the optimal set of features to cluster

and evaluate the correlation between the attribute features in the same cluster.

- Horizontal gradient of gravity anomaly

The gravity anomalous horizontal gradient value is calculated according to the following formula [5, 6, 9]:

$$GN = \sqrt{\left(\frac{\Delta g}{\partial x}\right)^2 + \left(\frac{\Delta g}{\partial y}\right)^2} \quad (1)$$

in which: GN is the gravity anomalous horizontal gradient value (mGal/m); Δg is the gravity anomaly value; ∂x , ∂y are derivatives in the direction of $0x$, $0y$.

The maximum value of the gravity anomalous horizontal gradient emphasizes the boundaries of anomalous features (fault systems, boundaries of geological features....); The horizontal gradient field highlights the boundaries of anomalies of different amplitudes, allowing visualization of the boundaries of all anomalies simultaneously.

- Evaluation of the statistical characteristics of the gravity anomaly field

In the statistical characteristics of the field (mathematical expectation, variance, eccentricity, asymmetrical deviation of the field...), we chose the field's variance value and according to the formula [15, 16, 17, 18, 19]:

$$\sigma^2 = \frac{1}{n} \cdot \frac{1}{m} \sum_{i=1}^n \sum_{j=1}^m (X_{ij} - \bar{X})^2 \quad (2)$$

where: σ is the variance of the gravity anomaly field (mGal²), X_{ij} is the gravity anomaly value (mGal), \bar{X} is the value of average gravity anomaly (mGal).

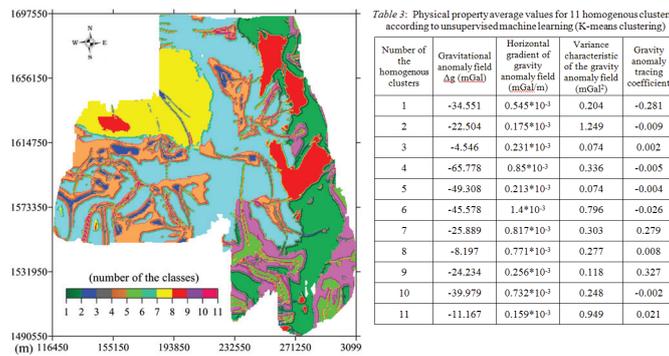


Fig. 5. The results of automatic division into 11 clusters with close characteristics of the properties of the gravity anomaly field in the central region of Vietnam

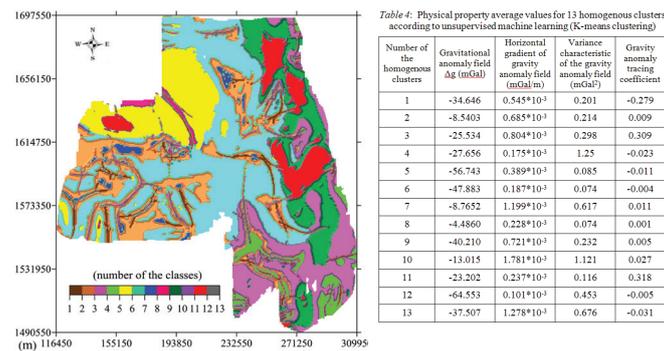


Fig. 6. The results of automatic division into 13 clusters with close characteristics of the properties of the gravity anomaly field in the central region of Vietnam

The maximum value of the variance carries the information of heterogeneous boundaries (fault systems, boundaries of geological features...).

- Tracing anomaly axis

The program is designed to trace anomalies of different energies and different directions [5, 6, 9], with the maximum value (equal to 1) corresponding to the axis of the positive anomaly, the minimum value (equal to -1) corresponding to the axis of the negative anomaly, and the value (equal to 0) corresponding to no anomaly.

3. Results and discussion

3.1. Result of processing input parameters of unsupervised learning method

As shown in the above analysis, in this study, we choose 4 input parameters to put into the K-mean clustering unsupervised learning algorithm, including regional gravity anomaly field (removed local field near the surface) (Fig.2a), applying formula (1) to evaluate the horizontal gradient characteristics of gravity anomaly, the results are shown in Figure 2(b); applying formula (2) to evaluate the statistical characteristic of gravity anomaly field, results are shown in Figure 2(c), and finally the results of tracing the anomaly axis (Fig.2c).

3.2. Result of applying the unsupervised learning method

Applying the K-means unsupervised learning algorithm with 4 input parameters, we got the results of automatic clustering of the central region of Vietnam into 07 clusters (Fig. 3), 09 clusters (Fig. 4), 11 clusters (Fig.5), 13 clusters (Fig.6), these clusters have close features on properties of gravity anomaly field, horizontal gradient value, variance and tracing

coefficients of anomaly axis on Table 1, Table 2, Table 3, and Table 4.

The results of automatic division into 07 clusters (Fig.3), 09 clusters (Fig.4), 11 clusters (Fig.5), and 13 clusters (Fig.6) show that the study area exists the complex cleavage physical properties in the Southwest, the South and the center of the area, compared to 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 [4, 21]. Thus, the results of applying the K-means unsupervised learning method are well-matched with the geological data.

In the North, Northwest, and East of the area (Fig.3, 4, 5, 6), the physical attributes of the field show 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.

Observations in Fig.3, Fig.4, Fig.5, and Fig.6 show 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.

The results of dividing the area into clusters having close features based on attributes of the gravity anomaly field reflect the homogeneity of the petrographic composition of the rock layers, which is a premise to increase the reliability of geologists and geophysicists' decisions, providing an intuitive and multi-dimensional view about the process of tectonic activities in the central region of Vietnam.

4. Conclusions

The results of processing and interpreting the data of Bouguer gravity anomalies in Central Vietnam led to some conclusions:

- The unsupervised learning algorithm (K-means) is effectively applied to the gravity anomaly data in Central Vietnam.
- The results of automatic division into 07 clusters, 09 clusters, 11 clusters, and 13 clusters having close characteristics of the properties of the gravity anomaly field completely coincide with the tectonic geology in the area.
- In the southwest, center, and south part of the study area maps show that the physical attributes of the field are quite complex, showing that in this area the rock layers are crumpled and broken in different directions, and this is the potential location to form endogenous mineral deposits of magma origin.
- In the north, northwest, and east part of the study area maps show little change in the physical properties of the field, reflecting that the rock layers in

this area are quite stable with the structural direction extending from the northwest to the southeast.

- The clustering results by means of the K-means unsupervised learning algorithm initially increase the reliability in the decisions of geologists and geophysicists about the interpretation of geological structures and the assessment of the origin of deep-hidden mineral deposits in the area.

Acknowledgments

This work is supported by the grassroots project of the University of Mining and Geology, Hanoi, Vietnam, no. T23-12. The authors would like to thank Geophysical Division, General Department of Geology and Minerals of Vietnam, for providing 1:100.000 scale data of the Bouguer gravity anomaly in the central region of Vietnam. Thanks to Professor, Ph.D. of Science Petrov Aleksey Vladimirovich, University of Geology and Exploration named Sergo Ordzhonikidze, Moscow, Russia (MGRI), for providing commercial software "COSCAD-3D" for data processing.

Literatura – References

1. Aline Melo, Yaoguo Li. Geology differentiation by applying unsupervised machine learning to multiple independent geophysical inversions. *Geophysical Journal International*, 2058-2078 (2021). <https://doi.org/10.1093/gji/ggab316>
2. Blakely, R.J. Potential theory in gravity and magnetic application. Cambridge University Press, 461 pages (1996). <https://www.cambridge.org/core/books/potential-theory-in-gravity-and-magnetic-applications/348880F23008E16E663D6AD14A41D8DE>
3. Carmine Cutaneo, Andrea Vitale, Maurizio Fedi. Unsupervised Boundary Analysis of potential field data: a machine learning method. Society of Exploration "First international Meeting for Applied Geoscience &Energy", 941-945 (2021). DOI:10.1190/segam2021-3594444.1
4. Hai Thanh Tran, Khin Zaw, Jacqueline A. Halpin, Takayuki Manaka, Sebastien Meffre, Chun-Kit Lai, Youjin Lee, Hai Van Le, Sang Dinh. The Tam Ky- Phuoc Son shear zone in central Vietnam: Tectonic and metallogenic implications. *Gondwana Research*, 26, 144-164 (2014). DOI: 10.1016/j.gr.2013.04.008
5. Hong T. Phan, Aleksey V. Petrov, Phuong M. Do, Giau M.lai, Luu T. Nguyen. Geological structure of central Vietnam by interpretation processing of gravitational survey data using the "COSCAD 3D" computer technology. *Geology and Exploration*, 77-90 (2020). <https://doi.org/10.32454/0016-7762-2020-63-5-77-90>
6. Hong P. T. 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, 126 pages (2022). [https://www.mgri.ru/science/scientific-and-innovative-activity/dissertation-council/download/%D0%A4%D0%B0%D0%BD%20%D0%A2.%D0%A5._%D0%90%D0%B2%D1%82%D0%BE%D1%80%D0%B5%D1%84%D0%B5%D1%80%D0%B0%D1%82_final%20\(22.04.2022\).pdf](https://www.mgri.ru/science/scientific-and-innovative-activity/dissertation-council/download/%D0%A4%D0%B0%D0%BD%20%D0%A2.%D0%A5._%D0%90%D0%B2%D1%82%D0%BE%D1%80%D0%B5%D1%84%D0%B5%D1%80%D0%B0%D1%82_final%20(22.04.2022).pdf)
7. <https://topex.ucsd.edu>
8. MacQueen, J. Some methods for classification and analysis of multivariate observations. Proceedings of the fifth Berkeley symposium on mathematical statistics and probability. Oakland, CA, USA (1967).
9. Nikitin A.A., Petrov A.V. Theoretical foundations of geophysical information processing. Study guide, Moscow 113 pages (2008). <https://www.geokniga.org/bookfiles/geokniga-nikitinteorosnovygeofizinfo2008.pdf>
10. Nguyen Truong Luu, Hoang Van Long, Phan Duc Le, Pham Van Hung, Nguyen Thien Cong, Tran Duan. 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 (2020). In Vietnam.
11. Nguyen Truong Luu et al.. 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, (2000). In Vietnam.
12. Nguyen Truong Luu et al. 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 (2014). In Vietnam
13. Nguyen Xuan Son et al. 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 (2000). In Vietnam.
14. Phan T. H., Petrov A.V., Do M. Ph. Processing and interpretation of gravity data anomalies in the central Vietnam using the «COSCAD 3D» computer technology. Young- earth sciences, MGRI-RGGRU, Moscow, 293-297 (2020). [https://www.mgri.ru/science/scientific-practical-conference/2020/TOM%204%20\(1\).pdf](https://www.mgri.ru/science/scientific-practical-conference/2020/TOM%204%20(1).pdf)
15. Petrov A.V. Methods of adaptive filtering and multivariate analysis of variance in the computer system for processing geophysical information "COSCAD 3D". Doctoral thesis, 192 pages (1997).
16. Petrov A.V., Yudin D.B., Soeli Hou. 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 (2010). http://www.kscnet.ru/kraesc/2010/2010_16/art13.pdf
17. Petrov A.V., Zinovkin S.V., Osipenkov D.Yu., Yudin D.B. Computer technology of statistical and spectral-correlation analysis of data "COSCAD 3D". *Geoinformatics*, 4, 7-13 (2011). <https://elibrary.ru/item.asp?id=17109032>
18. Petrov A.V., Yudin D.B. Computer technology of statistical and spectral - correlation analysis of areal geophysical information presented by profile observations "COSCAD-profile" Modern geophysical and geoinformation systems. *Earth sciences*, 2, UDC 551-214, 76-84 (2008)..
19. Petrov A.V., Trusov A.A. Computer technology of statistical and spectral-correlation analysis of three-dimensional geoinformation "COSCAD 3D". *Geophysics*, 4, 29-33 (2000). <https://elibrary.ru/item.asp?id=23719797>
20. Quyen Minh Nguyen, Quinglai Feng, Jian-Wei Zi, Tianyu Zhao, Hai Thanh Tran, Thanh Xuan Ngo, Dung My Tran, Hung Quoc Nguyen. 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 (2019). DOI:10.1016/j.gr.2019.01.002
21. Tran Van Tri, Nguyen Xuan Bao. Map of geological structure at 1:500 000 scale in the central region of Vietnam. Information Center, Geological Archives, Hanoi (2008). In Vietnam.