Identifying Correlation of Coal Seams in the Tien Hai Area, Northern Vietnam by Using Multivariate Statistic Methods

KHUONG The Hung[1, *], NGUYEN Phuong[1], NGUYEN Thi Cuc[1], PHAM Nhu Sang[1], NGUYEN Danh Tuyen[2]

[1] Hanoi University of Mining and Geology, 18 Vien street, Hanoi, Vietnam

[2] Vinacomin - Vietbac Geology Joint Stock Company, 30B Doan Thi Diem, Hanoi, Vietnam

Corresponding author: khuongthehung@humg.edu.vn

Abstract. In northern Vietnam, the Tien Hai area is considered a high potential area of coal deposits. Two hundred fifty-six geochemical coal samples of 13 cores in the Tien Hai area investigate coal seams and coal deposits to identify the correlation of coal seams. According to the statistical method and cluster analysis of geochemical samples, the results indicate that the Mg, V, As, Ca, Zn, Cr, Co, K, Na, Sr, Fe, Ge, Re, U, Mo, Th, and Ga elements are good indicator elements of the major and trace elements in coal. Most of them comply with the normal or lognormal distribution rules. Besides, the Yb, Sc, Ho, Er, Tm, Lu, Y, Tb, Pr, Dy, and Sm elements are also good indicator elements for rare earth elements in the region. Therefore, the selected elements are used to identify the correlation of the coal seams in the Tien Hai area. Based on the similarity degree between studied objects, the results of grouping boreholes in coal seams show that the correlation of coal seam TV2-11 is suitable and acceptable, the coal seams TV3-6a, TV3-6b, and TV3-6c can be grouped into the coal seam TV3-6. These results present that the models can help study geochemical coal samples and identify the correlation of the coal seams in the Tien Hai area. Additionally, the statistical analysis shows a remarkable degree to determine the correlation of the coal seams. Geochemical coal data can help to evaluate the indicator elements of the major, trace elements, and rare earth elements in coal seams and coal rashing of adjoining and pillar rocks in the Tien Hai area, northern Vietnam.

Keywords: Identifying correlation of coal seams, Multivariate statistic methods, Tien Hai area, Northern Vietnam

1. Introduction

Generally, the correlation of coal seams has been one of the most complicated issues that need further research to clarify. Up to the present, investigations on coal seam correlations have mainly focused on stratigraphic correlation and coal seam division, in which properties and composition of major, rare earth, and trace elements play a vital role. These important findings were obtained by geochemical data processing methods based on Dickinson & Suczek (1979) [1], which mainly rely on the mineral composition of sandstones to determine the provenance of sedimentary materials tectonic setting of the basin. Studies on sedimentary basins have proved that the mineral composition of modern sandstones of different tectonic settings exhibits systematic variations in functional correlation with the origin type and tectonic setting [2, 3, 4, 5]. Many similar studies on ancient sedimentary basins have confirmed the relationship between the mineral composition of sandstones, provenance, and the tectonic setting of sedimentary basins [1, 6, 7, 8]. As a result, the information about the chemical composition of major oxides, mainly rare earth and trace elements obtained by qualitative and quantitative data analysis methods, can identify the correlation of the coal seams.

In Vietnam, the Red River basin (RRB) has been recorded as a region with great potential for coal deposits [9, 10]. In the 2000s, the oil and gas prospecting in this area has gained many positive results about the coal resource potential in the Neogene sediments [11, 12]. However, the geological structure conditions are complicated, and coal seams are distributed deep underground, and most geological works through abundant only focus on oil and gas assessment, prospecting, and exploration. These facts have led to limited research on coals and associated minerals. Therefore, it is very difficult and, in some cases, impossible to identify the correlation of coal seams in the Tien Hai area by applying traditional methods. Accordingly, it is necessary to select appropriate methods and processes to identify the correlation of coal seams in the Red River coal basin in general and the Tien Hai area in particular by using geological and geophysical parameters based on multivariate statistic methods and specialized software on computers to contribute to solving actual requirements. This study aims to make a clear context for identifying the correlation of the coal seams in the Tien Hai area, northern Vietnam. This study will help us better understand the contribution of the major and trace elements and rare earth elements in this area and provide valuable tools for correlating

the coal seams in the Red River coal basin.

2. Geological framework

2.1 Structural and tectonic setting of the study area

The Red River coal basin is one part of the RRB in the Vietnam mainland (Fig. 1A). The RRB is about 650 km long, 150 km wide, and has a sediment thickness of more than 10 km at its depocenter. The Tien Hai area is located in the Red River coal basin (Fig. 1B).

The Red River fault zone, which runs northwest to southeast, is bounded by two major northwest to southeast trending faults that form a valley shape in the RRB. Chay River and Lo River faults are the two major faults that dominated the Tertiary sedimentation and deformation of the RRB (Fig. 1A).
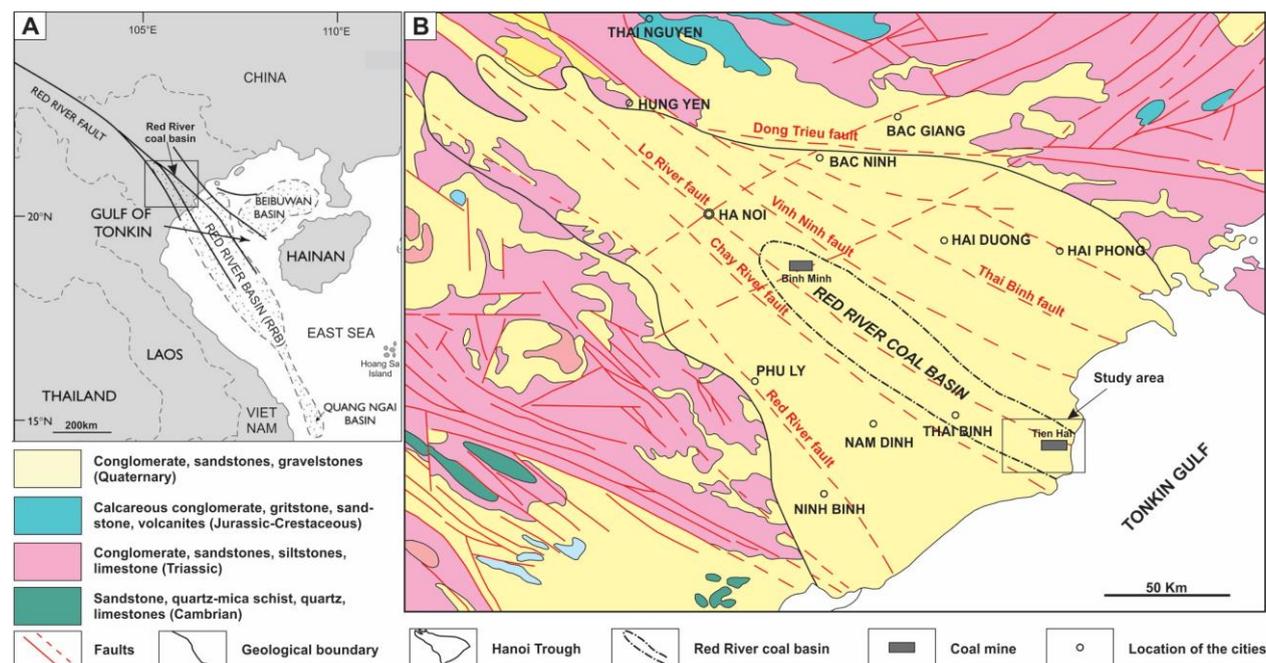
Fig. 1. A-Tectonic sketch map of the Red River basin, showing the location of the Red River coal basin [13], B-Geological map of the Red River coal basin, and location of the Tien Hai area [9, 14].

The RRB was formed and developed as a result of the Indochina block's large-scale extrusion and clockwise rotation induced by the collision of Indian and Eurasian plates. During the early Eocene, the plate began to rift slowly, resulting in forming the NE–SW trending proto-East Sea of Vietnam [11, 13]. Tectonic escape of the Indochina block to the southeast has been exacerbated by the continued collision of the Indian plate with the Eurasian plate during the late Eocene to early Oligocene (40 Ma). The Indochina block's sinistral-slip created a left lateral shear-rifting point followed by rapid clockwise rotation (15-20°). As a result, the forming of the RRB can be attributed to left-lateral displacement of around 200 to 800 km along with the Red River fault system. Strike-slip motions were mostly observed along the Lo River and Chay River faults. Grabens and half-grabens formed during the rifting process in the RRB, which was later filled in by fluvial and lacustrine deposits. When subsidence had to be maintained, the amount of water rose, and the lacustrine setting opened up into a marginal marine or shallow marine ecosystem to create more sedimentation space. The expansion out of the East Sea to the southeast in the late Oligocene, however, induced compression, resulting in the inversion and uplifting of hosts and grabens in the RRB. During the late-early Oligocene, active erosions cut large amounts of earlier deposited Oligocene sections, causing in a pop-up formation on overlying younger sediments. The late Oligocene angular unconformity that divided the syn-rift and post-rift portions were thought to be a significant breakup unconformity caused by erosion.

This basin was further depressed by continuous left-lateral transtension along with the Red River fault systems and thermal cooling after the rifting period. During the early Miocene, widespread subsidence resulted in transgression and back-stepping sedimentation. In the middle Miocene, though, the Sundaland

plate stopped the Indochina from drifting southeast [16, 17]. As a result, relative motions around Red River faults have shifted from left to right lateral. The strike-slip activity, which resulted in fault reversal and the forming of major inversion structures, led to the petroleum system's trapping mechanism in the RRB.

2.2 Stratigraphy

Four major Cenozoic formations have been found and divided in the Tien Hai area [18]. Accordingly, Cenozoic sediments originated from Paleogene with the Eocene-Oligocene Dinh Cao formation, covered by the Miocene formations including the Phong Chau, Phu Cu, and Tien Hung formations, and Pliocene Vinh Bao, Hoa My formations. Overlapping them are Pleistocene sediments of the Hai Duong formation and Holocene sediments of the Kien Xuong formation (Fig. 2). The survey results of the drilling core sample in the Tien Hai area presented sediments of 4 stratigraphic units as follows.
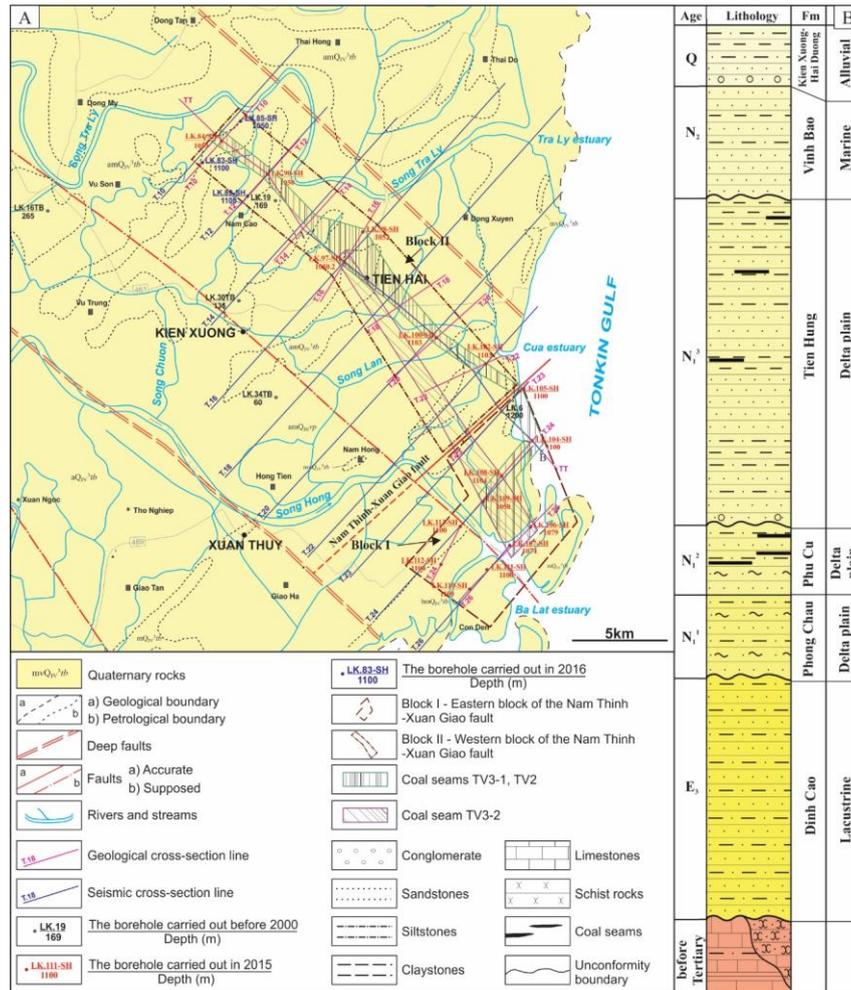
Fig. 2. A - Simplified geological map of the Tien Hai area [9]; B - Generalized stratigraphic column for the northeastern Red River coal basin [14].

The Tien Hung formation was found in all drilled holes that have been under construction in the Tien Hai area; however, only layer 3 of the Tien Hung formation was recorded from a depth of –319 m (LK.98-SH) to –443.3 m (LK.106-SH), at the average depth of –372 m. According to the borehole logs, the lithological composition of the third layer belongs to a coal-forming marsh, composed of large to fine-grained sandstone, siltstone, claystone, and coal. This layer contains 8-15 coal seams with thickness ranging from 0.8 m to 5.1 m, mainly from 2 m to 3m.

The Vinh Bao formation was first described and established in the borehole LK.3-SH [19]. This formation was recorded at the depth from –240 m to –510 m in the Vinh Bao district. The formation was found in all five boreholes in the Tien Hai area [14]. The core samples showed that the sediments of the Vinh Bao formation are characteristic of the lagoon facies, but in very few locations, there can be sand and

gravel in the alluvial-fluvial facies like in the borehole LK.105-SH. The composition of the sediments consists of siltstone interleaved with fine-grained sandstone or a few thin layers of light gray, ash-gray, green-felt claystone. The Vinh Bao formation covers unconformity on the sediments of the Tien Hung formation, and it is also covered unconformity by Quaternary sediments.

The Hai Duong formation was found in all five boreholes, at the depth from –78.4 m to –189.6 m (LK.102-SH), from –79.2 m to –238.6 m (LK.104-SH), from –137.2 m to –195.5 m (LK.106-SH). Its lithological composition includes pebbles, gravel, grit, sand, and a little sandy clay; however, the large-grained rock layer accounts for a higher proportion in the middle part of the formation. The sedimentary composition of the Hai Duong formation is characteristic of the continental alluvial facies. The average thickness of the Hai Duong formation is about 90 m, ranging from 58.3 m (LK.106-SH) to 159.4 m (LK.104-SH). Hai Duong formation covers unconformity on the sediments of the Vinh Bao formation.

The Kien Xuong formation consists mainly of clay sand, sandy clay, clay, powder, and large grain size. Clay sand has dirty gray, ash gray, reddish-brown gray colors, with many plant ruins. The mineral sand is composed of quartz, feldspar, and muscovite. Clay usually has light coffee and a relatively fine pink-brown color, sometimes containing small scales of muscovite. The average thickness of the formation is from -78.4 m (LK.102-SH) to –137.2 m (LK.106-SH). The formation covers conformity on the Hai Duong formation.

Structurally, the Tien Hai area belongs to a small part of the Khoai Chau-Tien Hai zone, limited by two main faults, namely the Vinh Ninh fault in the northeast and the Thai Binh fault in the southwest. The thickness of Cenozoic sediments in this zone is very large, and it can reach over 7,000 m, including Eocene to Quaternary formations (Figs. 2-3).
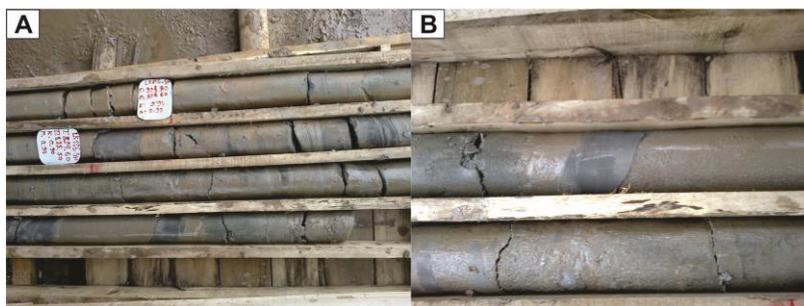
Fig. 3. The core samples of the borehole LK.106-SH in the Con Vanh, Tien Hai area, showing sandstone, siltstone, and gravel in the alluvial-fluvial facies at the depth from –823 m to –827 m [20].

2.3 Properties of coal in the Tien Hai area

The Tien Hai area mainly has brown coal, located in the general structure of the Khoai Chau-Kien Xuong-Tien Hai coal strip [18, 21]. According to the deep borehole and geophysical data, the brown coal distribution in the area is 60-100 km long, with an average length of 70 km, and 7-10 km wide with an average width of 8 km [14].

The borehole drilling successfully identified from 1 to 45 coal seams and coal lenses whose thickness ranges from 0.2 m to 5.0 m, with an average of 1.0 m. Coal seams exist in a large area where the distance between the coal seams in the cross-section is from 30 m to 50 m, even from 10 m to 100 m in some places. Coal seams widen gradually to the southeast, with the dip angle ranging from 8° to 15° and 10° on average. The dip angle of seams can increase from 25° to 40° when it closes to the fault plane. By the end of 2015, 5 coal seams which were concentrated mainly in the Tien Hung formation, were identified. In order from bottom to top of the stratigraphy are coal seam 1 (TV1), coal seam 2 (TV2), coal seam 3 (TV3), coal seam 4 (TV4), and coal seam 5 (TV5) that are distributed at the depth from –120 m to –1,750 m. The properties of coal seams are presented in Table 1.

Tab. 1. The properties of coal seams (from TV2 to TV5) [14].

No.	Borehole name	Coal seam name (amount)	Thickness (m)			Sounding rocks of coal seams	
			From	To	Average	Adjoining rocks	Pillar rocks
1	LK.84-SH	TV3 (7)	0.9	3.3	2.1	large grained sandstone	claystone
		TV2 (10)	0.6	10.0	5.3	coal clay	claystone
2	LK.90-SH	TV3 (8)	0.65	3.1	1.875	medium-grained sandstone	claystone
		TV2 (13)	0.65	9.9	5.275	medium-grained sandstone	claystone
3	LK.97-SH	TV4 (1)			0.9	claystone	claystone
		TV3 (12)	0.65	6.4	3.525	claystone	claystone
		TV2 (2)	0.6	8.6	4.6	gritstone	claystone
4	LK.98-SH	TV3 (7)	2.0	6.8	4.4	claystone	siltstone
		TV2 (3)	0.6	8.7	4.65	claystone	claystone
5	LK.100-SH	TV3 (6)	1.6	7.25	4.425	fine-grained sandstone	claystone
		TV2 (2)	3.8	5.8	4.8	claystone	claystone
6	LK.108-SH	TV5 (3)	0.45	4.6	2.525	medium-grained sandstone	siltstone
		TV4 (10)	1.0	7.7	4.35	medium-grained sandstone	claystone
		TV3 (1)			2.9	medium-grained sandstone	claystone
7	LK.109-SH	TV5 (2)	0.9	6.0	3.45	medium grained sandstone	claystone
		TV4 (9)	0.4	7.6	4.0	gritstone	claystone
		TV3 (4)	0.4	4.0	2.2	medium-grained sand	siltstone
8	LK.106-SH	TV5 (2)	0.75	6.5	3.625	medium-grained sandstone	claystone
		TV4 (10)	0.7	6.4	3.55	large grained sandstone	claystone
		TV3 (3)	0.75	8.7	4.725	claystone	claystone
9	LK.107-SH	TV5 (2)	0.5	7.5	4.0	medium grained sandstone	siltstone
		TV4 (5)	0.9	5.6	3.25	fine-grained sand-stone	siltstone
		TV3 (3)	0.4	4.6	2.5	medium grained	siltstone

No.	Borehole name	Coal seam name (amount)	Thickness (m)			Sounding rocks of coal seams	
			From	To	Average	Adjoining rocks	Pillar rocks
						sandstone	
10	LK.110-SH	TV5 (2)	1.1	4.1	2.6	claystone	siltstone
		TV4 (3)	1.2	4.1	2.65	fine-grained sandstone	siltstone
		TV3 (5)	0.85	2.3	1.575	siltstone	siltstone
11	LK.102-SH	TV3 (8)	1.45	5.8	3.625	claystone	claystone
		TV2 (6)	1.3	5.1	3.2	claystone	claystone
12	LK.104-SH	TV4 (6)	2.6	5.05	3.825	large grained sandstone	claystone
		TV3 (3)	1.9	3.9	2.9	large grained sandstone	claystone
13	LK.105-SH	TV4 (6)	0.9	6.0	3.45	gritstone	claystone
		TV3 (2)	2.3	2.6	2.45	fine-grained sandstone	claystone

The coal seam 1 (TV1) was identified in 4 boreholes which are distributed at the depth from –1,200 m to –1,900 m. The seam TV1 is maintained according to the strike direction, the dip format is from relatively stable to unstable, the average thickness of the TV1 is from 1.90 m to 3.27 m, and the average dip angle is 10°. The adjoining rocks of the coal seam are mainly composed of claystone, siltstone, and sometimes sandstone, while its pillar rocks consist mainly of claystone.

The coal seam 2 (TV2) was recorded in 11 boreholes and included 15 coal lenses distributed at the depth from –500 m to –1,400 m. According to the strike direction, the coal seam is maintained, and the dip format is from relatively stable to unstable. The average thickness of the seam TV2 is from 1.28 m to 6.46 m, and the average dip angle is 12°. The adjoining and pillar rocks of the coal seams are mainly composed of claystone, siltstone, sometimes sandstone, and gritstone.

The coal seam 3 (TV3) was found in 19 boreholes. The seam TV3 includes 22 coal lenses distributed at the depth from –295 m to –1,090 m. The coal seam is maintained in the strike direction, the dip format is relatively stable to unstable, the average thickness is from 0.39 m to 4.32 m, and the average dip angle is 13.22°. The coal seam's adjoining and pillar rocks consist mainly of claystone, siltstone, and sometimes sandstone.

The coal seam 4 (TV4) was found in 10 boreholes and consists of 10 coal lenses distributed at a depth of –376m to –1,015m. The seam is maintained in the strike direction, the dip format is relatively stable, the average thickness is from 1.34m to 5.27m, and the average dip angle is from 16° to 17°. The adjoining and pillar rocks of the coal seam are mainly composed of claystone, siltstone, sometimes sandstone, and gritstone.

The coal seam 5 (TV5) was determined in 5 boreholes constructed by the Intergeo Division [12]. The seam TV5 includes 3 coal lenses, distributed at the depth from –432 m to –701 m. The coal seam is maintained in the strike direction, the dip direction is not stable, the average thickness is from 0.39 m to 5.60 m, and the average dip angle is 12.73°. The adjoining and pillar rocks of the coal seam are mainly composed of siltstone and claystone.

The synthesis results of documents in the Tien Hai area show that the drilling works may only be limited at the upper coal beds of the third layer, belonging to the Tien Hung formation, which should be further clarified through research on geological cross-sections according to drilling and geophysical data, especially 2D reflective seismic measurements.

3.   Materials and Methods

The geochemical properties of the ancient and modern sandstone formations will clarify the identification marks of the bedrock and the tectonic setting in their composition, helping to speculate the redistribution of chemical elements during and after the sediment deposition process. In the scope of this study, the combination of correlation analysis and geological-geophysical multidimensional statistical methods is applied to identify the correlation of the coal seams in the Tien Hai area.

Based on the results of identifying correlation of coal seams in the Quang Ninh and Thai Nguyen-An Chau coal basins [22, 23] together with analysis results of characteristics of the geological structure, coal seams, core data, and actual geophysical measurements, the correlation models of coal seams in the study area were established in order of its geological and geophysical parameters. The proposed process is shown in Figure 4.
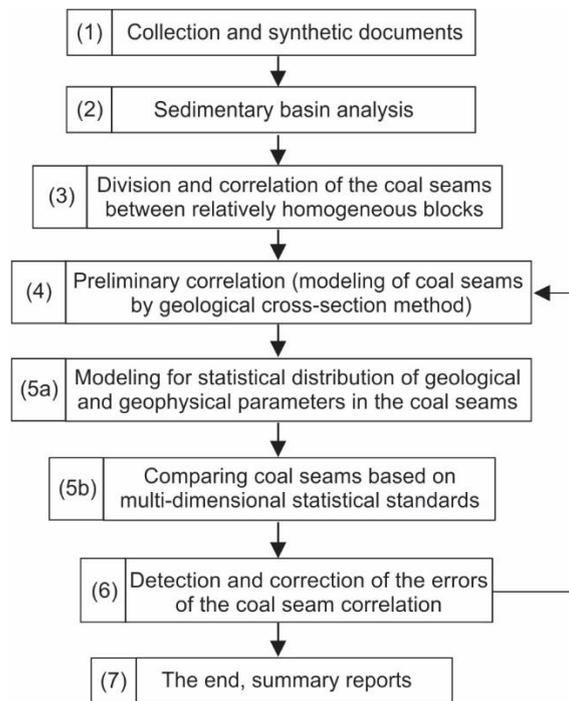
Fig. 4. Seven steps for correlation of coal seams at the Tien Hai area by using geological and geophysical parameters.

3.1 Multivariate statistics

One of the most commonly used algorithms to solve the coal seam correlation problem is the cluster analysis method, also known as dendrogram analysis, which includes the following steps.

Step 1: Determine the similarity $E(x_i, x_j)$ between each pair of objects (coal seams) according to the formula.

$$E(x_i, x_j) = Cov(x_i, x_j) = \frac{\sum_{p=1}^{k} a_{ip} \cdot a_{jp}}{\sqrt{\sum_{p=1}^{k} a_{ip}^2 \sum_{p=1}^{n} a_{jp}^2}} \tag{1}$$

in which i, j = 1,2…, n; k is the number of study properties, n is the number of study objects, and a_{ip}, a_{jp} are coordinates of the vectors x_i, x_j. The result will create a $Q_{m.m}$ matrix with terms {E_{ij}}, i, j = 1…, n.

Step 2: Find the arithmetic mean H from values reflecting the similarity E_{ij} according to the formula.

$$H = \frac{\sum_{i,j=1}^{m} E_{ij}}{n^2} \tag{2}$$

Step 3: Select $E_t^* = \frac{max}{t} E_{it}$; t=1,2,...,n; $t^1{}_i$; where i is the number of rows, t is the number of columns of the matrix $Q_{m.m}$.

Step 4: After selecting E_i^* values, compare E_i^* values with H. When $E_i^* > H$, the study subjects are classified into the same group. When $E_i^* < H$, the study subjects cannot be classified in the same group.

Step 5: Select $t = t^*$, in which $E_{it} = E_i^*$ (the name of column with value E_i^*).

Step 6: Formulate r sets with the following forms.

$$B_s = (t_s^*, i_1{}^s, ..., i_\mu{}^s..., i_{cs}{}^s; s = 1, ..., r) \tag{3}$$

where *r* is the number of different values of t_i^*; $i_1{}^s, ..., i_{cs}{}^s$ is the symbol of row with the maximum value.

Step 7: Start to synthesize B_1 from the second term $i^1{}_1$ compared to the first term of $t_2^*..., t_r^*$ of combinations $B_2..., B_r$. Concerning B_v, when $2 \le v \le r$, $i^1{}_1 = t_v$, $B_1{}^1 = B_i$, t_v^* has such a formula as.

$$B_1' = (t_1^*, i_1^1, ..., i_{l_1}^1, i_i^y, ..., i_{1v}^y) \tag{4}$$

That is, delete the combination B_v and change to the term $i^1{}_2$. If the first term of the combinations $B_2...$ B_r has no terms like $i^1{}_1$ continue to compare with the term $i^l{}_2$. Continue like this until all the terms $i^1{}_1$ of combination B_1 are considered.

Step 8: The number r', the number of remaining combinations after step 7 is performed, is reserved for the combination B'$_1$ (If new terms are not included in step 7).

Numbers from 1 to r'-1 are reserved for the remaining combinations and combinations with a new number s=1 after step 7. Steps 7 and 8 have been repeated until the combinatorial r* of combination B_s^* with $t_s^* \ne i_{\mu,s}{}^k$, k=1,...,n; μ=1,...,k.

Step 9: Compare pairs of combinations B_u^*, B_v^*; u, v = 1..., r*, starting from the second term, $i^u{}_1$ ($i^v{}_1$). Combine all combinations of the same terms.

Step 10: Calculate the properties of subsets A_s, s=1,...,r*; $\hat{X}_s = (\hat{\alpha}_{s_1}, \hat{\alpha}_{s_2}, ..., \hat{\alpha}_{s_n})$.

$$\hat{\alpha}_{sp} = \frac{\sum_{j,s=1}^{m_s} \alpha_{jsp}}{m_s} \tag{5}$$

with α_{jsp} is the Pth coordinate of the X_{js} vector corresponding to the a_{js} condition of the subset A_s; m_s is the number of objects in a subset.

Step 11: Calculate the similarity values between the subset pairs A_s, A_v, and s, v = 1 ÷ r*. Then, establish the matrix C with the terms D_{sv} by the formula.

$$D_{sv} = \frac{1}{m_s m_v} \sum_{i_s=1}^{m_s} \sum_{j_v=1}^{m_v} E_{i_s j_v} \tag{6}$$

in which, i_s and j_v are the number of the sum of subsets A_s, A_v; $E_{is,jv}$ is the similarity value between points a_i and a_j calculated in step 1; m_s and m_v are the number of objects in the combinations A_s and A_v, respectively.

3.2 Using multivariate statistics for mapping and comparison of coal seams

Assuming that 2 coal seam beds are belonging to 2 relatively homogeneous blocks (or 2 geological cross-sections according to the exploration route whose properties need to be compared by 2 document term matrices (original data table):

$$T_1 = \{\xi_1', \xi_2', ..., \xi_l', ..., \xi_m'\} \text{ and } T_1 = \{\xi_1'', \xi_2'', ..., \xi_l'', ..., \xi_m''\}$$

Each seam bed was controlled by the boreholes (corresponding points of view) $X_1, X_2, ..., X_{n1}$ (belonging block I) and $Y_1, Y_2, ..., Y_t, ..., Y_{n2}$ (belonging to block II). In which:

$$X_t = \left\{ X_{t_1}, X_{t_2}, ..., X_{t_j}, ..., X_{t_m} \right\}$$

$$Y_t = \left\{ Y_{t_1}, Y_{t_2}, ..., Y_{t_j}, ..., Y_{t_m} \right\}$$

In order to compare T_1 and T_2, it is necessary to test the hypothesis $H_0: \mu_1 = \mu_0$ or $H_1: \mu_1 \neq \mu_2$. The hypothesis H_0 can be tested according to multidimensional statistical standard (V), determined by the formula.

$$V = -(n_1 + n_2 - 2 - \frac{m}{2}) \ln \frac{|S_1|}{|S_o|} \tag{7}$$

where $|S_1|$, $|S_0|$ are determinants of the matrices S_1 and S_0 determined by the formula (1) and (2).

The hypothesis H_0 is accepted (that is, assuming that the coal seams in two blocks that need to be compared are similar, in other words, they are correlated) when: $V = \chi^2_{q,m}$ (in which $\chi^2_{q,m}$ is the Chi-square standard, q is significance level and m is the number of study properties or parameters).

Accordingly, the terms of the matrices S^0_{ij} and S^1_{ij} are calculated by the following formula.

$$S^o_{ij} = \frac{1}{n_1 + n_2 - 2} \left[\sum_{t=1}^{n_1} X_{t_i} X_{t_j} + \sum_{t=1}^{n_2} Y_{t_i} Y_{t_j} - \frac{1}{n_1 + n_2} \left(\sum_{t=1}^{n_1} X_{t_i} + \sum_{t=1}^{n_2} Y_{t_i} \right) \left(\sum_{t=1}^{n_1} X_{t_j} + \sum_{t=1}^{n_2} Y_{t_j} \right) \right] \tag{8}$$

$$S^1_{ij} = \frac{1}{n_1 + n_2 - 2} \left[\sum_{t=1}^{n_1} X_{t_i} X_{t_j} + \sum_{t=1}^{n_2} Y_{t_i} Y_{t_j} - \frac{1}{n_1} \sum_{t=1}^{n_1} X_{t_i} \cdot \sum_{t=1}^{n_1} X_{t_j} - \frac{1}{n_2} \sum_{t=1}^{n_2} Y_{t_i} \cdot \sum_{t=1}^{n_2} Y_{t_j} \right] \tag{9}$$

After determination of the multidimensional statistical standard (V), compare the V standard with the Chi-square standard ($\chi^2_{q,m}$).

3.3 Testing results of the coal seam correlation

The techniques described above use a mixture of geological and geophysical criteria to compare coal seams in two areas (or cross-sections); however, geologists often have to correlate several areas (cross-sections). They often use the multidimensional statistical standard proposed by Rao (1967) [24] to solve this task.

In order to correlate k geological forms (coal seams or coal seam beds); in which each coal seam or coal seam bed is characterized by a combination of geological and geophysical parameters and presented in the document term matrix in the form of $T_C = T_1, ..., T_L, ..., T_K$, with corresponding mathematical expectations $\mu_1, \mu_2,..., \mu_k$ and corresponding covariance matrices satisfying the condition: $\sum_1 = \sum_2 = ... = \sum_k = \sum_0$. Then, the results collected from the boreholes can be expressed as $X_u = \{X_{ct1}, X_{ct2}, ..., X_{ctj}\}$.

To confirm that the coal seams correlated on the cross-sections and correlation between blocks are suitable (or similar), it is often checked under the assumption: $H_0 = \mu_1 = \mu_2 = ... = \mu_k = \mu_0$. To test hypothesis H_0, the multidimensional statistical standard (V) is usually used and determined by the formula.

$$V = -\left(\sum_{i=1}^{k} n_1 - 1 - \frac{m+k}{2} \right) \ln \frac{|S_1|}{|S_o|} \tag{10}$$

The formula calculates the terms of the covariance matrices S_1 and S_0.

$$S^o_{ij} = \frac{1}{n-1} \left[\sum_{L=1}^{k} \sum_{t=1}^{n_t} X_{Lt_i} X_{Lt_i} - \frac{1}{n} \sum_{L=1}^{k} \sum_{t=1}^{n_t} X_{Lt_i} \cdot \sum_{L=1}^{k} \sum_{t=1}^{n_t} X_{Lt_i} \right] \tag{11}$$

$$S^1_{ij} = \frac{1}{n-1} \left[\sum_{L=1}^{k} \sum_{t=1}^{n_t} X_{Lt_i} X_{Lt_i} - \sum_{L=1}^{k} \frac{1}{n} \left(\sum_{t=1}^{n_L} X_{Lt_i} \cdot \sum_{t=1}^{n_L} X_{Lt_i} \right) \right]; N = \sum_{L=1}^{k} n_i \tag{12}$$

If $V \leq \chi^2_{q,m}$, the assumption is accepted.

4.   Results and discussions

Identifying the correlation of coal seams with a combination of geological and geophysical parameters is carried out following 7 steps (Fig. 4). The first step is to set up a database for each selected section, including the number of reflections on stratigraphic, morphological, and structural characteristics of coal seams, their material composition, etc. The second step carries out sedimentary basin analysis. The third step divides and correlates the coal seams between relatively homogeneous blocks. The fourth step models

the coal seams by using geological cross-sections. The fifth step estimates the statistical distribution of the coal seams' geological and geophysical parameters and then compares coal seams based on the multidimensional statistical standards. The sixth step includes detecting and correcting the errors of the coal seams correlation, and the last one summarizes results and reports.

4.1 The characteristics of statistical distribution elements

The rules of the statistical distribution of coal samples and related elements can be recognized by determining their statistical distribution models. Technical features and the element contents of coal in the Tien Hai area expressed through statistical characteristics reflecting coal quality are summarized in Table 2.

The analysis results showed the technical features of W^{pt} (moisture), A^d (ash), V^{ch} (volatile), Q^{ch} (quantity of heat), C^d (carbon), H^d (hydro), O^d (oxy), N^d (nitro), S^d (sulfur), and P (phosphor) in the coal seams with the average values of 16.72, 5.54, 46.70, 7125.61, 69.15, 4.29, 19.34, 0.89, 0.79, and 0.010, respectively (Tab. 2). They also present the coal seams belonging to the low sulfur coal group that is difficult to ignite. Based on the C^d, N^d, and O^d contents, the coal seams are arranged in the low-grade metamorphic ones. Most of the technical parameters of coal seams comply with the normal standard distribution rules; only the ash (A^d) and sulfur (S^d) contents comply with the standard lognormal distribution. The variation of the technical parameters ranges from very even to even, which indicates that coal quality is relatively homogeneous. Remarkably, the ash (A^d) and sulfur (S^d) contents are distributed fairly unevenly.

Similarly, the statistical characteristics of major and trace elements in coal show a clear change among element groups. Accordingly, the group of such elements as Na, V, Ge, As, and Ba has little change in content and complies with the normal standard distribution. On the contrary, the group of elements including Al, Ti, Ni, Zn, Mo, and Th has very strong variations on their contents. It complies with lognormal or gamma standard distribution rules. Moreover, the remaining groups of elements vary from quite even to uneven (Tab. 2).

Otherwise, the contents of the major and trace elements in coal rashing have a very clear change among element groups (Tab. 3). Accordingly, the group of elements including B, Ti, V, Cr, Ni, Zn, Ge, Sr, Tl, Th, and TR_2O_3 has displayed little change in their contents and complies with normal standard distribution models. The group of such elements as Ga, Ba, and Pb has extreme variations and complies with gamma distribution rules. The remaining groups of elements range from quite even to uneven. With this, the statistical characteristics of the contents of the rare earth elements do not affect human health in the coal mining process.

Moreover, most of the contents of the rare earth elements comply with the gamma distribution function, except for Sm, Eu, and Gd elements which comply with normal standard distribution rules. Therein, the contents of rare earth elements vary from uneven to very uneven (Tab. 3). These results indicate that the distribution of the contents of rare earth elements in coal rashing is not high. However, it can still create concentrated accumulations, which should be considered in the future coal mining process.

Tab. 2. Statistical characteristics and distribution models of the technical features and elements in coal seams.

Analyzed samples	Element	Mean (%)	Variance	Coefficient of variation (%)	Skewness	Kurtosis	Distribution model
Technical features	W^{pt}	16.72	3.51	21.01	0.19	–0.84	Normal standard
	A^d	5.54	3.14	56.70	0.94	0.87	Lognormal standard
	V^{ch}	46.70	2.06	4.42	0.35	–0.83	Normal standard
	Q^{ch}	7125.61	273.01	3.83	–0.54	–0.79	Normal standard
	C^d	69.15	3.60	5.20	0.05	–0.84	Normal standard
	H^d	4.29	0.27	6.27	0.23	–0.49	Normal standard
	O^d	19.34	2.26	11.67	–0.10	–0.58	Normal standard
	N^d	0.89	0.14	16.17	–0.58	0.40	Normal standard
	S^d	0.79	0.49	62.20	0.80	–0.19	Lognormal standard

Analyzed samples	Element	Mean (%)	Variance	Coefficient of variation (%)	Skewness	Kurtosis	Distribution model
	P	0.010	0.002	22.22	–0.199	–1.017	Normal standard
Major and trace elements	B	1.00	-	-	-	-	-
	Na	4090.42	1105.38	27.02	0.49	–0.97	Normal standard
	Mg	2007.30	1005.12	50.07	0.07	–0.43	Lognormal standard
	Al	2147.99	2674.92	124.53	2.55	6.78	Gamma
	K	1226.71	1043.03	85.03	1.58	2.23	Lognormal standard
	Ca	5370.13	2978.84	55.47	0.80	–0.04	Lognormal standard
	Ti	227.53	242.92	106.77	1.28	0.20	Gamma
	V	128.38	37.89	29.51	0.04	0.12	Normal standard
	Cr	16.69	9.44	56.60	1.76	3.17	Lognormal standard
	Mn	43.37	37.28	85.95	1.64	1.80	Lognormal standard
	Fe	3610.42	2769.99	2796.99	1.88	3.48	Lognormal standard
	Co	8.45	7.84	7.84	2.01	5.54	Lognormal standard
	Ni	52.75	71.90	71.90	3.73	15.70	Gamma
	Zn	171.55	213.64	213.64	3.65	13.70	Gamma
	Ga	12.96	9.45	9.45	1.85	3.32	Lognormal standard
	Ge	11.15	2.37	2.37	–3.14	12.92	Normal standard
	As	147.27	16.27	16.27	–0.94	1.22	Normal standard
	Sr	135.35	76.31	76.31	0.75	–0.09	Lognormal standard
	Mo	1.32	3.46	3.46	4.39	20.24	Gamma
	Pd	-	-	-	-	-	Normal standard
	Ba	130.99	64.12	64.12	1.14	1.03	Normal standard
	Re	161.64	129.70	129.70	0.53	–1.01	Lognormal standard
	Pt	2.84	2.52	2.52	2.23	5.81	Lognormal standard
	Hg	-	-	-	-	-	-
	Tl	-	-	-	-	-	-
	Pb	15.01	10.73	10.73	1.96	2.69	Lognormal standard
	Th	2.15	5.25	5.25	4.37	20.05	Gamma
	U	2.14	1.15	1.15	0.77	–0.69	Lognormal standard
	REE	0.40	0.34	0.34	1.85	2.56	Lognormal standard
Rare earth elements	La	397.99	517.08	129.92	1.90	2.59	Gamma
	Ce	813.08	1037.97	127.66	1.84	2.29	Gamma
	Pr	93.31	123.08	131.90	1.85	2.30	Gamma
	Nd	342.85	451.89	131.80	1.84	2.26	Gamma

Analyzed samples	Element	Mean (%)	Variance	Coefficient of variation (%)	Skewness	Kurtosis	Distribution model
	Sm	926.16	317.28	34.26	0.73	0.99	Normal standard
	Eu	149.27	50.29	33.69	0.14	−1.14	Normal standard
	Gd	281.72	115.17	40.88	0.96	0.81	Normal standard
	Tb	11.31	13.75	121.49	2.53	6.99	Gamma
	Dy	50.19	80.29	159.99	2.81	8.77	Gamma
	Ho	10.10	17.38	172.10	2.99	9.89	Gamma
	Er	26.30	47.88	182.06	3.04	10.15	Gamma
	Tm	3.57	6.42	179.83	3.21	11.46	Gamma
	Sc	13.71	31.96	233.11	4.09	18.09	Gamma
	Yb	8.96	25.70	286.72	4.19	18.61	Gamma
	Y	253.36	453.68	179.07	2.91	9.17	Gamma
	Lu	3.49	6.40	183.10	3.25	11.77	Gamma

Tab. 3. Statistical characteristics and distribution models of the elements in coal rashing.

Analyzed samples	Element	Mean (%)	Variance	Coefficient of variation (%)	Skewness	Kurtosis	Distribution model
	B	4.27	0.59	13.80	−0.08	−1.06	Normal standard
	Ti	3900.42	1343.94	34.46	−0.64	−0.63	Normal standard
	V	116.85	47.07	40.28	−0.74	−0.30	Normal standard
	Cr	109.68	32.45	29.58	−0.57	0.00	Normal standard
	Co	22.09	11.19	50.64	2.12	6.63	Lognormal standard
	Ni	68.75	23.68	34.44	1.12	3.59	Normal standard
	Zn	148.60	54.70	36.81	0.03	−0.53	Normal standard
	Ga	13.85	15.48	111.74	2.18	6.16	Gamma
	Ge	3.62	1.33	36.66	−0.34	−1.19	Normal standard
	Sr	135.60	56.13	41.39	3.01	12.04	Normal standard
Major and trace elements	Pd	-	-	-	-	-	-
	Ba	807.30	1063.86	131.78	3.14	11.49	Gamma
	Re	92.01	89.50	97.27	0.86	−0.76	Lognormal standard
	Os	-	-	-	-	-	-
	Ir	-	-	-	-	-	-
	Pt	-	-	-	-	-	-
	Pb	14.34	16.80	117.17	0.75	−0.81	Gamma
	Tl	0.78	0.36	46.06	0.78	0.75	Normal standard
	Th	17.88	5.38	30.12	−0.90	0.31	Normal standard
	U	5.57	4.09	73.42	3.21	14.68	Lognormal standard
	TR_2O_3	0.51	0.13	26.09	−0.93	0.39	Normal standard
Rare earth elements	La	495.02	257.10	51.94	0.02	−0.83	Lognormal standard
	Ce	493.35	285.36	57.84	0.15	−1.32	Lognormal standard
	Pr	493.38	331.77	67.24	0.26	−1.54	Lognormal standard

Analyzed samples	Element	Mean (%)	Variance	Coefficient of variation (%)	Skewness	Kurtosis	Distribution model
	Nd	379.53	225.42	59.39	0.54	−0.37	Lognormal standard
	Sm	427.92	269.01	62.87	0.56	−0.45	Lognormal standard
	Eu	259.23	129.59	49.99	1.62	2.06	Normal standard
	Gd	492.44	282.30	57.33	0.08	−1.29	Lognormal standard
	Tb	76.01	25.87	34.04	0.42	1.12	Normal standard
	Dy	382.49	132.23	34.57	0.49	0.92	Normal standard
	Ho	67.69	22.13	32.69	0.35	0.86	Normal standard
	Er	174.46	56.74	32.53	0.34	0.88	Normal standard
	Yb	0.00	-	-	-	-	-
	Tm	24.53	8.02	32.71	0.42	1.13	Normal standard
	Lu	9.41	11.15	118.51	0.71	−1.17	Gamma
	Sc	395.63	156.40	39.53	0.13	−0.38	Normal standard
	Y	355.97	239.60	67.31	0.67	−0.27	Lognormal standard

4.2 Logic models for selecting good indicator elements

Associations of elements as the best indicator elements were selected to eliminate unnecessary elements, reduce computational mass, remove fake anomalous contents of elements, and improve efficiency for the correlation of coal seams. As a result, 256 coal samples (including 25 major and trace elements and 16 rare earth elements) were analyzed and processed.

As a result, the major and trace element concentrations are Mg > V > As > Ca > Zn > Cr > Co > K > Na > Sr > Fe > Ge > Re > U > Mo > Th > Ga > Al > Ti > Ba > Ni > Pt > Mn > REE > Pb for coal seams in the study area (Tab. 4). In particular, 17 elements including Mg, V, As, Ca, Zn, Cr, Co, K, Na, Sr, Fe, Ge, Re, U, Mo, Th, and Ga elements account for more than 90%, representing a clear association for coal seams. Therefore, these elements can be selected as the good indicator ones for coal seams, whereas the remaining elements only account for less than 10%. As the whole, rare-earth elements in coal include 11 elements that also account for more than 90%, namely, Yb, Sc, Ho, Er, Tm, Lu, Y, Tb, Pr, Dy, and Sm, while the remaining elements make up less than 10%. These rare earth elements can also be selected as good indicator elements.

Tab. 4. Frequency analysis of the content of the elements for the coal samples in the study area.

Major and trace elements				Rare earth elements			
Element	Amount of information (AI)	Information combination (IC)	Probability [%]	Element	Amount of information (AI)	Information combination (IC)	Probability [%]
Mg	0.307	0.307	27.6	Yb	0.325	0.325	31.49
V	0.283	0.418	37.5	Sc	0.313	0.451	43.70
As	0.277	0.501	45.0	Ho	0.309	0.547	53.00
Ca	0.276	0.572	51.4	Er	0.309	0.628	60.85
Zn	0.259	0.628	56.4	Tm	0.308	0.700	67.83
Cr	0.256	0.678	60.9	Lu	0.308	0.764	74.03
Co	0.249	0.722	64.9	Y	0.245	0.803	77.81
K	0.237	0.760	68.3	Tb	0.238	0.837	81.11
Na	0.234	0.795	71.4	Pr	0.238	0.870	84.30

Major and trace elements				Rare earth elements			
Element	Amount of information (AI)	Information combination (IC)	Probability [%]	Element	Amount of information (AI)	Information combination (IC)	Probability [%]
Sr	0.230	0.828	74.4	Dy	0.232	0.901	87.31
Fe	0.229	0.859	77.2	Sm	0.227	0.929	**90.02**
Ge	0.221	0.887	79.7	Gd	0.225	0.956	92.64
Re	0.220	0.914	82.1	Eu	0.204	0.977	94.67
U	0.216	0.939	84.4	Ce	0.197	0.997	96.61
Mo	0.215	0.963	86.5	La	0.195	1.016	98.45
Th	0.213	0.987	88.7	Nd	0.183	1.032	100
Ga	0.201	1.007	**90.5**				
Al	0.194	1.026	92.2				
Ti	0.191	1.043	93.7				
Ni	0.187	1.060	95.2				
Ba	0.187	1.076	96.7				
Pt	0.155	1.087	97.7				
Mn	0.152	1.098	98.7				
REE	0.138	1.106	99.4				
Pb	0.122	1.113	100				

Note: Amount of information (AI) and Information combination (IC) are mentioned in Hung (et al., 2020) [25].

4.3 Established correlation coefficients between the good indicator elements

The results of correlation analysis can be used to form the pair correlation matrix of the best indicator elements in the coal samples of the study area and each coal seam. The elements of the pair correlation matrix among the good indicator elements are presented in Tables 5, 6, 7, 8, 9, 10, and 11. Among the indicator element associations of the coal seams, the Ca-Mg-Sr, Cr-V-Ga-K-As, and Th-Mo elements display close associations (Tab. 5). The calculated results of each coal seam's associations are also similar to those in the coal deposit. Still, their relation levels vary, particularly in the coal seam TV3-2. Two associations of major and trace elements (including Mg-Ca-Sr, and K-V-Cr-Zn-Mo-Th associations) are established. However, these associations have a closely inverse relationship with each other (Tab. 6). For major and trace elements of the coal seam TV3-1, the Mg-Ca-Cr-K and V-Sr-Mo-Th associations are established. Still, their elements are inversely related; Zn content is inversely related to all other elements (Tab. 7). The contents of the major and trace elements in coal seam TV2 showed two associations as good indicator elements: Ca-Mg-Sr-Mo and V-Zn-K-Th associations. Especially, Cr has a loosely inverse relationship with all elements, except Mo, which has a relatively close inverse relationship with all other elements (Tab. 8). Besides, rare earth elements in coal like the Ho-Er-Tm-Lu-Dy-Sc-Yb-Y association have a very close relationship that is significant in the correlation of coal seams in the study area (Tab. 9).

Herein, correlation coefficient analysis for elements as good indicators in the coal rashing is presented in Tables 10 and 11. Three associations (including V-Zn -Cr, Co-Zn-Ge, and Ge-Co) have a relatively close direct relationship between major and trace elements. The Ge-Co association has an inverse relationship with the Ge element (Tab. 10). Concerning rare earth elements, the results of correlation analysis proved that the group of elements, including Ho-Er-Tm-Dy-Tb, has a very close correlation and is considered an association of elements that go together (Tab. 11).

Generally, the correlation results clearly show that associations of major and trace elements as good indicator elements in the coal deposit and coal seams have relatively clear differences. These associations are often inversely related. Thus, the associations mentioned above of major and trace elements and rare earth elements are important groups that correlate coal seams in the study area.

Tab. 5. The correlation coefficient for major and trace elements as indicator elements in the coal seams.

	Na	Mg	K	Ca	V	Cr	Fe	Co	Zn	Ga	Ge	As	Sr	Mo	Re	Th	U
Na	1																
Mg	**0.475**	1															
K	0.061	0.035	1														
Ca	**0.432**	**0.867**	-0.101	1													
V	-0.075	0.036	0.244	-0.007	1												
Cr	0.058	-0.034	**0.766**	-0.048	**0.571**	1											
Fe	0.066	0.15	0.179	-0.139	0.266	0.147	1										
Co	0.199	0.187	0.251	0.206	0.037	**0.389**	0.246	1									
Zn	0.16	**0.402**	0.333	**0.433**	0.236	**0.437**	-0.041	0.27	1								
Ga	0.196	-0.055	**0.687**	-0.244	**0.533**	**0.769**	**0.394**	0.255	0.055	1							
Ge	0.051	**-0.369**	0.036	**-0.415**	0.227	0.086	0.051	-0.176	-0.065	0.283	1						
As	-0.07	-0.057	**-0.367**	-0.038	**0.645**	-0.025	0.027	-0.311	0.062	0.004	**0.503**	1					
Sr	**0.394**	**0.726**	0.029	**0.875**	0.138	0.115	-0.162	0.281	**0.362**	-0.021	**-0.393**	-0.054	1				
Mo	0.136	-0.002	0.263	0.017	**-0.492**	-0.049	-0.033	0.184	-0.114	0.04	-0.075	**-0.66**	0.039	1			
Re	0.22	**0.353**	0.293	0.199	0.313	**0.389**	0.135	-0.03	0.144	**0.449**	-0.332	-0.015	0.33	-0.184	1		
Th	-0.003	-0.007	0.284	0.001	-0.334	0.011	0.041	0.212	-0.099	0.082	-0.256	**-0.645**	0.08	**0.941**	-0.047	1	
U	0.275	0.171	-0.011	0.106	-0.115	-0.032	-0.167	-0.006	-0.152	0.158	**-0.443**	-0.313	0.127	0.235	**0.431**	0.338	1

Note: Significant values of correlation coefficients are highlighted in bold-typed.

http://doi.org/10.29227/IM-2021-02-11

Received: 04 Jun 2021, Accepted: 14 Sep 2021, Published: 10 Nov 2021

Tab. 6. The correlation coefficient for major and trace elements as indicator elements in the coal seam TV3-2.

	*Mg	*K	Ca	*V	*Cr	*Zn	*Sr	*Mo	Th
*Mg	1.00								
*K	−0.33	1.00							
Ca	**0.69**	−0.89	1.00						
*V	−0.36	**0.99**	−0.91	1.00					
*Cr	−0.70	**0.70**	−0.92	**0.73**	1.00				
*Zn	−0.51	**0.78**	−0.70	**0.76**	0.39	1.00			
*Sr	0.86	−0.72	**0.89**	−0.73	−0.78	−0.82	1.00		
*Mo	−0.01	**0.87**	−0.72	**0.88**	**0.67**	0.38	−0.38	1.00	
Th	−0.35	**0.99**	−0.91	**0.99**	**0.74**	**0.74**	−0.72	**0.89**	1.00

Tab. 7. The correlation coefficient for major and trace elements as indicator elements in the coal seam TV3-1.

	*Mg	*K	Ca	*V	*Cr	*Zn	*Sr	*Mo	Th
*Mg	1.00								
*K	0.26	1.00							
Ca	**0.99**	0.26	1.00						
*V	−0.65	−0.72	−0.69	1.00					
*Cr	**0.67**	**0.52**	**0.67**	−0.59	1.00				
*Zn	−0.11	0.32	−0.07	−0.16	0.32	1.00			
*Sr	−0.64	−0.71	−0.67	**0.98**	−0.52	−0.13	1.00		
*Mo	−0.64	−0.67	−0.68	**0.96**	−0.70	−0.30	**0.92**	1.00	
Th	−0.53	−0.58	−0.56	**0.81**	−0.26	−0.01	**0.83**	**0.69**	1.00

Tab. 8. The correlation coefficient for major and trace elements as indicator elements in the coal seam TV2.

	*Mg	*K	Ca	*V	*Cr	*Zn	*Sr	*Mo	Th
*Mg	1.00								
*K	0.40	1.00							
Ca	**0.82**	−0.14	1.00						
*V	0.11	**0.75**	−0.33	1.00					
*Cr	−0.41	−0.02	−0.38	0.07	1.00				
*Zn	−0.16	**0.57**	−0.48	**0.78**	−0.02	1.00			
*Sr	**0.85**	0.03	**0.91**	−0.17	−0.39	−0.50	1.00		
*Mo	0.29	−0.19	**0.53**	−0.26	−0.51	−0.20	0.46	1.00	
Th	0.26	**0.88**	−0.19	**0.83**	0.13	**0.80**	−0.17	−0.22	1.00

Tab. 9. The correlation coefficients for the rare earth elements as indicators in coal seams.

	Pr	Sm	Tb	Dy	Ho	Er	Tm	Sc	Yb	Y	Lu
Pr	1										
Sm	0.31	1									
Tb	**0.884**	0.313	1								
Dy	**0.813**	0.232	**0.987**	1							
Ho	**0.763**	0.199	**0.971**	**0.996**	1						

144   *Journal of the Polish Mineral Engineering Society, No.2, Vol.1, 2021*

	Pr	Sm	Tb	Dy	Ho	Er	Tm	Sc	Yb	Y	Lu
Er	**0.737**	0.190	**0.960**	**0.992**	**0.999**	1					
Tm	**0.711**	0.186	**0.951**	**0.986**	**0.995**	**0.998**	1				
Sc	**0.692**	0.179	**0.927**	**0.947**	**0.953**	**0.951**	**0.958**	1			
Yb	0.464	0.112	**0.755**	**0.793**	**0.813**	**0.815**	**0.844**	**0.904**	1		
Y	**0.738**	0.179	**0.958**	**0.990**	**0.998**	**0.999**	**0.995**	**0.940**	**0.801**	1	
Lu	**0.704**	0.183	**0.947**	**0.983**	**0.994**	**0.996**	**0.999**	**0.959**	**0.849**	**0.993**	1

Tab. 10. The correlation coefficient for trace elements as good indicators in the coal rashing.

	V	Cr	Co	Zn	Ge	Re	Th
V	1						
Cr	**0.509**	1					
Co	0.093	0.022	1				
Zn	**0.323**	**0.404**	**0.365**	1			
Ge	0.077	0.059	**0.651**	–0.094	1		
Re	–0.101	0.282	–0.229	**0.440**	–0.534	1	
Th	–0.044	0.103	–0.184	0.096	–0.344	0.327	1

Tab. 11. The correlation coefficient for rare earth elements as good indicator elements
in the coal rashing.

	La	Ce	Nd	Gd	Tb	Dy	Ho	Er	Tm	Sc
La	1									
Ce	–0.068	1								
Nd	–0.363	0.057	1							
Gd	–0.005	0.136	0.25	1						
Tb	0.306	0.167	0.239	0.281	1					
Dy	0.319	0.248	0.205	0.299	**0.8**	1				
Ho	0.344	0.196	0.186	0.257	**0.812**	**0.936**	1			
Er	0.327	0.198	0.194	0.256	**0.793**	**0.914**	**0.993**	1		
Tm	0.322	0.2	0.184	0.26	**0.781**	**0.892**	**0.98**	**0.993**	1	
Sc	0.043	0.198	0.163	0.412	0.12	0.118	0.178	0.225	0.265	1

4.4 Division and correlation of coal seams between relatively homogeneous blocks

The results of calculating the similarity $E(x_i, x_j)$ between each pair of coal seams and grouping the coal seams between relatively homogeneous blocks (blocks I and II) are summarized in Tables 12 and 13.

Comparing similarity coefficient H with similarity coefficient E (Tabs. 12 and 13) shows that the coal seam TV3-6 is unsuitable. Thus, it is necessary to repeat from step 1 to step 10 of the multivariate statistic methods. The comparison results are presented in Table 14, which shows that the coal seam TV3-6 is unreasonable; hence it should be considered and adjusted.

Tab. 12. Similarity coefficients (E) between coal seams.

Coal seam	Similarity coefficient E	Similarity coefficient H				
			Group I	Group II	Group III	Group IV
TV3-6	66	**Group I**	100	63	65	
		Group II	63	100	84	
		Group III	65	84	100	
TV2-11	79	**Group I**	100	85	76	89
		Group II	85	100	82	90

	Group III	76	82	100	80
	Group IV	89	90	80	100

Tab. 13. The grouping results of boreholes in coal seams.

Coal seam		Borehole groups
TV3-6	Group I	LK.84-SH, LK.100-SH, LK.104-SH, LK.107-SH, LK.108-SH, LK.109-SH
	Group II	LK.97-SH (SH97/14, SH97/15), LK.106-SH
	Group III	LK.97-SH (SH97/16), LK.102-SH
TV2-11	Group I	LK.84-SH, LK.102-SH, LK.108-SH
	Group II	LK.97-SH (SH90/27), LK.97-SH, LK.98-SH, LK.100-SH
	Group III	LK.90-SH (SH90/29, SH90/30), LK.107-SH
	Group IV	LK.104-SH, LK.106-SH, LK.109-SH

Tab. 14. Similarity coefficient of the coal seam TV3-6.

	Group 1	Group 2
Group 1 (I)	100	66
Group 2 (II)+III	66	100

The coal seam TV2-11 is as acceptable as the preliminary correlation step; in other words, the correlation of coal seams conducted in step 4 is reasonable. Moreover, the calculation results show that the coal seams TV3-6a, TV3-6b, and TV3-6c (LK.97-SH) can be grouped into the seam TV3-6, which has a relatively complex structure containing many intercalated layered materials. Besides, the coal seams TV2-11a, TV2-11b, and TV2-11c can be grouped into the coal seam TV2-11 (LK.90-SH and LK.100-SH).

From Tables 11, 12, and 13, it can be seen that the preliminary correlation of main coal seams and coal lenses on 02 relatively homogeneous blocks (block I and block II) of the Nam Thinh-Giao Xuan fault zone is suitable and acceptable. However, the preliminary correlation in step 4 between the 02 blocks in the western and eastern blocks of the Nam Thinh-Giao Xuan fault is not appropriate and needs further consideration.

To detect and correct the errors of the coal seam correlation, many quantitative problems can be used; among them is the multidimensional statistical method proposed by Rodinov (1968) [26]. This suggests that the multidimensional statistical standard (V) of the coal seams can follow the equations mentioned above (10), (11), and (12).

Tab. 15. Comparison of coal seams between the eastern block (block I) and western block (block II) of the Nam Thinh - Giao Xuan fault.

Coal seam comparison		Block II		
		TV3-2	TV3-1	TV2
Block I	TV3-1	13.8	229.4	276.8
	TV2	17.6	29.2	22.1

Comparison results showed the Chi-square standard ($\chi^2_{0.5,28}$) of 15.6, which is greater than the multidimensional statistical standard (V) of 13.8 (Tab. 15). This indicates that the seam TV3-1 in the western block (block I) corresponds to the seam TV3-2 in the eastern block (block II) of the Nam Thinh - Giao Xuan fault zone. These results are also completely different from the seam TV3-1 in the eastern block of Nam Thinh-Giao Xuan fault as the results of preliminary correlation in step 4.

Based on the results of step 5 and geological and geophysical data, the unreasonable correlation of coal seams on the geological cross-sections is adjusted. Then step 5a and step 5b are repeated many times until the multidimensional statistical comparison problem (step 5) satisfy comparable standards.

5. Conclusions

Based on 256 geochemical coal and rashing samples of 13 boreholes, several statistical, multivariate, and correlation analysis methods were used to identify the correlation of the coal seams from the Tien Hai area in Northern Vietnam. Throughout this study, the conclusions can be drawn as follows.

The Mg, V, As, Ca, Zn, Cr, Co, K, Na, Sr, Fe, Ge, Re, U, Mo, Th, and Ga elements are strong indicator elements of the major and trace elements in coals and most of them comply with the normal or lognormal distribution, according to the frequency analysis results. Besides, for rare earth elements in coal samples, the Yb, Sc, Ho, Er, Tm, Lu, Y, Tb, Pr, Dy, and Sm elements are also good indicator elements, suggesting that these elements can be selected as the indicator ones for identifying correlation of the coal seams. Additionally, the correlation matrix and cluster analyses can divide the element associations into coal seams (i.e., the Mg-Ca-Sr, and K-V-Cr-Zn-Mo-Th associations of the major and trace elements) V-Zn-Cr, Co-Zn-Ge, and Ge-Co associations in rare earth elements) in the Tien Hai area.

Next, the similarity degree between studied objects is used to group boreholes in coal seams. Generally, the results of identifying the correlation of coal seam TV2-11 are suitable and acceptable; the coal seams TV3-6a, TV3-6b, and TV3-6c can be grouped into the coal seam TV3-6.

Finally, the statistical analysis of the geological and geophysical data from core samples shows a remarkable degree of correlation of the coal seams in the region due to the association with the good indicator elements that are established during the statistical process. Furthermore, geochemical coal data can help to evaluate the indicator elements of the major and trace elements, and rare earth elements in coal seams, and coal rashing of adjoining and pillar rocks in the Tien Hai area, northern Vietnam.

6. Acknowledgments

7. References

1. Dickinson, W.R., and Suczek, C.A., 1979. Plate tectonics and sandstone compositions, American Association of Petroleum Geologists Bulletin, 110, 1268-1280.

2. Dickinson, W.R., and Valloni, R., 1980. Plate settings and provenance of sands in modern ocean basins, Geology, 8, 82-86.

3. Potter, E., 1978. Petrology and chemistry of modern big river sands, Journal of Geology, 86, 423-449.

4. Valloni, R. and Maynard, J.B., 1981. Detrital model of Recent deep-sea sands and their relation to tectonic setting: a first approximation, Sedimentology, 28: 75-83.

5. Zhifei, L., Yingchun, W., Shuzheng, N., Xu, J., Rongfang, Q., Daiyong, C., 2019. The differences of element geochemical characteristics of the main coal seams in the Ningdong coalfield, Ordos Basin, Journal of Geochemical Exploration, 202, 77-91.

6. Crook, A.W., 1974. Lithogenesis and geo-tectonics: the significance of compositional variations in flysch arenites (graywackes), in Dott, R.H., and Shaver, R.H., eds., Modern and ancient geosynclinal sedimentation, SEPM Special Publication 19, 304-310.

7. Dyke, M.V., Klemetti, T., Wickline, J., 2020. Geologic data collection and assessment techniques in coal mining for ground control, International Journal of Mining Science and Technology, 30(1), 131-139.

8. Schwab, L., 1975. Framework mineralogy and chemical composition of continental margin-type sandstone: Geology, 3, 487-490.

9. Hoang Ngoc Ky (ed.). Report on the geological mapping and mineral resources in the Hai Phong-Nam Dinh area at a scale of 1:200.000, Department of Geology and Mineral Resources of Vietnam (in Vietnamese), 1999.

10. Tran Van Tri & Vu Khuc (ed.). Geology and Earth Resources of Vietnam. General Department of Geology and Minerals of Vietnam, Publishing House for Science and Technology, Hanoi, 2011.

11. Huyen, T.N., Cuong, D.T., Hieu, T.N., 2019. Non-structural traps in the post-rift succession of Phu Khanh Basin: Classification and Depositional History, Journal of Mining and Earth Sciences 60 (3): 1-9.

12. Nga, H.L., Dong, N.P., Huy, Q.B., 2019. Organic petrology and Rock-Eval characteristics in selected coal samples of the Cau Formation, block 07 Nam Con Son Basin, Journal of Mining and Earth Sciences 60(3): 10-17.

13. Nielsen, L.H., Mathiesen, A., Bidstrup, T., Vejbñk, O.V., Dien, P.T., Tiem, P.V., 1998. Modeling of hydrocarbon generation in the Cenozoic Red River basin, Vietnam: a highly prospective basin, Journal of Asian Earth Science 17(1): 269-294.

14. Dong Van Giap (ed.). Report on the results of Investigation and Assessment of coal resources in the mainland of Red River coal basin and adjacent area, Department of Geology and Mineral Resources of Vietnam (in Vietnamese), 2005.

15. Golonka, J., Krobicki, M., Pająk, J., Giang, N.V., and Zuchiewicz, W. Global Plate Tectonics and Paleogeography of Southeast Asia, Faculty of Geology, Geophysics and Environmental Protection, AGH University of Science and Technology, Arkadia, Krakow, Poland, 2006.

16. Hall, R., 1996. Reconstructing Cenozoic SE Asia. In: Hall, R., Blundell, D. J. (eds). Tectonic evolution of Southeast Asia, vol 106, Geological Society of London Special Publication, London, 203-224.

17. Lee, T.Y., Lawver, L.A., 1995. Cenozoic plate reconstruction of Southeast Asia, Tectonophysics 251:85-138.

18. Vu Xuan Doanh. Report on the degree of coal-bearing in the Hanoi trough (Hung Yen - Thai Binh area), Vietnam Institute of Geosciences and Mineral Resources (in Vietnamese), 1982.

19. Golovenok, V.K., Le Van Chan, 1966. Sediments and formation conditions of Neogene-Quaternary sediments in the Hanoi trough, Vietnam Petroleum Institute, Hanoi (in Vietnamese).

20. Nguyen Van Tinh, Nguyen Phuong, Do Van Nhuan, Nguyen Quoc Phi, Tang Dinh Nam, Dong Van Giap, Dang My Cung, Dang Tran Nhu Thuy. Research on establishing the geological and geophysical parameters as the basis for connection and correlation of the coal seams in the Red River coal basin, a case study of the Tien Hai area, northern Vietnam, Vietnam Institute of Geosciences and Mineral Resources, Vietnam Ministry of Natural Resources and Environment (in Vietnamese), 2015.

21. Ngo Tat Chinh. Report on the results of prospecting mineral deposits in the Khoai Chau-Chau Giang-Hung Yen area, Department of Geology and Mineral Resources of Vietnam (in Vietnamese), 1987.

22. Dong Van Nhi, Nguyen Phuong, 1988. Using geological and geophysical dimensional statistics for identifying correlation of the coal seams in the Quang Ninh area, northeastern Vietnam, Journal of Mineral Economics and Raw Materials (in Vietnamese).

23. Tran Manh Quang. Selecting a reasonable combination of borehole geophysical methods for studying and applying the correlation of the coal seams based on the multivariate statistical models of geologics and geophysics (an example from the Nga Hai coal mine). PhD thesis, Hanoi University of Mining and Geology (in Vietnamese), 1988.

24. Rao, C.R., 1967. Linear statistical inference and its applications, Annals of Mathematical Statistics, 38(1), 281-284.

25. Hung, K.T., Sang, P.N., Phuong, N., Linh, V.T., & Sang, B.V., 2020. Statistical evaluation of the geochemical data for prospecting polymetallic mineralization in the Suoi Thau-Sang Than region, Northeast Vietnam, Geology, Geophysics and Environment, 46(4), 285-299.

26. Rodionov, D.A. Statisticeskie resenija v geologii.: Izd. "Nedra", Moskva, 231 pp (in Russian), 1968.