Journal of Petroleum Science and Engineering 185 (2020) 106616

Contents lists available at ScienceDirect

Journal of Petroleum Science and Engineering

journal homepage: http://www.elsevier.com/locate/petrol

A mathematical method to identify and forecast coal texture of multiple and thin coal seams by using logging data in the Panguan syncline, western Guizhou, China

Ziqi Zhao [a,b], Shu Tao [a,b,*], Dazhen Tang [a,b], Shida Chen [a,b], Pengfei Ren [a,b]

[a] School of Energy Resources, China University of Geosciences (Beijing), Beijing, 100083, PR China
[b] Coal Reservoir Laboratory of National Engineering Research Center of Coalbed Methane Development & Utilization, Beijing, 100083, PR China

ARTICLE INFO

Keywords:
Coal texture
Multiple and thin coal seam
Geophysical logging
Linear discriminant analysis

ABSTRACT

There are abundant coalbed methane (CBM) resources in the western Guizhou Province, China. The high efficiency development of CBM in western Guizhou is restricted to a great extent by multiple and thin coal seams, the high in-situ stress and the complex coal texture. By using the mathematical principle of Linear Discriminant Analysis (LDA) method, 1508 logging data obtained from three CBM wells are related to their geophysical logging responses. Four types of well-logging curves of the natural gamma ray (GR), densities (DEN), acoustic logging (AC) and deep lateral resistivity (LLD) were chosen to analyze coal texture. The mathematical method reflecting the relationship between typical logging data and coal texture was established, which has a high accuracy for identifying the coal texture of medium-thin coal seams. The results show that tectonic coal in western Guizhou is developed and complicated. The coal texture predicted by the established method is in good agreement with that of actual coal core, which plays an important role in the joint and efficient development of CBM in the multiple and thin coal seams in western Guizhou.

1. Introduction

As an alternative clean energy, coalbed methane (CBM) has become an important component of unconventional natural gas (Fu et al., 2009a; Tao et al., 2012, 2019a; Xu et al., 2016). Permeability of coal reservoirs is an important factor affecting CBM yield (Connell et al., 2010; Tao et al., 2014, 2019b), while coal texture largely determines coal reservoir permeability and closely related to production efficiency (Rutqvist and Stephansson, 2003; Min et al., 2004; Hou et al., 2017; Wang et al., 2018). Moreover, coal texture can also be used to predict gas concentration (Ju and Li, 2009). Therefore, it is very important to clarify the changes of coal texture in coal reservoir for the development and evaluation of CBM.

Coal texture can be divided into undeformed coal, cataclastic coal, granulated coal and mylonitized coal (Xue et al., 2012; Ren et al., 2018; Fu et al., 2009b). Cataclastic coal seam is the most suitable type in the process of CBM development, followed by undeformed coal, while granulated and mylonitized coals are the worst (Ren et al., 2018). Nowadays, the recognition of coal texture is mainly based on the visual observation and description of core samples to define the degree of coal damage. During the CBM development, due to the high cost of obtaining coal cores (Xu et al., 2016), the accurate prediction of coal texture can't be realized in areas where the core is undrilled, while geophysical well logging is an economic and convenient method (Li et al., 2016). According to the support of a huge mass of well logs such as caliper logging (CAL), natural gamma ray (GR), densities (DEN), acoustic logging (AC) and deep lateral resistivity (LLD), the well logging provides continuity evaluation of formation characteristics (Yegireddi and Bhaskar, 2009; Siregar et al., 2017), which can be used to deeply understand the changes of formation property and to establish a link between logging data and coal texture. The logging response of coal is characterized by low natural gamma, low density, high acoustic time difference, and high resistivity (Rai et al., 2004; Xu et al., 2016; Tao et al., 2018a, b; Luo et al., 2019). With the increase of coal destruction degree, the values of natural gamma and density tend to decrease, while the values of acoustic time difference and resistivity tend to increase (Teng et al., 2015; Ren et al., 2018; Wang et al., 2018).

Previous researchers used various logging data to calculate coal permeability and ash yield, to identify coal facies, coal macrolithotypes and coal property (Yegireddi and Bhaskar, 2009; Li et al., 2016; Zhou

* Corresponding author. School of Energy Resources, China University of Geosciences (Beijing), Beijing, 100083, PR China.
E-mail address: taoshu@cugb.edu.cn (S. Tao).

https://doi.org/10.1016/j.petrol.2019.106616
Received 29 January 2019; Received in revised form 19 September 2019; Accepted 24 October 2019
Available online 29 October 2019

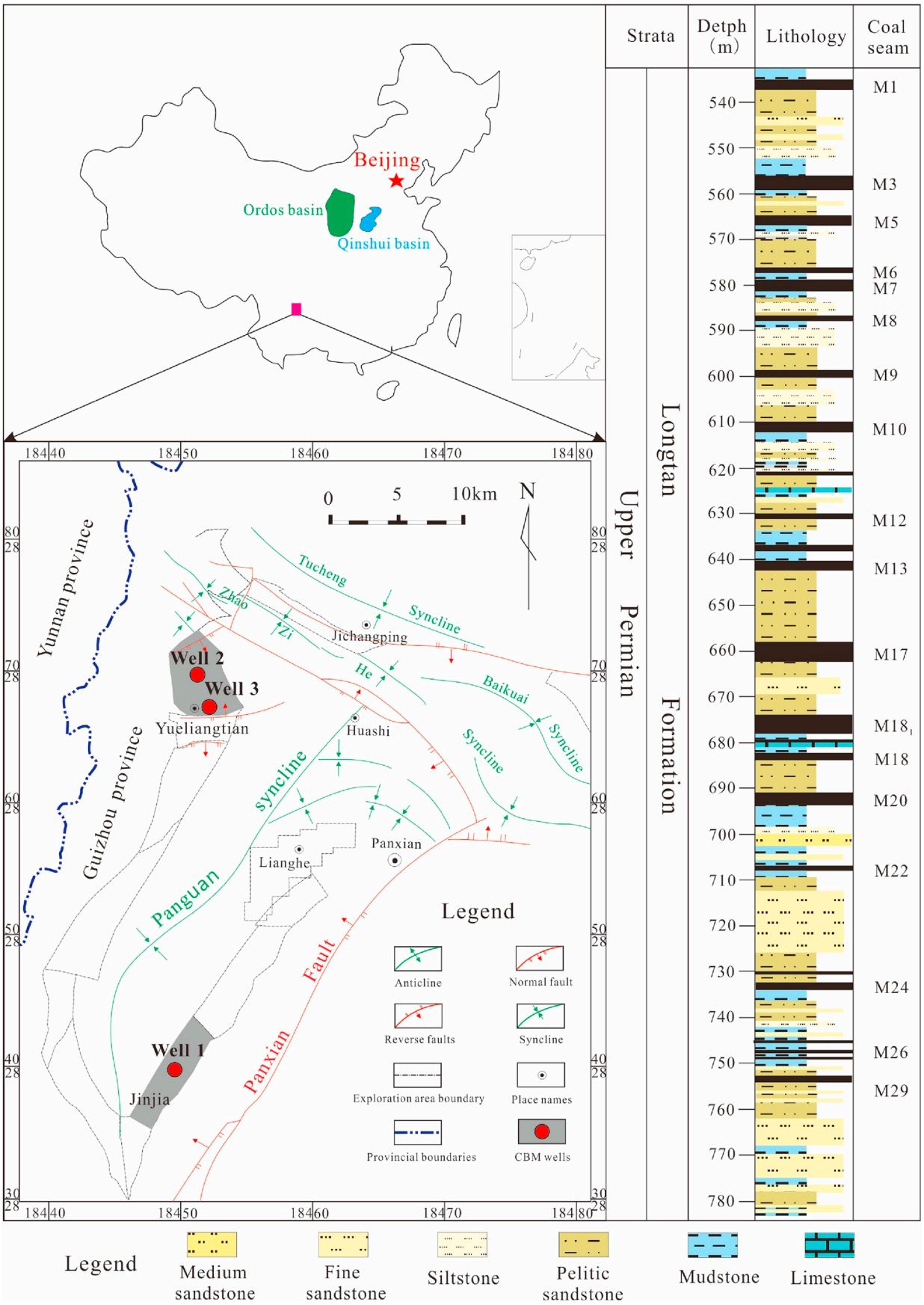

Fig. 1. Location of the Panguan syncline and the longitudinal distribution of coal seams in Well 1.

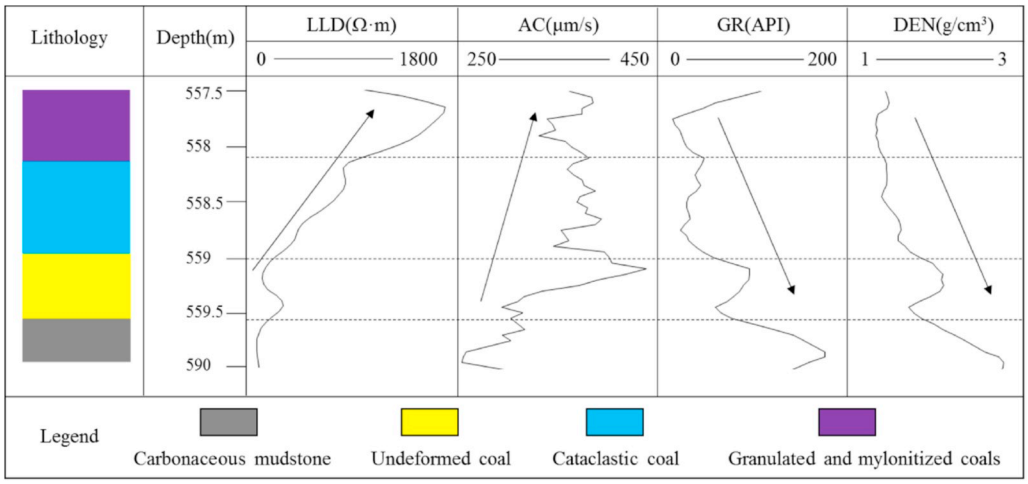

Fig. 2. Logging response characteristics of different coal texture in Well 1.

and O'Brien, 2016; Siregar et al., 2017; Tao et al., 2019c), and coal texture can be characterized as well (Fu et al., 2009a; Teng et al., 2015; Li et al., 2016; Wang et al., 2018). At the same time, the effective prediction of coal texture in adjacent wells with similar coal-forming environment, tectonic background and material composition can be predicted (Fu et al., 2009a). However, the common identification of coal texture is mainly aimed at medium-thick coal seams, lacking in-depth study on thin coal seams. In addition, the in-situ stress condition in western Guizhou is complex (Chen et al., 2017, 2018; Zhou et al., 2019), and the coal seams are characterized by multiple layers (Qin et al., 2014; Shen et al., 2016), many of the previous evaluation methods are not applicable.

What's more, compared with previous unsupervised classification methods, such as cluster analysis and principal component analysis, LDA method can use prior knowledge and experience of categories, in other words, through the sample data of the known coal development status in this area, the unknown coal body structure can be deduced. Different from previous empirical formula method, for the actual research area, LDA establishes the calculation formula suitable for the region, which can more accurately restore the geological conditions of the area and realize the accurate classification of coal texture.

The current study takes the multiple and thin coal seams from the Panguan syncline in western Guizhou as research object. Based on the analysis and selecting of logging parameters, a mathematics method was used to identify and predict the coal texture of these special coal seams in the study area, which provides basis for the reasonable layer selection and efficient development of CBM in western Guizhou.

2. Geological setting

Panguan syncline is located in the west of Guizhou Province, which covers an area of approximately 605 km^2 (Chen et al., 2018), belonging to the southwest margin of Yangzi block. The Yangzi block underwent multi-period tectonic activities including the Indosinian, Yanshan and Himalayan movements, resulting in numerous folds, faults, and complex coal texture. Moreover, it controls the sedimentary evolution of coal seams and the distribution of CBM reservoirs in the current area.

Series of Late Permian coal seams mainly developed in the upper and lower part of Longtan Formation in the Panguan syncline. The thickness of Longtan Formation is about 180 m, containing more than 40 layers of coal, and 9 of them are local minable coal seams. The thickness of most single coal seam is around 1 m. The coal-bearing ratio is 13.98% on average and the coal rank changes greatly ($0.89\% < R_o < 2.18\%$). In this study, the wells in Jinjia (Well 1) and Yueliangtian (Well 2 and Well 3) Blocks are selected to do a further research (Fig. 1).

3. Method and procedure

3.1. Well logging parameter selection

To achieve the quantitative and accurate identification of coal texture, the logging response of different coal texture should be taken as the basis, and the logging data are also be used for comparison and division. With the increase of coal destruction degree, the logging curve also varies regularly (Fu et al., 2009b; Yao et al., 2011; Ren et al., 2018). Tectonic coals (cataclastic coal, granulated and mylonitized coals) can amplify logging diameter which is reflected in lower GR and DEN, higher AC and LLD in logging curves (Fu et al., 2009b; Teng et al., 2015; Wang et al., 2018).

Considering that single logging curve is susceptible to drilling fluid and engineering conditions, if the classification of coal texture is conducted according to the data of single logging curve, its accuracy will be greatly affected. Therefore, this study selects various logging curve as the classification basis to eliminate the influence of single-factor error and improves the classification accuracy. Because the CAL is easily affected by drilling engineering, the GR, DEN, LLD and AC logging parameters are selected as the indexes for discriminant analysis. As shown in Fig. 2, the LLD and AC show positive correlation with progressive damage of coal, while the GR and DEN show negative correlation. Moreover, the carbonaceous mudstone is obviously different from other curves.

3.2. The procedure of identification method by LDA theory

Linear Discriminant Analysis (LDA) is a kind of multivariate statistical analysis method (Cai et al., 2018; Wen et al., 2018), which can distinguish each sample accurately based on the thought of variance analysis, also known as Fisher Linear Discriminant. Firstly, LDA categorizes the known sample data, and then select the variables that can describe the classified samples more comprehensively. Secondly, according to a certain discriminant criterion, the undetermined coefficient of the discriminant function is determined based on the data of known sample data (training sample set), and one or more discriminant functions are established. The category of undetermined samples (test sample set) can be determined by the discriminant function with little loss of information.

Meanwhile, LDA uses projection method to project points in the p-dimensional space into the one-dimensional space to reduce the dimension (Ye et al., 2017; Abuzeina and Al-Anzi, 2018). In the original coordinate system, the samples which are difficult to be classified are distinguished after the projection. The result of discrimination is able to

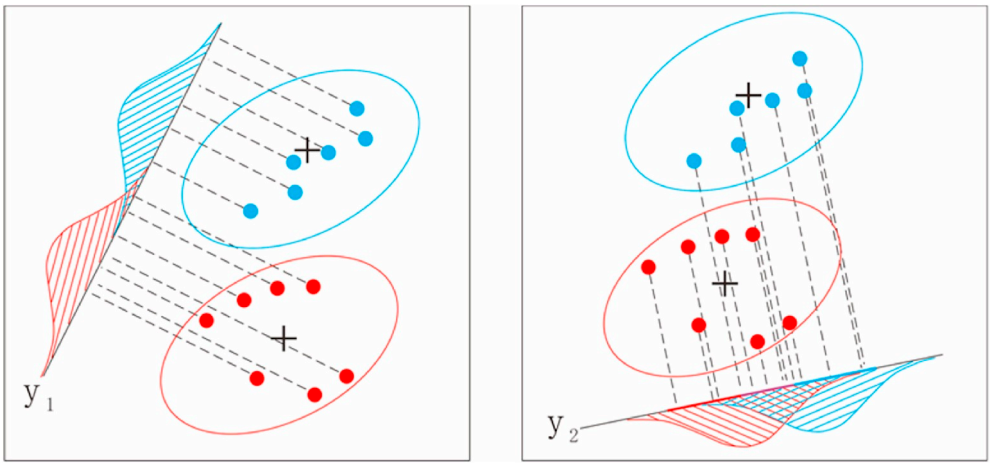

Fig. 3. Two-dimensional projection diagram (y_1 and y_2 are different projection directions of data).

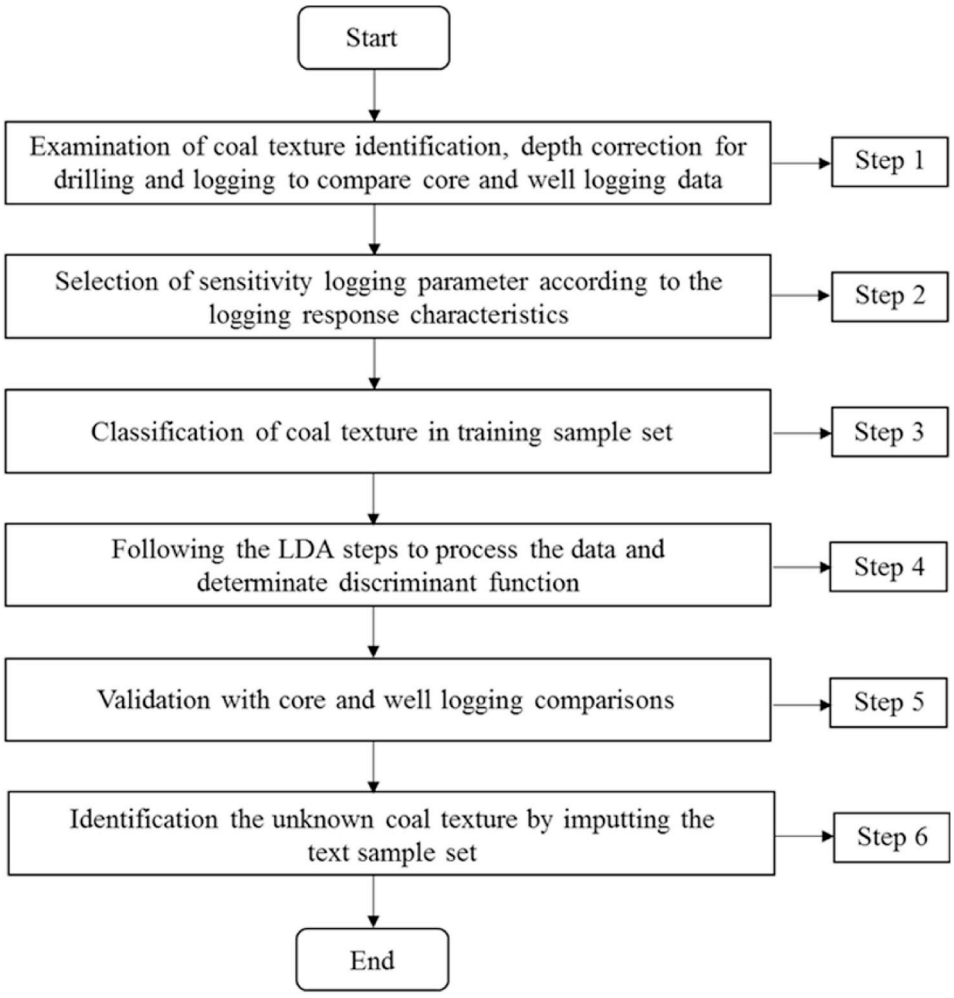

Fig. 4. Workflow of coal texture identification.

maximize the separation of the groups and make the discrete form within groups minimum (Wen et al., 2018). Using a two-group general type in Fig. 3 as an example, it can be seen that the classification of projection on y_1 is obviously better than that of y_2. For a new sample, it can be classified by the distance between the unknown sample and the known sample, and higher discrimination accuracy can be obtained easily.

SPSS (Statistical Product and Service Solutions) software is used for discriminant analysis to achieve precise inversion of coal texture according to the four variables of GR, LLD, DEN and AC (Nie and Norman, 1975; Guo, 1999; Su et al., 2000). SPSS is a software package used for statistical analysis, data mining, predictive analysis and decision support tasks. It has played a great role in various fields of scientific research. The software version used in this article is IBM SPSS Statistics 20.0, the

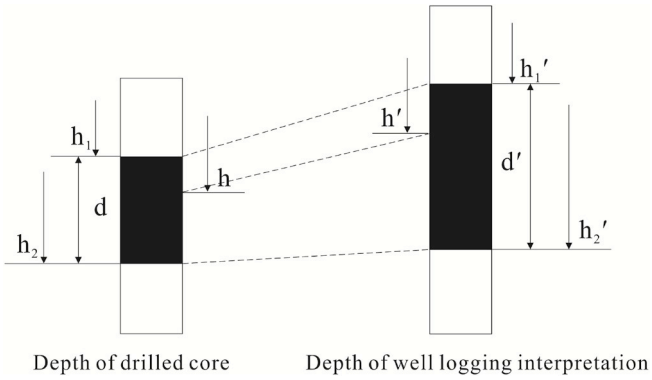

Fig. 5. Depth correction for drilling and logging after Fu et al. (2009a).

main function module is "Analyze"→ "Classify"→ "Discriminate". The sequential process is shown in Fig. 4 according to the LDA by using the SPSS software. The main steps are: (1) The data of different coal texture samples (training sample sets) determined by classification are selected as the known information. (2) Linear discriminant functions are determined by analyzing LLD, AC, GR, DEN. (3) For the unknown coal textures sample data (testing sample set), the coal textures are classified according to the discriminant criterion.

The detailed mathematical process is as follows (Gao, 2004):

(1) Projection dimension reduction

A group named G_t (t = 1 ,…,*k*) including 4 types (undeformed coal, cataclastic coal, granulated and mylonitized coals, and carbonaceous mudstone) is established, and each type has 4 variables of GR, DEN, LLD, and AC. $X_i^t = (x_{i1}^t, \ldots, x_{i4}^t)'$ (t = 1 , …, k; i = 1, …, n_t), $a = (a_1, \ldots, a_4)'$ is any vector in four-dimensional (GR, DEN, LLD, AC) space, and u(x) = $a'X$ is the projection in the X direction in the normal direction of ***a***. As a result, these logging data become one-dimensional variable.

$$\overline{X} = \frac{1}{n}\sum_{t=1}^{k}\sum_{j=1}^{n_t} X_j^t \tag{1}$$

$$\overline{X}^k = \frac{1}{n_k}\sum_{j=1}^{n_k} X_j^k \tag{2}$$

where $\overline{X}^k$ and $\overline{X}$ are respectively the sample mean and the total sample mean of G_t.

(2) Calculating the inner-class distance and between-class distance after the projection

Unary variance analysis was performed on unary data of group k

$$B_0 = \sum_{t=1}^{k} n_t (a'\overline{X}^t - a'\overline{X})^2 = a'Ba \tag{3}$$

where B_0 is the sum of squares between groups (The distance between different coal textures)

$$\text{That is: } B = \sum_{t=1}^{k} n_t (\overline{X}^t - \overline{X})(\overline{X}^t - \overline{X})' \tag{4}$$

$$A_0 = \sum_{t=1}^{k}\sum_{j=1}^{n_t}\left(a'X_j^t - a'\overline{X}^t\right)^2 = a'Aa \tag{5}$$

where A_0 is the sum of squares within the group (The distance between same coal textures)

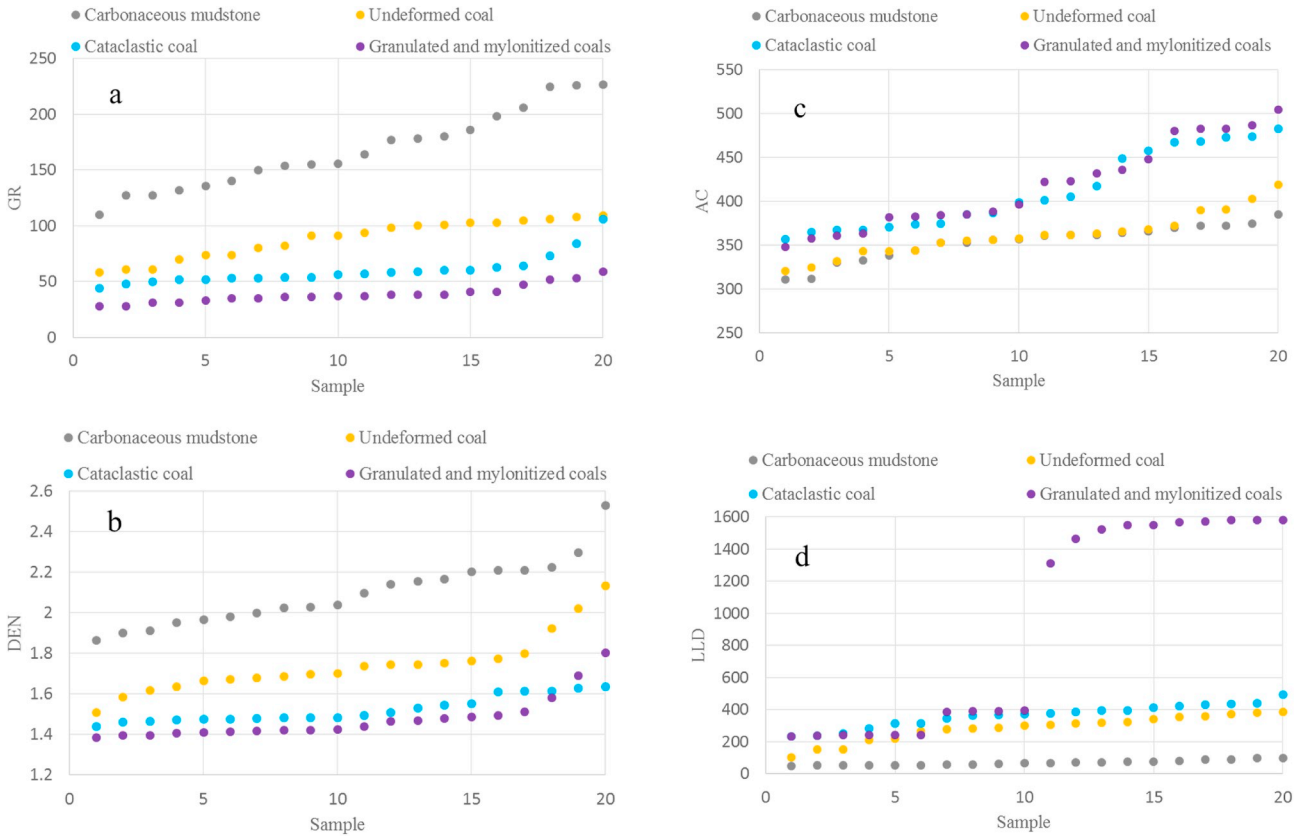

Fig. 6. Different logging values of coal texture in training samples set.

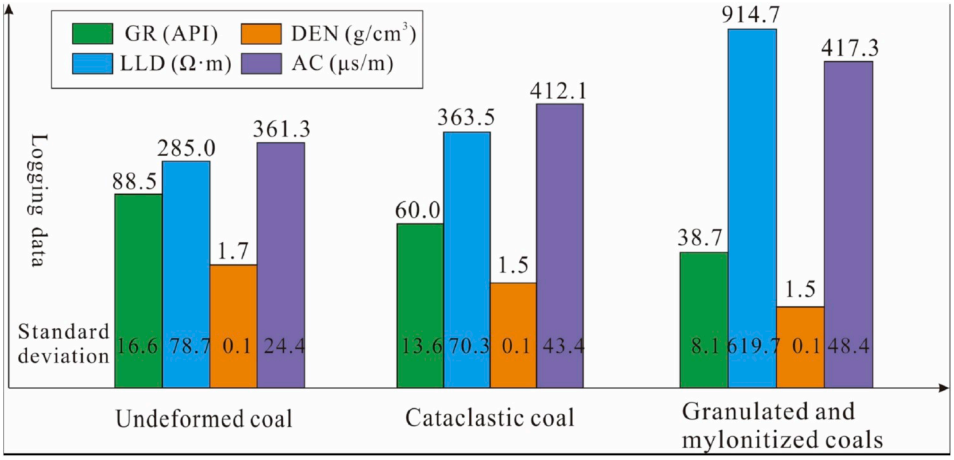

Fig. 7. Average value of logging data of coal texture in training samples set.

Table 1
Training sample set data.

Case Number	Actual Group	Predicted Group	Discriminant Function scores			Case Number	Actual Group	Predicted Group	Discriminant Function scores		
			F1	F2	F3				F1	F2	F3
1	A	A	−1.134	−0.813	0.299	41	C	C	4.947	−0.043	2.049
2	A	A	−1.525	−1.817	0.21	42	C	C	5.352	0.289	2.346
3	A	A	−0.753	−2.041	0.342	43	C	C	5.118	0.183	2.603
4	A	A	−0.331	−1.944	0.487	44	C	C	4.756	0.128	2.767
5	A	A	−0.028	−1.697	0.406	45	C	C	5.531	1.753	1.696
6	A	A	−0.421	−1.484	0.303	46	C	C	5.52	1.888	1.691
7	A	A	−0.999	−1.241	0.119	47	C	C	5.863	2.261	1.52
8	A	A	−0.772	−0.716	−0.275	48	C	C	5.707	2.422	1.503
9	A	A	−0.89	−0.424	−0.499	49	C	C	5.712	2.456	1.461
10	A	A	−3.195	−0.773	0.59	50	C	C	5.593	2.646	1.35
11	A	A	−2.082	−0.391	0.414	51	C	B*	3.71	−0.334	−0.98
12	A	A	−1.067	−0.143	0.114	52	C	C	3.804	−0.432	−0.974
13	A	A	−1.054	−0.293	0.074	53	C	C	4.078	−0.337	−1.049
14	A	A	−1.64	−1.744	0.558	54	C	C	4.067	−0.21	−1.077
15	A	A	−1.456	−1.351	0.641	55	C	C	4.383	0.163	−1.743
16	A	A	−1.233	−0.77	0.541	56	C	C	4.634	0.125	−1.764
17	A	A	−1.404	−0.4	0.472	57	C	C	4.618	0.037	−1.727
18	A	A	−1.778	−0.369	0.492	58	C	C	5.219	0.196	−1.893
19	A	A	−1.091	−0.446	0.349	59	C	C	4.659	−0.079	−1.705
20	A	A	−1.427	−0.478	0.413	60	C	B*	3.711	−0.316	−1.459
21	B	B	2.142	−1.13	0.364	61	D	D	−4.489	−0.429	0.123
22	B	B	1.995	−1.452	0.302	62	D	D	−5.888	−0.16	0.236
23	B	B	1.191	−1.929	0.489	63	D	D	−4.888	−1.707	1.04
24	B	B	0.984	−1.763	0.421	64	D	D	−3.68	−0.382	−0.089
25	B	B	0.927	−1.713	0.367	65	D	D	−5.151	1.268	−0.369
26	B	B	1.174	0.005	−0.449	66	D	D	−5.697	1.049	−0.173
27	B	B	1.598	−0.026	−0.532	67	D	D	−4.835	0.802	−0.22
28	B	B	0.723	0.46	−0.415	68	D	D	−4.245	0.603	−0.189
29	B	A*	0.152	−1.408	0.6	69	D	D	−4.231	−1.302	0.26
30	B	B	1.316	−0.142	−0.508	70	D	D	−4.235	−0.583	0.035
31	B	B	1.874	−0.057	−0.699	71	D	D	−7.556	2.496	−0.146
32	B	B	2.261	0.39	−0.982	72	D	D	−8.004	2.409	−0.1
33	B	B	3.153	0.536	−1.283	73	D	D	−7.549	2.507	−0.257
34	B	B	3.482	0.702	−1.487	74	D	D	−6.303	1.742	−0.254
35	B	B	3.331	0.786	−1.593	75	D	D	−5.985	1.363	−0.125
36	B	B	3.394	0.652	−1.651	76	D	D	−7.21	1.757	−0.142
37	B	C*	3.749	0.586	−1.75	77	D	D	−5.998	1.071	−0.135
38	B	B	3.435	0.348	−1.63	78	D	D	−4.551	0.536	−0.132
39	B	A*	0.361	−0.525	0.175	79	D	D	−4.783	0.483	−0.125
40	B	A*	0.345	−1.014	0.277	80	D	D	−5.008	−0.295	0.086

Notes: groups “A,B,C,D” refer to “undeformed coal”, “cataclastic coal”, “granulated and mylonitized coals”, and “carbonaceous mudstone”, respectively. The meaning of “A,B,C,D” in follow figures and tables are the same as Table 1. “Actual Group” refers to the actual classification of coal texture in coal core. “Predicted Group” is the classification of the LDA. The symbol “*” represents as misjudged case.

Table 2
Typical discriminant function abstracts.

Function	Eigenvalue	Variance (%)	Cumulative (%)	Canonical Correlation	Significant
1	15.507	96.3	96.3	0.969	0.000
2	0.498	3.1	99.4	0.577	0.000
3	0.104	0.6	100.0	0.307	0.024

Notes: "Variance" and "Cumulative" refer to the percentage of the original information that can be interpreted. When the significant is lower than 0.05, it refers to significant difference.

Table 3
Unstandardized canonical discriminant function coefficients.

Variable	Discriminant Function Coefficients		
	F1	F2	F3
GR (API)	−0.042	0.028	−0.001
LLD (Ω·m)	0.003	0.002	0.002
DEN (g/cm^3)	−0.488	−1.467	1.026
AC (μs/m)	0.027	0.009	−0.008
Constant	−7.080	−4.377	0.620

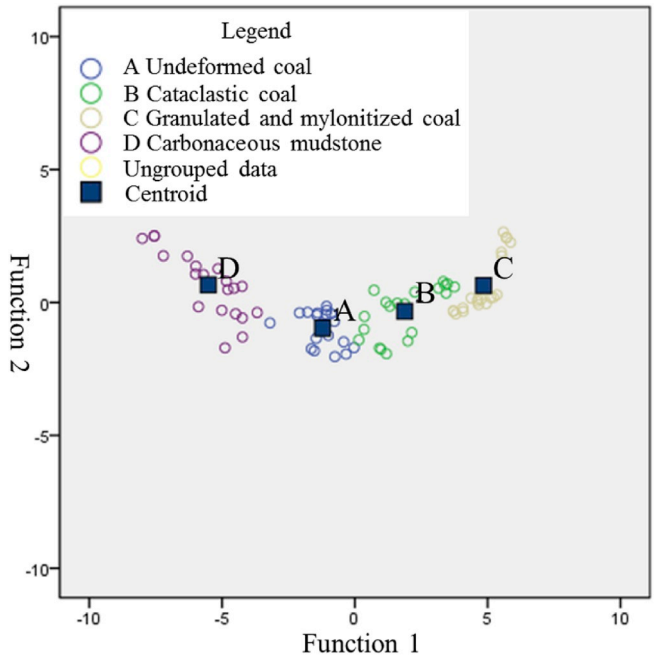

Fig. 8. Typical discriminant function scatter diagram. The X-axis and Y-axis refer to F1 and F2, respectively. The circle "O" represents data of 80 training sample sets. Four different color circles mean they are divided into four groups, which surround the four centroids of "A, B, C, D". (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

$$\text{That is: } A = \sum_{t=1}^{k}\sum_{j=1}^{n_t}\left(X_j^t - \overline{X}^t\right)\left(X_j^t - \overline{X}^t\right)', \tag{6}$$

(3) Construction of the target function

In order to get the best $\boldsymbol{a}$ (optimum projection direction), B_0 is to be large enough, and A_0 is to be as small as possible, therefore, the ratio is as follow:

$$\frac{a'Ba}{a'Aa} \underline{\underline{\text{def}}} \Delta(a) \tag{7}$$

(4) Calculation of $\boldsymbol{a}$ and getting linear discriminant function

Under Fisher's rule, the maximum eigenvalue λ_1 is corresponding to the eigenvector of $\boldsymbol{a}$. The discrimination function is $Y = u_1(X)$.

(5) Discriminate criterion

For an ungrouped sample X (unknown coal texture), its texture is evaluated according to the Mahalanobis distance:

$$\frac{|u_1(X) - \overline{u_1}^{(i1)}|}{\sigma_{i1}} = \min_{t=1,\ldots,k} \frac{|u_1(X) - \overline{u_1}^{(t)}|}{\sigma_t} \tag{8}$$

It needs to choose the greatest efficiency of discriminant function. If there is a unique i1, then judging $X \in G_{i1}$. Otherwise, the discriminant function with lower efficiency is chosen. If there is a unique i2, then judging $X \in G_{i2}$ and so on. Sequentially, the accurate identification of the coal texture of ungrouped samples can be realized.

4. Results and discussion

4.1. Classification of coal texture in training sample set

Core description is a direct method to identify coal texture. Due to the influence of drill stem, there is a deviation between the drilling depth and the depth from logging interpretation (Fu et al., 2009a). Therefore, correction of deviation plays an important role in improving the accuracy of identifying coal texture. The calibration can be calculated from the following relationship in Fig. 5, giving

$$\frac{h_2 - h}{h_2' - h'} = \frac{d}{d'} = \frac{h_2 - h_1}{h_2' - h_1'} \tag{9}$$

where d and d′ are thicknesses of the coal seam measured by drilling and log interpretation. h_1, h_2 and h are the depths of roof, floor, and sampling point of the coal seam obtained during the core. h_1', h_2′and h′ are the depths of roof, floor, and sampling point of the coal seam measured during logging interpretation.

The identification accuracy of logging data in vertical direction can reach 0.05 m, and the classification results of coal texture are checked by drilling core description. 80 coal cores and their geophysical logging data are selected as training sample sets from Well 1 in Jinjia area, Panguan syncline, where the data are divided into four groups of undeformed coals, cataclastic coals, granulated and mylonitized coals, and carbonaceous mudstone, with 20 groups of data in each category. The logging data of GR, DEN, LLD and AC are great difference in different coal textures. The four types of logging curves all have a regular variation trend shown in Fig. 6 and Fig. 7, and the value range of each curve changes obviously, which is consistent with Fig. 2, indicating that these data can be used as the characteristic variable of sample points.

4.2. Logging application according to the LDA theory

80 data of four groups in training sample set are processed according to LDA theory by SPSS software (Table 1), and the typical discriminant function abstracts are as follow (Table 2). The eigenvalue of function 1 is 15.507, which is much larger than that of function 2, suggesting that the classification effect of function 1 is better than that of function 2. The significance of both function 1 and function 2 is less than 0.05, indicating that the discriminant results are significantly difference. Conversely, the classification effect of function 3 is not obvious, thus it has very little information. Actually, Table 2 shows that a joint discriminant function 1 and 2 can explain enough original information (99.4%), showing a higher level of superiority. The two discriminant functions mean that SPSS software reduces all the data to two dimensions eventually.

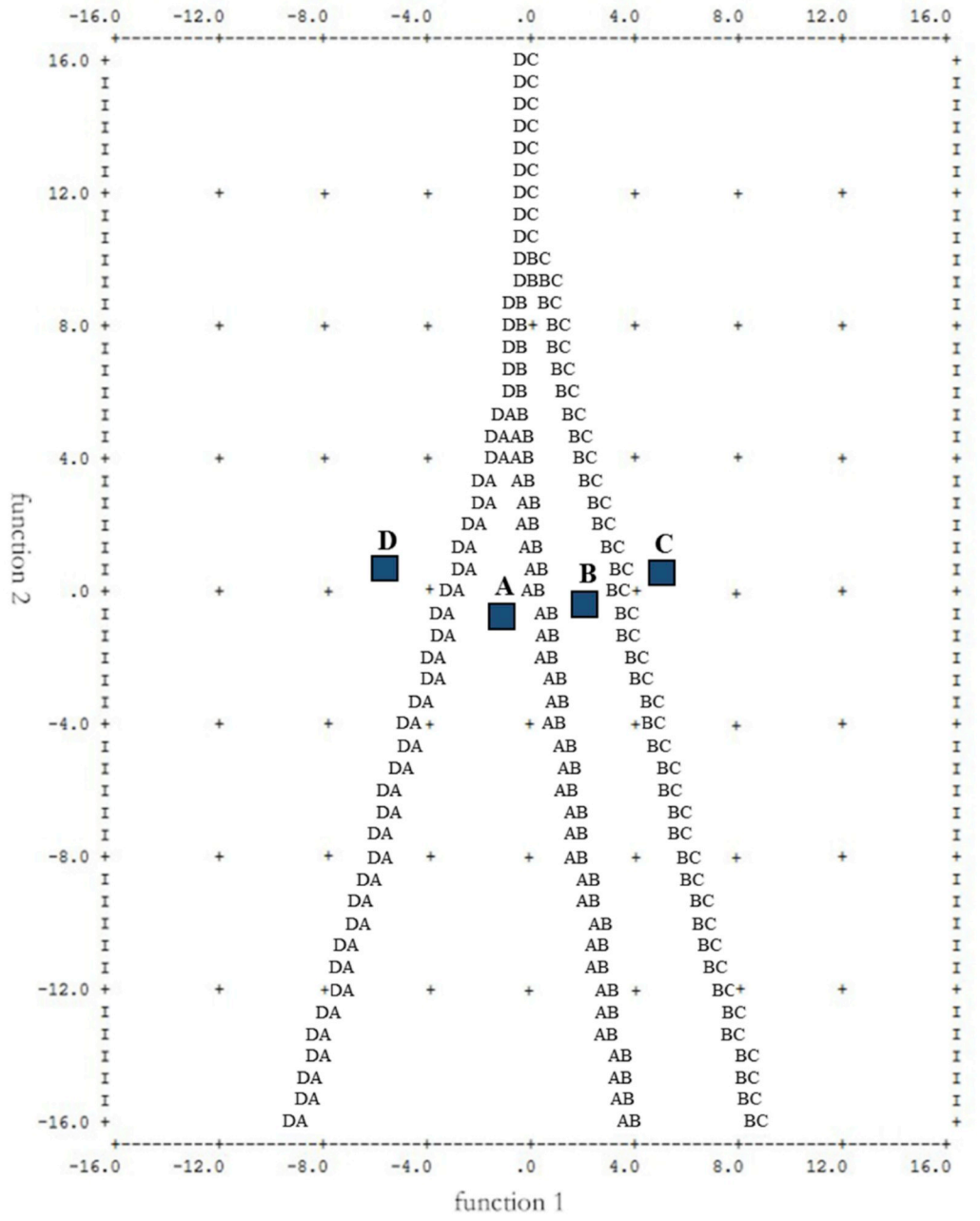

Fig. 9. Typical discriminant function territorial map. The X-axis and Y-axis refer to F1 and F2, respectively. The whole region is divided into four regions of "A, B, C, D", and the boundaries of four regions are defined by Mahalanobis distance.

Table 4
Training sample set classification results.

Actual Group	Predicted Group Member				Total
	A	B	C	D	
A	20	0	0	0	20
B	3	16	1	0	20
C	0	2	18	0	20
D	0	0	0	20	20

Table 5
Cross-validation results.

Actual Group	Predictive Group Member				Total
	A	B	C	D	
A	20	0	0	0	20
B	3	15	2	0	20
C	0	3	17	0	20
D	0	0	0	20	20

Discriminant function coefficients can be concluded by SPSS software showing in Table 3, and the original data of each sample point can be calculated to get the corresponding function coordinate value. The advantage is that the original data does not need to be processed. The original data can be directly substituted into the function to obtain the function value and to determine its location.

Function 1:

$$F_1 = -0.042*\mathrm{GR} + 0.003*\mathrm{LLD} - 0.488*\mathrm{DEN} + 0.027*\mathrm{AC} - 7.080$$

Function 2:

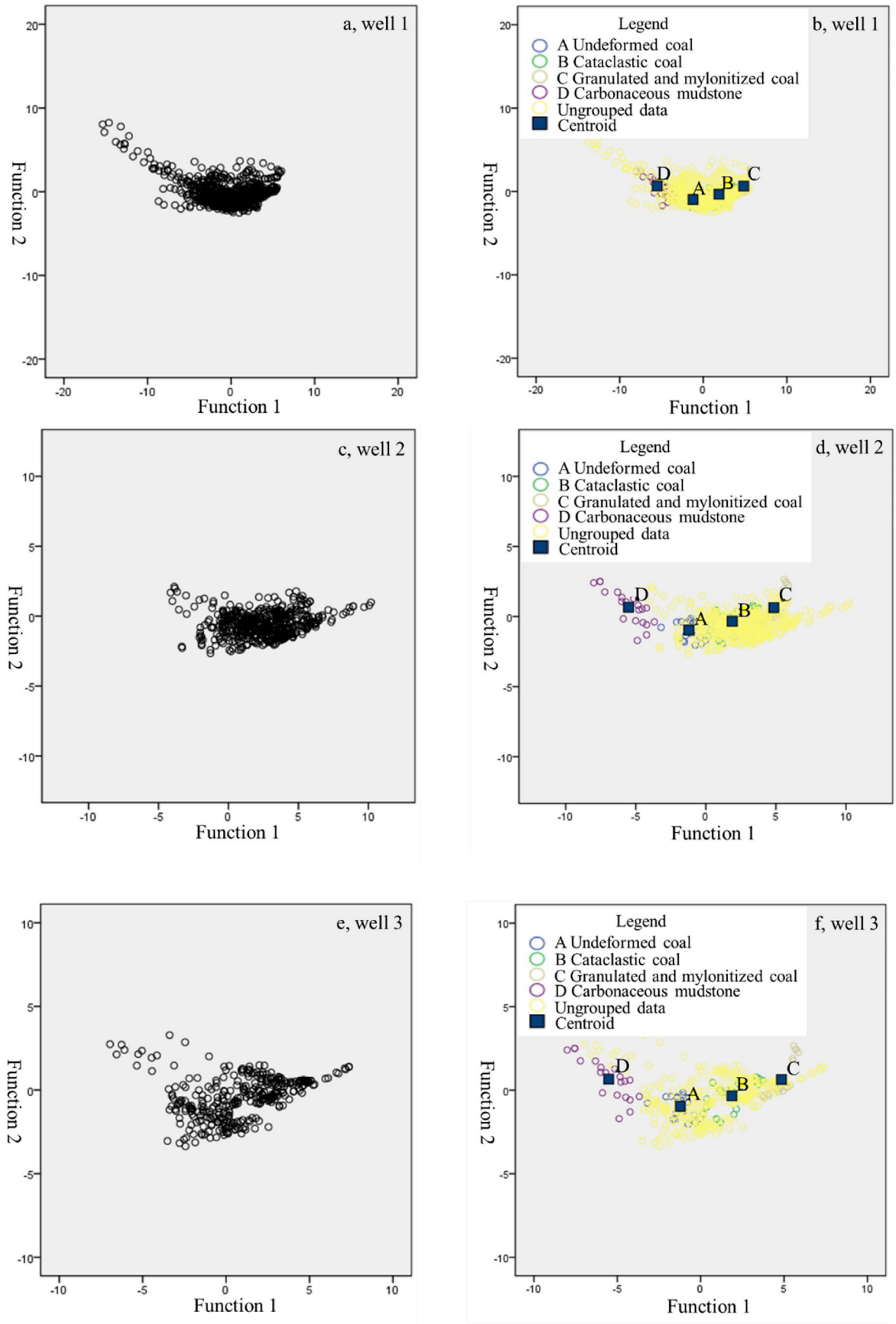

Fig. 10. Projection distribution diagram and discriminant analysis classification diagram of ungrouped samples of three Wells. The black circles in figures a, c, and e correspond to the yellow circles in figures b, d and f. The black circles in figures a, c and e correspond the projections of the test sample set data (ungrouped data) in Well 1, Well 2 and Well 3, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

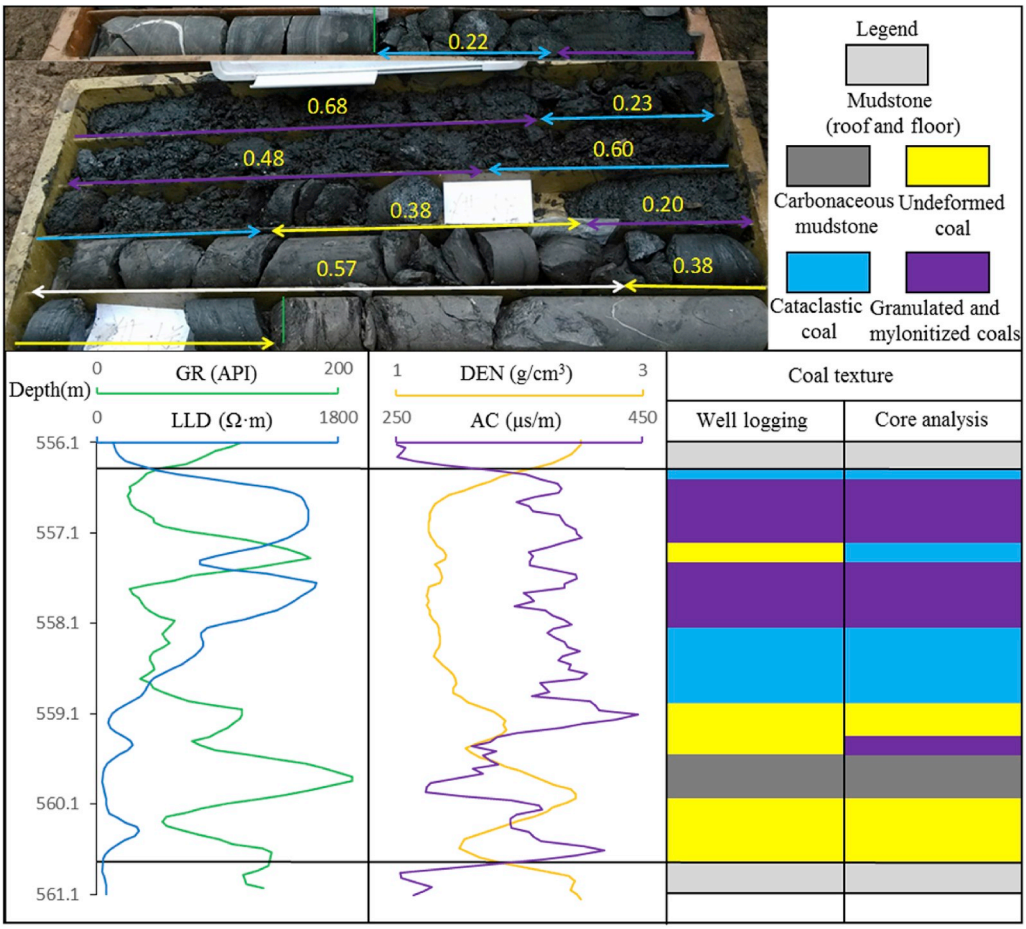

Fig. 11. Comparison of core identification and well logging identification of No. 3 coal seam (4.22 m) in Well 1 for coal texture.

$$F_2 = 0.028*\mathrm{GR} + 0.002*\mathrm{LLD} - 1.467*\mathrm{DEN} + 0.009*\mathrm{AC} - 4.377$$

Function 3:

$$F_3 = -0.001*\mathrm{GR} + 0.002*\mathrm{LLD} + 1.026*\mathrm{DEN} - 0.008*\mathrm{AC} + 0.620$$

All four groups of sample points are distributed around the group centroids in Fig. 8, where the centroid coordinates are A (−1.214, −0.967), B (1.879, 0.335), C (4.849, 0.640) and D (−5.514, 0.661), respectively. They are clearly distinguished in the coordinate system identified by F1 and F2. The area surrounded by numbers and boundaries is divided into four regions in Fig. 9. As for the unclassified samples, the classification of each sample point can be accurately determined by calculating the discriminant score of F1 and F2.

As shown in Table 4, the discrimination results of all undeformed coals and carbonaceous mudstones are consistent with actual coal cores in Well 1, Panguan syncline. Among 20 cataclastic coal samples, 16 of them are accurately discriminated, but three of them are classified into undeformed coal, and one is classified into granulated and mylonitized coal. Meanwhile, Among 20 granulated and mylonitized coals samples, 18 of them are correct, but two of them are classified into cataclastic coal. Thus, 76 samples are accurately identified based the built method and the classification accuracy is 92.5%. In order to further verify the accuracy of the classification results from training set, the classification results are cross-validated (Table 5), where each case is classified by a function derived from other cases. It is found that 90% of the cases are classified correctly. For the part of the data that is different from the training set classification, it is mainly because the cataclastic coal belongs to the transition between the undeformed coal and the granulated coal. When the degree of destruction is slightly weak, the coal structure is relatively complete, which can be easily identified as undeformed coal by visual observation. Conversely, cataclastic coal may be recognized as granulated coal when the degree of destruction is slightly strong. The reason for the error of the granulated and mylonitized coals is the same. Overall, the discrimination results have high accuracy, indicating that the discriminant analysis results can be used for prediction.

4.3. Validation with core and well logging comparisons

Based on the training sample set data, the method is applied to the fully cored Well 1 in the Panguan syncline. After the projection of logging data of all coal seams, the discriminant analysis was carried out according to the discriminant criterion. It can be seen that the grouped data is distributed around the four groups of centroids. The aggregation of the same category of samples is strong, which is helpful for the identification of coal texture (Jemwa and Aldrich, 2012). By substituting each point data into the above two discriminant functions, the obtained coordinates can be distinguished from the classification (Fig. 10a and b). The discriminant analysis of Well 2 and Well 3 in the Panguan syncline were carried out in the same way (Fig. 10c, d, e, f). In general, the data points have a good classification effect in the figure, and are divided into four categories in the space.

Compared with the core identification results of No. 3 coal seam in Well 1 and No. 12 coal seam in Well 2, the logging identification results have good consistency, and the transition boundaries of coal texture are divided clearly. The upper part of No. 3 coal seam mainly develops tectonic coal, and the bottom is almost undeformed coal. Carbonaceous mudstone with a thickness of about 0.5 m is sandwiched in the undeformed coal (Fig. 11). The No. 12 coal seam is greatly affected by in-situ stress, with the coal texture of mainly cataclastic coal, and the

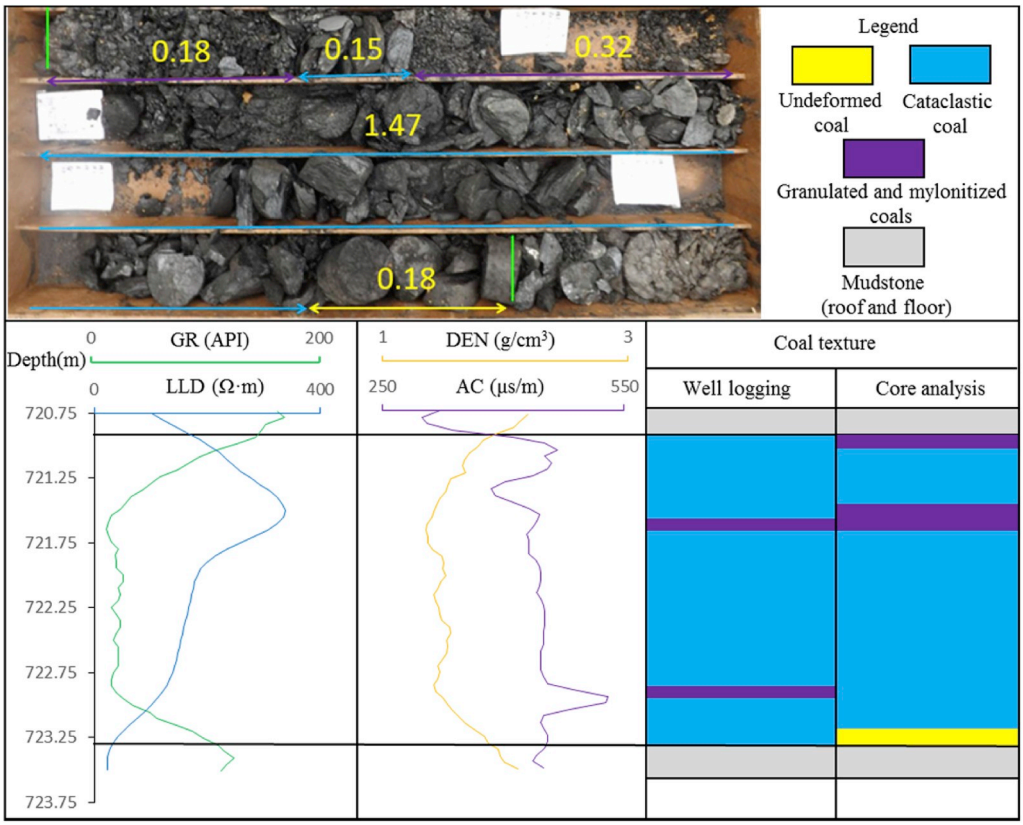

Fig. 12. Comparison of core identification and well logging identification of No. 12 coal seam (2.36 m) in Well 2 for coal texture.

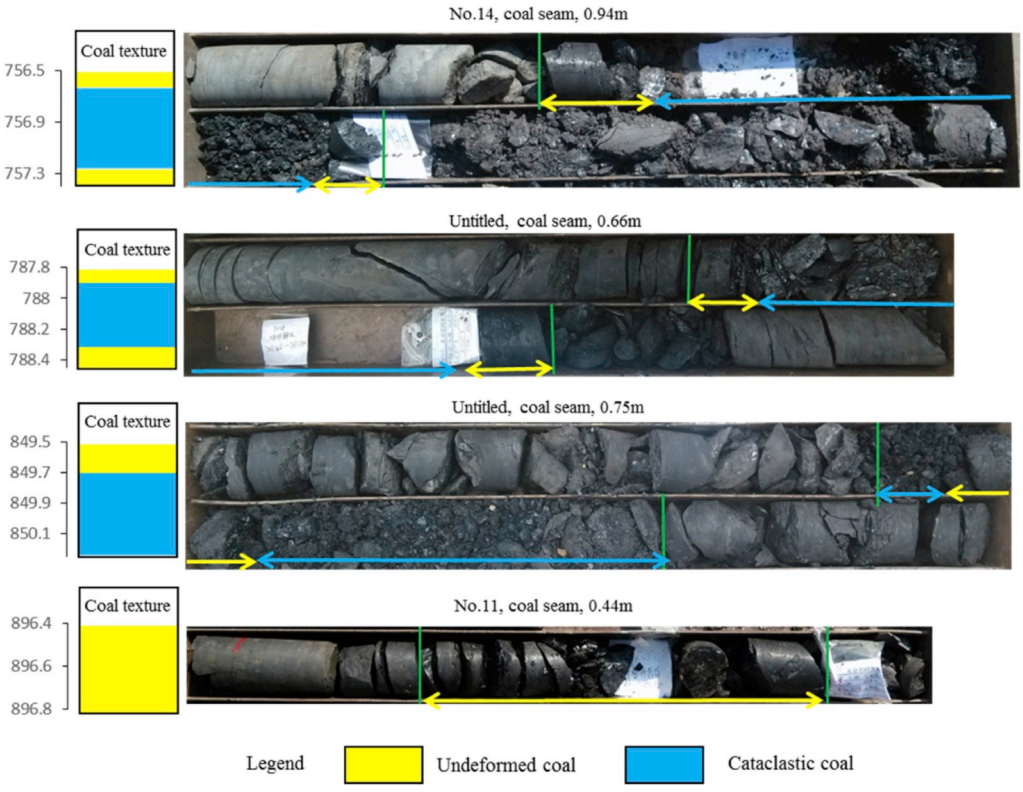

Fig. 13. Well logging identification of multiple and thin coal seams (less than 1 m) in Well 3.

Table 6
The situation of coal texture in wells 1, 2 and 3.

Well	Predicted Group Member				Total
	A	B	C	D	
Well 1	265	230	70	105	670
Well 2	70	310	121	7	508
Well 3	132	136	48	14	330

development of granulated and mylonitized coals is mainly controlled by the strong structural deformation (Fig. 12). As can be seen in Fig. 13, lots of coal seams in Well 3 develop undeformed and cataclatic coals, though there are some errors, most of them are consistent with core observations.

Table 6 shows the classification of 1508 data points from three wells in the Panguan syncline after discriminant analysis. All the logging data and classification results of coal texture are checked by coal core description and comparison, and accurate inversion of coal texture is realized. Among them, Well 1 mainly develops undeformed and cataclastic coals and the thickness of them is about 24.75 m, with multiple layers of carbonaceous mudstone interposed. Well 2 mainly develops cataclastic coals with the total thickness of about 15.5 m, followed by the undeformed coal, but carbonaceous mudstone is rarely to be seen. The undeformed and cataclastic coals are more common than granulated and mylonitized coals in Well 3, and the thickness of undeformed and cataclastic coals is about 13.4 m. The tectonic coal (mainly cataclastic coal) is mainly distributed in the Nos. 3, 7, 8, 9, 10, 12, 17, 18_1 coal seams of Well 1, in the Nos. 10, 12, 18_1 coal seams of Well 2, and in the Nos. 12, 14, 15, 18_1 coal seams of Well 3. Most of them are the stable and minable coal seam in the Panguan syncline.

In conclusion, LDA is suitable for inversion of multiple and thin coal seams in the Panguan syncline. At the same time, it is also suitable for thick coal seams and can accurately identify the carbonaceous mudstone in this area. Compared with other inversion methods, LDA has certain advantages: (1) It retains as much information as possible of the data sample and has no special requirements for the overall distribution. (2) Based on the determination of Mahalanobis distance, the LDA theory can accurately determine the boundary of different coal textures and reduce the occurrence of artificial errors. (3) LDA can be used not only for dimensionality reduction, but also for classification, which selects the projection direction with the best classification performance. (4) Different from the unsupervised dimensionality reduction methods such as cluster analysis and principal component analysis, LDA can use the prior knowledge and experience of categories in the dimensionality reduction process. (5) Compared with some previous empirical formulas, LDA assigns weights to different logging curves according to the actual response characteristics of different logging parameters in the study area.

5. Conclusions

To identify different thickness of coal texture, especially the multiple and thin coal seam in the Panguan syncline, four types of logging parameters are reasonably applied, and the identification method of coal texture based on LDA is established. The identification method can take full advantage of logging data and improve accuracy as far as possible. The method generally shows a good consistency, and the agreement is acceptable with limited error. There are some possible reasons, one is the coal may be damaged while drilling, so we might make the mistake of identifying the destruction degree of coal core by visual observation. In addition, the changes of in situ stress, geotemperature field, and pressure field and gas-water interaction may inevitably affect the accuracy of logging data.

By comparison with the core description and drilling log, this method is not only applicable to the thick coal seam, but also has high accuracy for the identification of multiple and thin coal seam in the Panguan syncline. Moreover, it can effectively distinguish the carbonaceous mudstone sandwiched in the coal seam. Tectonic coals (cataclastic coal, granulated and mylonitized coal) are mainly distributed in the Nos. 10, 12, and 18_1 coal seams in the Panguan syncline, and the tectonic coals are more developed within the thicker coal seam.

Acknowledgements

This work is supported by the National Natural Science Foundation of China (41772132, 41502157, 41530314), the National Major Science and Technology Project of China (2016ZX05044-001), and the Geological Survey Projects of China Geological Survey (121201021000150014, 12120115008201).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.petrol.2019.106616.

References

Abuzeina, D., Al-Anzi, F.S., 2018. Employing Fisher discriminant analysis for Arabic text classification. Comput. Electr. Eng. 66, 474–486.
Cai, W., Guan, G.Y., Pan, R., Zhu, X.M., Wang, H.S., 2018. Network linear discriminant analysis. Comput. Stat. Data Anal. 117, 32–44.
Chen, S.D., Tang, D.Z., Tao, S., Xu, H., Li, S., Zhao, J.L., Ren, P.F., Fu, H.J., 2017. In-situ stress measurements and stress distribution characteristics of coal reservoirs in major coalfields in China: implication for coalbed methane (CBM) development. Int. J. Coal Geol. 182, 66–84.
Chen, S.D., Tang, D.Z., Tao, S., Xu, H., Zhao, J.L., Fu, H.J., Ren, P.F., 2018. In-situ stress, stress-dependent permeability, pore pressure and gas-bearing system in multiple coal seams in the Panguan area, western Guizhou, China. J. Nat. Gas Sci. Eng. 230, 258–265.
Connell, L.D., Lu, M., Pan, Z.J., 2010. An analytical coal permeability model for tri-axial strain and stress conditions. Int. J. Coal Geol. 84, 103–114.
Fu, X.H., Qin, Y., Wang, G.G.X., Rudolph, V., 2009. Evaluation of gas content of coalbed methane reservoirs with the aid of geophysical logging technology. Fuel 88, 2269–2277.
Fu, X.H., Qin, Y., Wang, G.G.X., Rudolph, V., 2009. Evaluation of coal structure and permeability with the aid of geophysical logging technology. Fuel 88, 2278–2285.
Gao, H.X., 2004. Application of Multivariate Statistical Analysis. Peking University Press.
Guo, Z.G., 1999. Methods of Social Statistical Analysis- the Application of SPSS. China Renmin University Press.
Hou, H.H., Shao, L.Y., Guo, S.Q., Li, Z., Zhang, Z.J., Yao, M.L., Zhao, S., Yan, C.Z., 2017. Evaluation and genetic analysis of coal structures in deep Jiaozuo Coalfield, northern China: investigation by geophysical logging data. Fuel 209, 552–566.
Jemwa, G.T., Aldrich, C., 2012. Estimating size fraction categories of coal particles on conveyor belts using image texture modeling methods. Expert Syst. Appl. 39, 7947–7960.
Ju, Y.W., Li, X.S., 2009. New research progress on the ultrastructure of tectonically deformed coals. Prog. Nat. Sci. 19 (11), 1455–1466.
Li, C.C., Liu, D.M., Cai, Y.D., Yao, Y.B., 2016. Fracture permeability evaluation of a coal reservoir using geophysical logging: a case study in the Zhengzhuang area, southern Qinshui Basin. Energy Explor. Exploit. 34 (3), 378–399.
Luo, C.J., Zhang, D.F., Lun, Z.M., Zhao, C.P., Wang, H.T., Pan, Z.J., Li, Y.H., Zhang, J., Jia, S.Q., 2019. Displacement behaviors of adsorbed coalbed methane on coals by injection of SO_2/CO_2 binary mixture. Fuel 247, 356–367.
Min, K.B., Rutqvist, J., Tsang, C.F., Jing, L.R., 2004. Stress-dependent permeability of fractured rock masses: a numerical study. Int. J. Rock Mech. Min. Sci. 41 (7), 1191–1210.
Nie, Norman, H., 1975. SPSS: Statistical Package for the Social Sciences. Mcgraw-Hill Book Company, 1975.
Qin, Y., Tang, D.Z., Liu, D., 2014. Geological evaluation theory and technology progress of coal reservoir dynamics during coalbed methane drainage. Coal Sci. Technol. 41, 80–88 (in Chinese with an English abstract).
Rai, D.K., Roy, S., Roy, A.L., 2004. Evaluation of Coal Bed Methane through Wire Line Logs Jharia Field: A Case Study. 5th Conference & Exposition on Petroleum Geophysics, Hyderabad, India, pp. 910–914.
Ren, P.F., Xu, H., Tang, D.Z., Li, Y.K., Sun, C.H., Tao, S., Li, S., Xin, F.D., Cao, L.K., 2018. The identification of coal texture in different rank coal reservoirs by using geophysical logging data in northwest Guizhou, China: investigation by principal component analysis. Fuel 230, 258–265.
Rutqvist, J., Stephansson, O., 2003. The role of hydromechanical coupling in fractured rock engineering. Hydrogeol. J. 11 (1), 7–40.
Shen, Y.L., Qin, Y., Guo, Y.H., Yi, T.S., 2016. Yuan XX, Shao YB. Characteristics and sedimentary control of a coalbed methane-bearing system in Lopingian (late Permian) coalbearing strata of western Guizhou province. J. Nat. Gas Sci. Eng. 33, 8–17.

Siregar, I., Niu, Y.F., Mostaghimi, P., Armstrong, R.T., 2017. Coal ash content estimation using fuzzy curves and ensemble neural networks for well log analysis. Int. J. Coal Geol. 181, 11–22.

Su, J.M., Fu, R.H., Zhou, J.B., Zhang, L.H., 2000. SPSS for Windows. Publishing House of Electronics Industry.

Tao, S., Chen, S.D., Pan, Z.J., 2019. Current status, challenges, and policy suggestions for coalbed methane industry development in China: a review. Energy Sci. Eng. 00, 1–16.

Tao, S., Chen, S.D., Tang, D.Z., Zhao, X., Xu, H., Li, S., 2018. Material composition, pore structure and adsorption capacity of low-rank coals around the first coalification jump: a case of eastern Junggar Basin, China. Fuel 211, 804–815.

Tao, S., Pan, Z.J., Chen, S.D., Tang, S.L., 2019. Coal seam porosity and fracture heterogeneity of marcolithotypes in the Fanzhuang Block, southern Qinshui Basin, China. J. Nat. Gas Sci. Eng. 66, 148–158.

Tao, S., Pan, Z.J., Tang, S.L., Chen, S.D., 2019. Current status and geological conditions for the applicability of CBM drilling technologies in China: a review. Int. J. Coal Geol. 202, 95–108.

Tao, S., Tang, D.Z., Xu, H., Gao, L.J., Fang, Y., 2014. Factors controlling high-yield coalbed methane vertical wells in the Fanzhuang Block, Southern Qinshui Basin. Int. J. Coal Geol. 134, 38–45.

Tao, S., Wang, Y.B., Tang, D.Z., Xu, H., Lv, Y.M., He, W., Li, Y., 2012. Dynamic variation effects of coal permeability during the coalbed methane development process in the Qinshui Basin, China. Int. J. Coal Geol. 93, 16–22.

Tao, S., Zhao, X., Tang, D.Z., Deng, C.M., Meng, Q., Cui, Y., 2018. A model for characterizing the continuous distribution of gas storing space in low-rank coals. Fuel 233, 552–557.

Teng, J., Yao, Y.B., Liu, D.M., Cai, Y.D., 2015. Evaluation of coal texture distributions in the southern Qinshui basin, North China: investigation by a multiple geophysical logging method. Int. J. Mining Sci. Technol. 140, 9–22.

Wang, Y.J., Liu, D.M., Cai, Y.D., Yao, Y.B., Zhou, Y.F., 2018. Evaluation of structured coal evolution and distribution by geophysical logging methods in the Gujiao Block, northwest Qinshui basin, China. J. Nat. Gas Sci. Eng. 51, 210–222.

Wen, J., Fang, X.Z., Cui, J.R., Fei, L.K., Yan, K., Chen, Y., Xu, Y., 2018. Robust sparse linear discriminant analysis. IEEE Trans. Circuits Syst. Video Technol. 1-1.

Xu, H., Tang, D.Z., Mathews, J.P., Zhao, J.L., Li, B.Y., Tao, S., Li, S., 2016. Evaluation of coal macrolithotypes distribution by geophysical logging data in the hancheng block, eastern margin, Ordos basin, China. Int. J. Coal Geol. 165, 265–277.

Xue, G.W., Liu, H.F., Li, W., 2012. Deformed coal types and pore characteristics in Hancheng coalmines in Eastern Weibei coalfields. Int. J. Mining Sci. Technol. 22 (5), 681–686.

Yao, J.P., Sima, L.Q., Zhang, Y.G., 2011. Quantitative identification of deformed coals by geophysical logging. J. China Coal Soc. 36 (1), 94–98 (in Chinese with English abstract).

Ye, H.S., Li, Y.J., Chen, C., Zhang, Z.H., 2017. Fast Fisher discriminant analysis with randomized algorithms. Pattern Recognit. 72, 82–92.

Yegireddi, S., Bhaskar, G.U., 2009. Identification of coal seam strata from geophysical logs of borehole using Adaptive Neuro-Fuzzy Inference System. J. Appl. Geophys. 67 (1), 9–13.

Zhou, B.Z., O'Brien, G., 2016. Improving coal quality estimation through multiple geophysical log analysis. Int. J. Coal Geol. 167, 75–92.

Zhou, J., Zhang, L., Li, X., Pan, Z., 2019. Experimental and modeling study of the stress-dependent permeability of a single fracture in shale under high effective stress. Fuel 257, 116078.