

# The classification in metamorphic rocks using modified fuzzy cluster analysis from geophysical log data: evidence from Chinese Continental Scientific Drilling Main Hole

Huaijie Yang<sup>1</sup> · Heping Pan<sup>1</sup> · Miao Luo<sup>1</sup> · Gang Li<sup>1</sup> · Jing Yao<sup>1</sup>

Received: 23 October 2014 / Accepted: 5 May 2015 / Published online: 28 May 2015  
© The Author(s) 2015. This article is published with open access at Springerlink.com

**Abstract** Lithology is one of the most important data in evaluating reservoir, and is mainly carried out by cores recovery in laboratory which is very expensive, and its interpretation is time consuming. Accurate identification of lithology is fundamentally crucial to evaluate reservoir from geophysical log data. Pattern recognition and statistical analysis have been proved to be the most powerful methods for constructing optimal model in lithology recognition. To address this issue, a fast and practical K-means clustering algorithm is proposed in order to better deal with lithology recognition from geophysical log data. Based on the traditional K-means clustering algorithm, Euclidean distance is replaced by Mahalanobis distance; the initial cluster centers are acquired from the average of characteristic values but not selected randomly, in addition, adding weight value in each characteristic value of the objective function, and thus a lithology recognition model named modified K-means clustering is established. The method is applied to identify the Chinese Continental Scientific Drilling Main Hole (CCSD-MH) metamorphic rocks. Compared with the traditional K-means clustering, the accurate rate of the modified K-means clustering in lithologic identification has improved for the same 45 samples, raised 11.11 %. According to the modified K-means cluster algorithm, nine kinds of lithology cluster centers are acquired from 45 samples. The classes of the samples can be determined by analyzing the hamming

approach degree curves, which is calculated by the undetermined samples and 9 cluster centers. The predicted results and the core recovery are exactly the same by comparison. The hamming approach degree can identify the whole well of CCSD-MH lithology effectively and accurately. This model may be made applications to other areas.

**Keywords** Lithology recognition · CCSD-MH · K-means clustering · Hamming approach degree · Cluster center · Geophysical log data · Weight value

## Introduction

Fuzzy theory was proposed by cybernetic professor L. A. Zadeh in University of California in 1965 (Gao 2004) and has been widely used in the natural sciences and social sciences fields in the following 50 years. Fuzzy clustering analysis is a branch of fuzzy mathematics, and its range of applications involves time series prediction (Ryoke et al. 1995), neural networks training (Karayiannis and Mi 1997), nonlinear system identification (Runkler et al. 1996), parameter estimation (Gath and Geva 1989), medical diagnosis (Bezdek and Fordon 1979), weather forecast (Newton 1992), food classification (Windham 1985), and water quality analysis (Mukherjee 1995).

The limitations of traditional fuzzy clustering analysis are several controlling factors, such as the choice of initial cluster centers, the correlation between samples, the trade-off between iteration times, and solutions accuracy. To solve these problems, many researchers had proposed many modified algorithms, such as K-means clustering, C-means clustering, fuzzy clustering neural network, and fuzzy clustering genetic.

✉ Heping Pan  
panpinge@163.com  
Huaijie Yang  
yhj870624@126.com

<sup>1</sup> Institute of Geophysics and Geomatic, China University of Geosciences, Wuhan 430074, Hubei Province, China

Overview of worldwide, the workers had made many researches about the lithology identification of CCSD-MH; however, the database of CCSD-MH core data was still incomplete and inaccurate. Xu et al. (2006) analyzed magnetic susceptibility and density of different rocks from CCSD-MH in the depth Section 0–2000 miles, identified the lithology with SPSS statistical software. Jing et al. (2007) summed up 11 kinds of eclogites into 6 kinds based on multivariate statistic methods. Gu et al. (2009) constructed the lithology recognition model combining the logging response and several well logs of different rocks with the method of cluster analysis and stepwise discriminant analysis. Luo and Pan (2010) used core-log correlation and cross-plotting methods, and the results allowed the authors to conclude that the lithology is mainly comprised orthogneiss, paragneiss, eclogite, amphibolite, and ultramafic rocks. Bosch et al. (2013) used fuzzy logic for lithology prediction from well log data of the German Continental Deep Drilling Program (KTB). Results showed that this fuzzy logic-based method was suited for rapidly and reasonably suggesting a lithology column from KTB well log data. The above authors heavily focused on approaches such as visual inspection, cross-plotting technology, and discriminate function analysis, and not formed a method that can neatly identify the main units and refine the classification of the CCSD-MH whole well.

Reservoir evaluation needs the data of many kinds of rocks, which have a much more different porosity and permeability. The well logs have varieties of responses based on different kinds of rocks' characteristic. And the lithology data is mainly carried out by cores recovery in lab which is very expensive and its interpretation is time consuming, so accurate identification of lithology from geophysical well log data plays a significant role in reservoir evaluation.

In this study, a fast and practical K-means clustering algorithm was proposed in order to better deal with lithology recognition of CCSD-MH from geophysical log data. Based on the traditional K-means clustering algorithm, Euclidean distance was replaced by Mahalanobis distance, and the initial cluster centers were acquired from the average of characteristic values, in addition, added weight value in each characteristic value of the objective function. The model was applied to classify CCSD-MH metamorphic rocks and get the cluster centers of each class. The cluster centers, as well as weight values, were used to calculate the hamming approach degree, which can neatly identify the main units and refine the classification of the CCSD-MH whole well.

## Modified K-means clustering

### Cluster center

Let us choose  $m$  objects, and each object has  $n$  characteristic values that may be classified into  $z$  classes. According to the fuzzy theory, the fuzzy matrix involving the above objects can be constructed as  $X = [x_{ij}]$ , ( $i = 1, 2, \dots, m$ ;  $j = 1, 2, \dots, n$ ). The cluster center matrix is defined as  $C = [c_{kj}]$ , and  $k$  ( $k = 1, 2, \dots, z$ ),  $j$  ( $j = 1, 2, \dots, n$ ). And  $m \geq z$ . The two matrixes and their compositions are as follows:

$$X = \{X_1, X_2, \dots, X_m\} \text{ and } X_i = \{x_{i1}, x_{i2}, \dots, x_{in}\} \quad (1)$$

$$C = \{C_1, C_2, \dots, C_z\} \text{ and } C_k = \{c_{k1}, c_{k2}, \dots, c_{kn}\}. \quad (2)$$

The traditional K-means algorithm will increase iteration times if initial cluster centers selected inappropriately, and may easily fall into local optimums. In order to alleviate this problem, we acquired the initial cluster centers from the average of characteristics values in matrix  $X$  based on the theory of cluster center (Liao 2013).

Therefore, any elements  $C_k = \{c_{k1}, c_{k2}, \dots, c_{kn}\}$  of matrix  $C$  can be defined as follows:

$$c_{k1} = \frac{k}{z} \frac{1}{n} \sum_{j=1}^n x_{j1} \quad c_{k2} = \frac{k}{z} \frac{1}{n} \sum_{j=1}^n x_{j2}, \dots, c_{kn} = \frac{k}{z} \frac{1}{n} \sum_{j=1}^n x_{jn}. \quad (3)$$

### The weight value

Traditional K-clustering takes the same contribution for cluster results by different characteristic values, which will be susceptible to noise data resulting in the inaccuracy of cluster result. However, the contribution of each characteristic value to the cluster result is different. Therefore, a weighted value scheme is proposed to deal with the relationship between various samples data and cluster centers in the objective function. The weight value of each characteristic in fuzzy matrix  $X$  can be obtained as follows:

Firstly, we introduce a standard to evaluate the comparison of every two groups of characteristic values:

$$b_{ij} = \frac{x_{ij} - x_{\min}}{x_{\max} - x_{\min}}, \quad (4)$$

where  $x_{\min}$  and  $x_{\max}$  refer to the maximum and minimum of the given each group of characteristic values.

Secondly, the weight  $w_j$  is defined as information entropy for the  $j$ th group characteristic value, which can be evaluated by (Feng et al. 2010; Zhang et al. 2010)

$$H_j = -\frac{1}{\ln m} \left( \sum_{i=1}^m f_{ij} \ln f_{ij} \right), \tag{5}$$

where  $f_{ij}$  is the weight of the  $j$ th index for the  $i$ th sample in all samples which can be estimated using normalized judgment matrix elements  $b_{ij}$ , giving (Zhang et al. 2010)

$$f_{ij} = \frac{1 + b_{ij}}{\sum_{i=1}^m (1 + b_{ij})}. \tag{6}$$

Finally, the weight  $w_j$  can evaluate the contribution of each characteristic value to the lithology classification results, which is given by (Feng et al. 2010; Zhang et al. 2010)

$$w_j = \frac{1 - H_j}{n - \sum_{j=1}^n H_j}, \text{ and } \sum_{j=1}^n w_j = 1. \tag{7}$$

### Objective function

Based on the traditional K-means clustering algorithm, Euclidean distance is replaced by Mahalanobis distance, in addition, added different weight values in each characteristic value. The weight is given a higher value if the contribution of its corresponding characteristic value has quite a higher influence to the lithology recognition, while the redundant characteristic values are given much smaller values. Therefore, the objective function of modified K-means cluster algorithm can be defined as

$$F(X, C) = \min \left( \sum_{k=1}^z \sum_{i=1}^m \sqrt{w_1(x_{i1} - c_{k1}) \sum_{i=1}^{-1} (x_{i1} - c_{k1})^T + \dots + w_n(x_{in} - c_{kn}) \sum_{i=1}^{-1} (x_{in} - c_{kn})^T} \right), \tag{8}$$

where  $\sum$  is the covariance matrix and  $w_j$  is the weight of each characteristic value.

If the  $\sum$  was a singular matrix,  $\sum$  can be written as the product of three pieces,  $\sum = A^T G A$ . The pseudo-inverse of  $\sum$  can be solved through the inverse of  $G$  (Wu et al. 2011). Uniquely to Euclidean distance, Mahalanobis distance takes the correlation of every two groups of characteristic values into account and solves the repetitive information reused that leads to the clustering results inaccuracy. The weigh value of each characteristic value is taken as the parameter for reflect the different contributions of every group of characteristic value to the clustering results, and thus the calculated clustering center can be much closer to the actual results.

The modified K-means clustering follows

- (1) Given the maximum iteration times  $T$  or threshold value  $\varepsilon$ , and the cluster center number  $z$ . The initial iterative value  $t = 1$ .
- (2) Calculating the cluster centers and weight values with Eqs. (3) and (7), respectively, and the covariance matrix.
- (3) Calculating the distances between the undetermined samples and every cluster centers with Eq. (8), and assigning the sample to the class, whose cluster center is the nearest to this sample.
- (4) Calculating the new cluster centers of every new class.
- (5) According to Eq. (8), calculating the objective function  $J(X, C)$ . It will not stop operating until  $|J(X, C)^{(t)} - J(X, C)^{(t-1)}| \leq \varepsilon$  or  $T \leq t$ . If so, output the cluster centers of each class; otherwise, make  $t = t + 1$ , backward to step 3, and go on iterating.

### Method application

The drill site of the CCSD is located near Maobei village (N34°25', E118°40'), about 17 km southwest of Donghai in the southern segment of the Sulu UHP terrane (Wang et al. 2013), as shown in Fig. 1. Major goals of the CCSD, as outlined by Xu et al. (1998), include (1) to reveal the crustal structure of convergent plate boundaries, (2) to

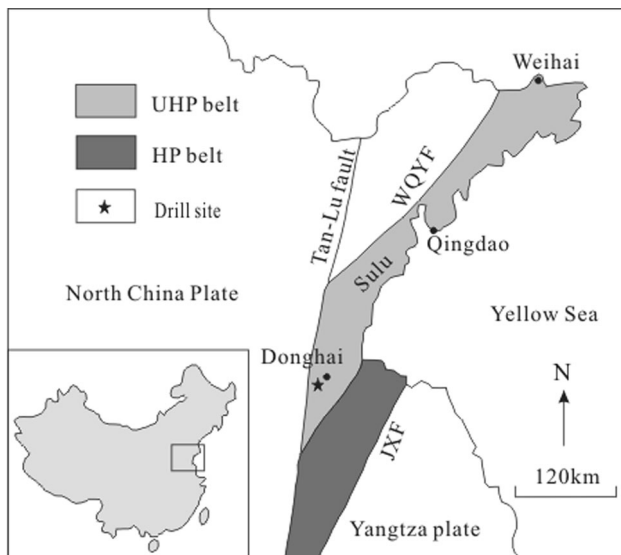
provide constraints on crust-mantle interactions and mantle behavior during deep subduction of continental crust, and (3) to investigate fluid evolution during UHP metamorphism. The CCSD-MH was completed at its final depth of 5158 m in March 2005.

The lithology types have been confirmed by core recovery of 112 drill hole samples in laboratory from 220 to 4000 miles, are mainly eclogite, paragneiss, orthogneiss, amphibolites, serpentinite, and granite (Zhang et al. 2000, 2003; Ya et al. 2000; Liu et al. 2004, 2005).

In this paper, we chose several well logs and core recovery materials of CCSD-MH from the depth Section 100–2000 miles for method development, which will identify the lithology of the whole well from the depth

100–5118.2 miles. In order to demonstrate the method capacity in identifying lithology, we selected 9 kinds of rocks that almost distribute in the whole well based on the log representations and core recovery materials. There are seven well logs that are selected by analyzing the corresponding relationship between the logs and lithology; they are deep induction resistivity (RILD), compensated neutron (CNL), interval transit time (AC), compensated density (DEN), natural gamma (GR), uranium (U), and thorium (TH).

The fuzzy matrix mentioned above contains 45 samples, 9 kinds of lithology, and each lithology has five samples. We identified the lithologies both using the traditional K-means cluster and modified K-means algorithms, and output the cluster centers of each lithology (Table 1 and Table 2).



**Fig. 1** Simplified geological location of CCSD-MH by a star symbol (Ni et al. 2013; Zhang et al. 2006)

The parameter setting of traditional K-means cluster algorithm is as follows: input 45 samples, threshold  $\varepsilon = 0.001$ , and cluster numbers  $z = 9$ . The initial cluster centers are selected randomly and were the former  $z$  samples. The classification accuracy of 45 samples was 77.78 %. Output the cluster centers of each lithology (Table 1).

The parameter setting of modified K-means algorithm is as follows: input 45 samples, threshold  $\varepsilon = 0.001$ , and cluster numbers  $z = 9$ . In order to make a comparison, the modified K-means cluster algorithm tested the same samples with the traditional K-means cluster algorithm, and the classification accuracy was 88.89 %. Compared with the traditional K-means clustering, the accurate rate of the modified K-means clustering in lithologic identification has improved, raised 11.11 %. Output the cluster centers of each lithology (Table 2).

## Results and discussion

### Verification of method prediction

According to the modified K-means cluster algorithm, 9 cluster centers were acquired from the 45 samples. The information entropy used to evaluate the contribution of each well logs on identifying the CCSD-MH lithology was calculated with Eq. (7). After that, calculating the hamming approach degree between the undetermined samples and 9 cluster centers, which is to determine the class of the samples. The computation of the hamming approach degree is as follows (Zhang et al. 2010):

$$H = 1 - \sum_{j=1}^n w_j |x_j - c_k| \quad k = 1, 2, \dots, z, \quad (9)$$

**Table 1** Cluster centers of each lithology using traditional K-means cluster algorithm

No	RILD ( $\Omega$ m)	CNL (%)	AC ( $\mu$ s/ft)	DEN ( $\text{g}/\text{cm}^3$ )	GR (API)	U ( $10^{-6}$ )	TH ( $10^{-6}$ )
1	1069.26	12.06	46.17	3.10	39.55	0.91	2.22
2	247.94	10.70	51.13	3.03	46.67	0.76	3.30
3	2754.18	1.00	54.79	2.75	150.22	2.94	18.92
4	1531.05	-0.61	52.54	2.59	123.99	0.90	7.92
5	5081.61	-0.61	52.46	2.59	216.95	5.91	27.25
6	360.71	9.74	47.95	2.98	47.70	1.26	3.55
7	765.36	42.62	56.47	2.66	12.35	0.52	1.90
8	3936.09	-0.71	52.91	2.61	215.64	7.68	20.64
9	437.71	11.16	53.00	2.92	50.67	0.73	4.27

1 Rutile eclogite, 2 phengite eclogite, 3 retrograded eclogite, 4 paragneiss, 5 orthogneiss, 6 amphibolites, 7 serpentinite, 8 chlorite amphibolites and 9 moyite

**Table 2** Cluster centers of each lithology using modified K-means cluster algorithm

No	RILD ( $\Omega$ m)	CNL (%)	AC ( $\mu\text{s}/\text{ft}$ )	DEN ( $\text{g}/\text{cm}^3$ )	GR (API)	U ( $10^{-6}$ )	TH ( $10^{-6}$ )
1	1022.12	7.95	42.87	3.29	20.45	0.93	1.11
2	233.31	11.00	51.68	3.05	44.57	0.55	3.06
3	2754.18	1.00	54.79	2.75	150.22	2.94	18.92
4	1520.33	-0.53	52.42	2.59	130.95	0.96	7.90
5	5081.61	-0.61	52.46	2.59	216.95	5.91	27.25
6	341.23	9.38	46.87	2.98	47.85	1.62	3.16
7	791.64	43.07	56.54	2.66	13.04	0.53	1.95
8	3936.09	-0.71	52.91	2.61	215.64	7.68	20.64
9	418.46	10.84	52.15	2.94	51.22	0.69	4.56

1 Rutile eclogite, 2 phengite eclogite, 3 retrograded eclogite, 4 paragneiss, 5 orthogneiss, 6 amphibolites, 7 serpentinite, 8 chlorite amphibolites and 9 moyite

where  $H$  was the hamming approach degree of the undetermined samples and one cluster center.  $x_j$  represents the undetermined sample, and  $c_k$  represents the cluster center of one class.  $w_j$  is the weight of  $j$ th well log.

Hamming approach degree is defined to the similar degree between the undetermined samples and standard samples, and the greater its value indicates that the undetermined samples are closer to the standard samples, that is to say, the undetermined samples and the standard samples are owned to the same class. In this paper, 27 undetermined samples are selected in the depth Section 100.0–5118.2 miles (the whole well).

As shown in Fig. 2a, it is the hamming approach degree curve of 27 undetermined samples and cluster center 4 (Table 2); the hamming approach degrees of number 10, number 11, and number 12 undetermined samples are the biggest, 0.984, 0.985, and 0.976, respectively, almost 1. We can conclude that the three samples belong to the class 4 and are paragneiss. Figure 2b shows the curve of 27 undetermined samples and cluster center 7; and number 19, number 20, and number 21 undetermined samples' hamming approach degree are the biggest, 0.981, 0.979 and 0.979, respectively. These three samples belong to the class 7, and are serpentinite.

Through the classification of the former 6 samples, it can be concluded that number 1, number 2, and number 3 are rutile eclogite; number 4, number 5, and number 6 are phengite; number 7, number 8, and number 9 are retrograded eclogite; number 13, number 14, and number 15 are orthogneiss; number 16, number 17, and number 18 are amphibolites; number 22, number 23, and number 24 are chlorite amphibolites; and number 25, number 26, and number 27 are moyite. According to hamming approach degree curves, we determine the lithology of 27 samples, as shown in Table 3. The predicted results and the core recovery are exactly the same by comparison.

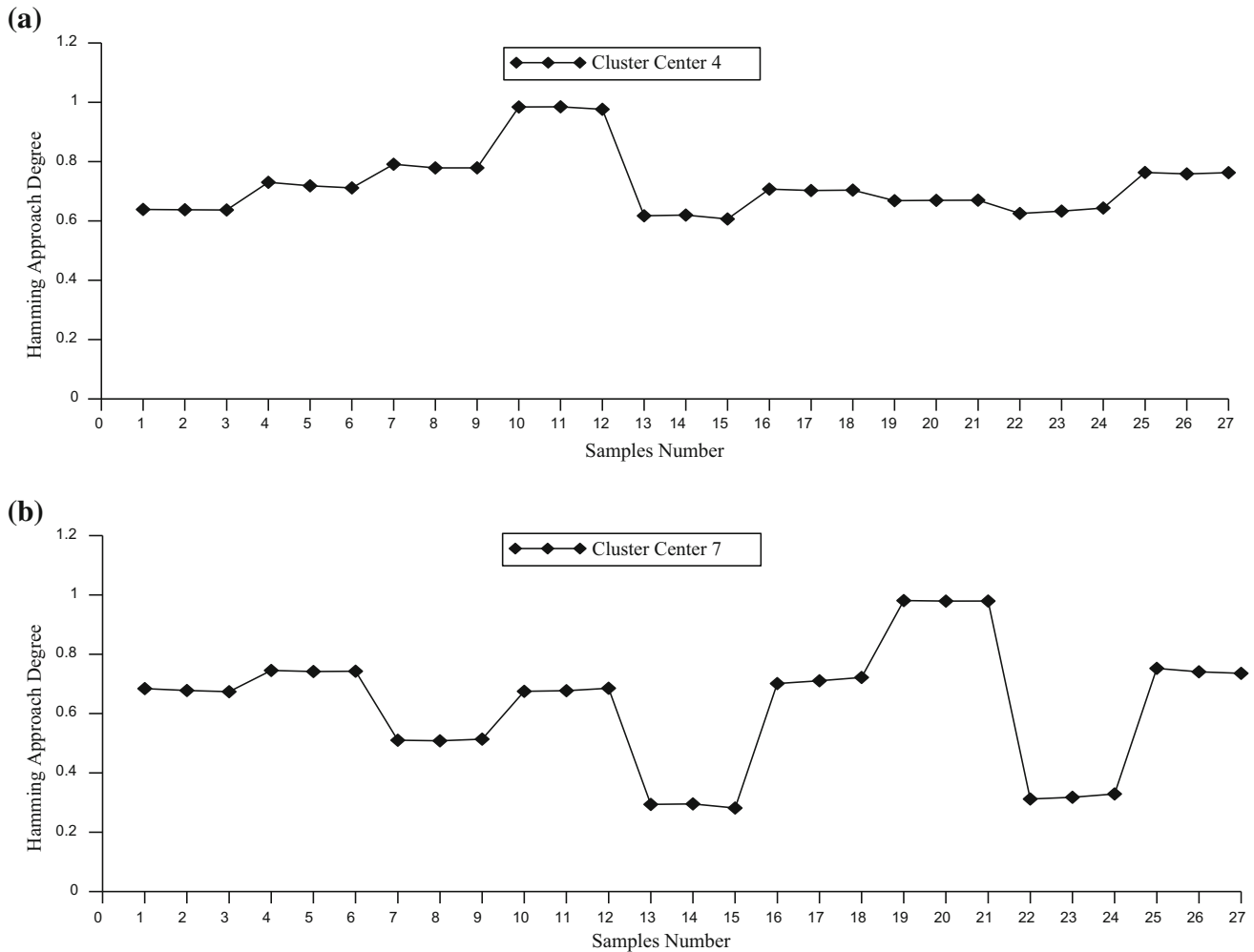
As shown in Fig. 2, the curves of different lithologies have sudden changes; the reason would be both the theoretical model and logging data. In addition, the approach degree of the same lithology also exists slight differences. However, the model could meet the needs of lithology identification for the overall prediction results. Therefore, the cluster centers acquired from the modified K-means clustering algorithm, as well as the hamming approach degree can identify CCSD-MH lithology effectively and accurately, which could also be made applied to other areas.

### Well logging graphic of lithology

The lithology analysis figure resulted from the traditional K-means clustering and modified K-means clustering for visual identification of CCSD-MH lithology, as shown in Figs. 3 and 4. With these two figures, the lithology of CCSD-MH can be easily distinguished.

As shown in Fig. 3, the section of CCSD-MH was divided into 15 parts from the core recovery, while the section could be divided into 11 and 13 parts based on the traditional and modified K-means clustering, respectively.

According to the core recovery, there are five kinds of lithology, and two fracture zones. There are also five kinds of lithology, and no fracture zone displayed based on the identification methods of the traditional K-means cluster and modified K-means cluster. The great error section is from 1600.0 to 1616.7 miles which corresponds to chlorite amphibolites, and there is no fracture zone from the two fuzzy clusters. According to the traditional K-means cluster, there are no chlorite amphibolites between the depth of 1601.52 and 1614.40 miles, and all is phengite eclogite, while the modified K-means cluster has good correspondence with the core recovery and presents the same lithology at the same depth position. At the end depth of the phengite eclogite from the core recovery, there is a short section of orthogneiss, while the corresponding



**Fig. 2** Hamming approach degree curve of 27 samples and cluster center 4 and cluster 7

position of the traditional K-means cluster and modified K-means cluster is amphibolites. Several small errors exist between the depth 1615.4 and 1640.7 miles at the transition of two layers from the traditional K-means cluster and the core recovery, while the modified K-means cluster has a good correspondence with the core recovery, except that a fracture zone presents from the core recovery.

Three kinds of lithology and five fracture zones are shown in Fig. 4 from the depth 1700.0 to 1750.0 miles based on the core recovery. According to the modified K-means clustering, the section could be divided into nine parts, and this section was divided into eight parts based on the traditional K-means clustering, one phengite eclogite layer is neglected.

The great error section is from 1705.2 to 1716.5 miles, there are two thin phengite eclogite layers and one fracture zone, while there is no phengite eclogite presented with the two methods in the corresponding depth. Because of the thin layer, the well logging response of phengite eclogite

was affected by the surrounding rocks and did not present the characteristic of phengite eclogite layers, so they were not properly identified. In general, the lithology of fracture zones is the same with the surrounding rocks. The density of fracture zone is about 1.55–1.67 g/cm<sup>3</sup> which is so low that the entropy information is small; thus its contribution to the lithology recognition is low. The phengite eclogite layer was not identified with the method of traditional K-means clustering in the depth Section 1728.3–1730.7 miles, and there were several slight depth errors in the rock transition section. Therefore, the modified K-means cluster has better correspondence with the core recovery than the traditional K-means cluster.

### Clustering analysis

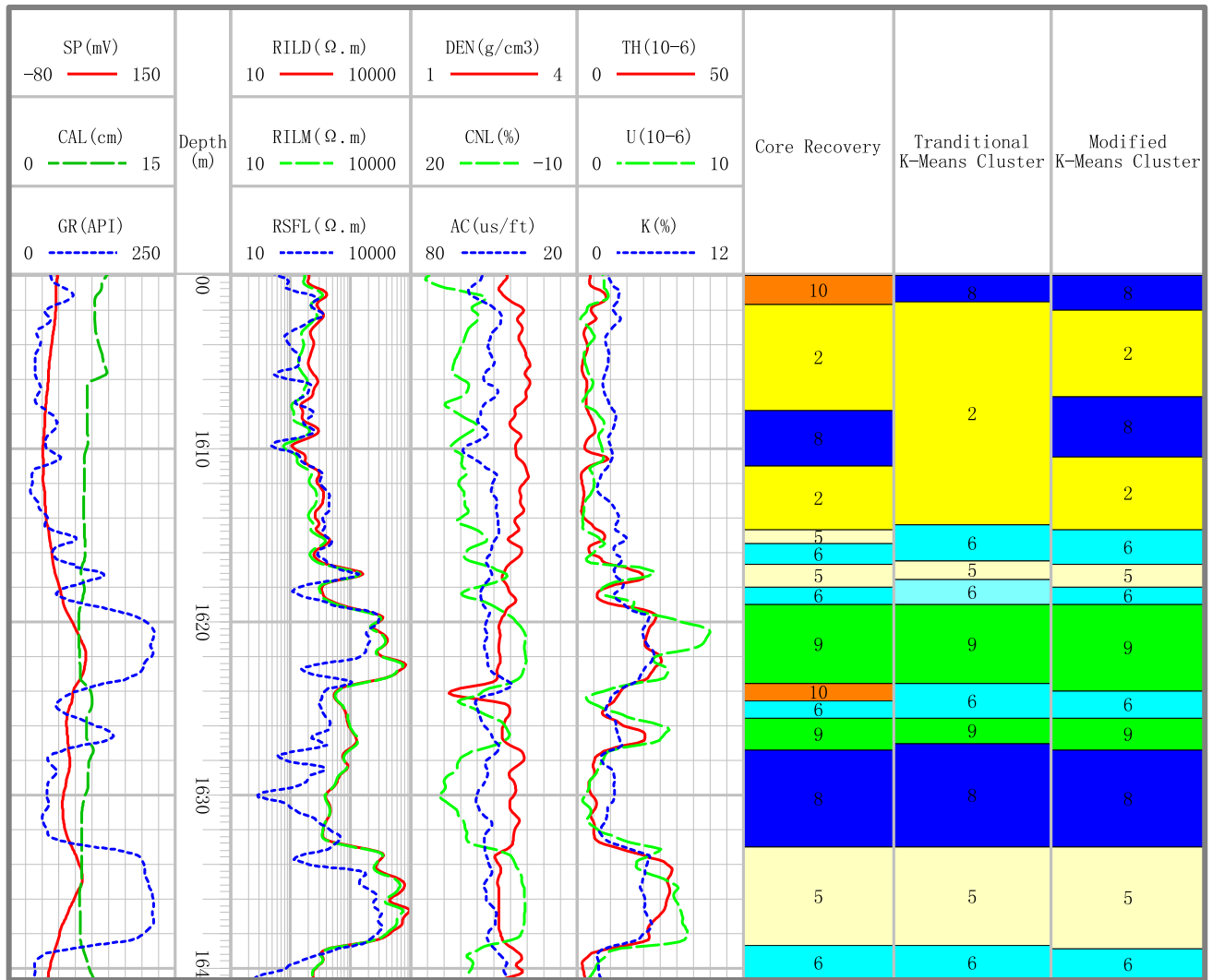
In order to verify the striking effect that modified K-means cluster algorithm has brought about, the same 45 samples were used by the traditional K-means cluster algorithm to



**Table 3** The undetermined samples and corresponding lithology

No	RILD (Ω m)	CNL (%)	AC (μs/ ft)	DEN (g/ cm <sup>3</sup> )	GR (API)	U (10 <sup>-6</sup> )	TH (10 <sup>-6</sup> )	Lithology 1	CC1	CC2	CC3	CC4	CC5	CC6	CC7	CC8	CC9	Maximum	Lithology 2
1	1101.680	7.860	42.690	3.280	20.700	0.930	1.090	1	0.972	0.856	0.497	0.639	0.223	0.864	0.684	0.258	0.823	0.972	1
2	1073.670	7.900	41.470	3.280	20.450	0.940	1.030	1	0.977	0.850	0.497	0.638	0.223	0.859	0.678	0.258	0.817	0.977	1
3	957.570	8.060	41.070	3.300	19.970	0.950	0.980	1	0.980	0.846	0.497	0.637	0.223	0.855	0.674	0.257	0.813	0.980	1
4	234.310	11.020	51.650	3.060	44.540	0.520	3.160	2	0.825	0.981	0.588	0.730	0.314	0.935	0.745	0.348	0.963	0.981	2
5	226.690	11.140	51.490	3.070	43.030	0.460	3.000	2	0.832	0.992	0.576	0.718	0.302	0.927	0.742	0.336	0.955	0.992	2
6	219.780	11.300	51.280	3.080	41.380	0.430	2.770	2	0.835	0.995	0.569	0.711	0.295	0.922	0.743	0.329	0.949	0.995	2
7	2505.490	0.030	53.190	2.570	171.340	2.810	22.330	3	0.451	0.541	0.936	0.791	0.734	0.569	0.510	0.762	0.591	0.936	3
8	2498.580	0.320	54.570	2.650	160.870	2.950	21.890	3	0.482	0.572	0.967	0.779	0.695	0.600	0.508	0.734	0.622	0.967	3
9	2638.080	0.740	55.500	2.750	148.300	3.080	20.540	3	0.531	0.621	0.937	0.779	0.644	0.649	0.514	0.683	0.671	0.937	3
10	1483.410	-0.350	52.360	2.590	125.970	1.010	7.980	4	0.634	0.723	0.775	0.984	0.571	0.724	0.675	0.606	0.774	0.984	4
11	1596.880	-0.720	52.550	2.580	123.340	0.860	7.900	4	0.633	0.722	0.772	0.985	0.569	0.718	0.677	0.602	0.773	0.985	4
12	1477.460	-0.200	51.930	2.590	158.780	1.180	7.860	4	0.634	0.726	0.768	0.976	0.560	0.719	0.685	0.595	0.776	0.976	4
13	4611.170	-0.600	52.750	2.590	216.860	5.880	27.210	5	0.248	0.337	0.750	0.617	0.953	0.366	0.294	0.912	0.388	0.953	5
14	4482.050	-0.650	52.660	2.590	217.470	5.970	26.900	5	0.250	0.340	0.752	0.620	0.951	0.368	0.295	0.914	0.390	0.951	5
15	5019.150	-0.700	51.990	2.600	218.020	6.050	26.710	5	0.242	0.331	0.738	0.607	0.962	0.360	0.282	0.909	0.382	0.962	5
16	347.220	9.800	46.920	3.010	47.170	1.670	3.260	6	0.833	0.931	0.601	0.707	0.327	0.984	0.701	0.362	0.924	0.984	6
17	359.110	9.240	46.070	3.000	43.330	1.510	2.440	6	0.849	0.929	0.590	0.703	0.315	0.985	0.711	0.350	0.916	0.985	6
18	350.860	8.790	45.310	2.970	40.540	1.350	1.800	6	0.857	0.916	0.583	0.704	0.309	0.975	0.722	0.344	0.915	0.975	6
19	751.511	42.852	56.460	2.662	12.870	0.515	1.992	7	0.675	0.762	0.517	0.668	0.252	0.725	0.981	0.291	0.776	0.981	7
20	728.653	41.869	56.373	2.661	11.790	0.520	1.838	7	0.678	0.763	0.518	0.669	0.253	0.727	0.979	0.292	0.778	0.979	7
21	766.935	41.584	56.389	2.658	10.650	0.534	1.679	7	0.679	0.760	0.517	0.670	0.254	0.725	0.979	0.292	0.776	0.979	7
22	3890.460	-0.770	52.270	2.620	218.610	7.850	20.980	8	0.264	0.353	0.769	0.625	0.856	0.382	0.312	0.973	0.404	0.973	8
23	3967.650	-0.660	53.230	2.600	214.010	7.600	20.470	8	0.268	0.357	0.774	0.633	0.855	0.386	0.318	0.968	0.408	0.968	8
24	3671.190	-0.610	53.330	2.600	213.280	7.450	20.460	8	0.278	0.368	0.786	0.644	0.851	0.396	0.329	0.956	0.418	0.956	8
25	418.760	10.880	51.960	2.950	50.380	0.650	4.520	9	0.802	0.948	0.621	0.763	0.347	0.928	0.752	0.381	0.994	0.994	9
26	398.030	10.480	51.120	2.970	51.300	0.620	4.860	9	0.806	0.952	0.616	0.759	0.342	0.934	0.741	0.376	0.984	0.984	9
27	381.150	10.260	50.600	2.960	52.800	0.670	5.120	9	0.805	0.943	0.620	0.763	0.346	0.933	0.736	0.381	0.978	0.978	9

1 Rutile eclogite, 2 phengite eclogite, 3 retrograded eclogite, 4 paragneiss, 5 orthogneiss, 6 amphibolites, 7 serpentinite, 8 chlorite amphibolites and 9-moyite; Lithology 1 represents core recovery; Lithology 2 represents recognition result; CCi represents the hamming approach degree of samples and cluster center  $i$  ( $i = 1, 2, 3, \dots, 9$ )



**Fig. 3** Well logging graphic of CCSD-MH lithology analysis in the depth Section 1600.0–1640.7 m. 1 Rutile eclogite, 2 phengite eclogite, 3 retrograded eclogite, 4 paragneiss, 5-orthogneiss, 6 amphibolites, 7 serpentinite, 8 chlorite amphibolites, 9 moyite and 10 fracture zone

identify the lithology of CCSD-MH. The iteration times, the threshold value, and the cluster numbers were given as  $T = 1250$ ,  $\epsilon = 0.001$ , and  $z = 9$ , respectively. The results are shown in Table 4.

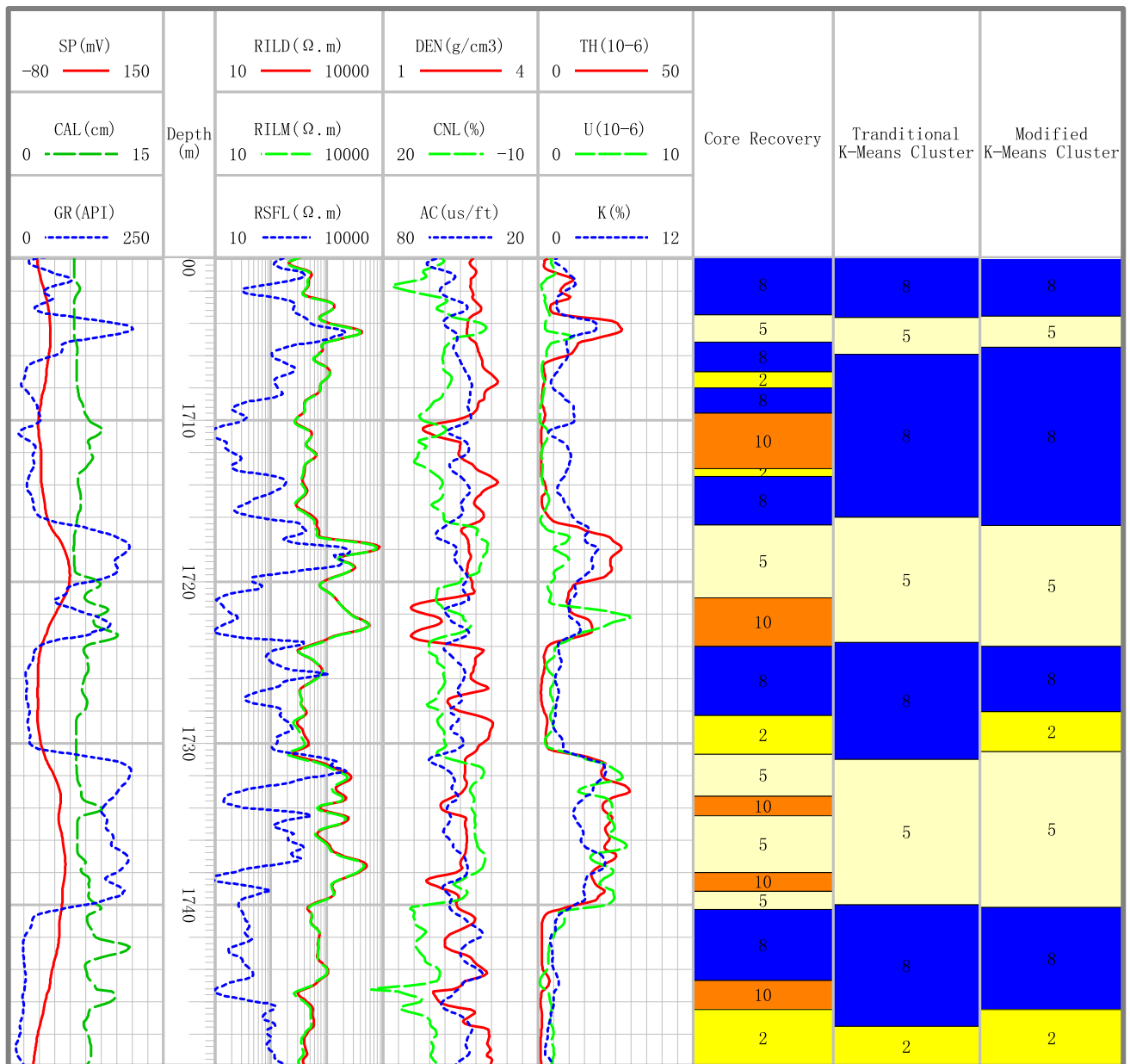
As shown in Table 4, the accuracy of modified K-means cluster algorithm is higher than traditional K-means cluster algorithm. This is mainly because of the cluster centers that the modified K-means cluster algorithm calculated used Eq. (3), which reduces the iteration times and avoids falling into local optimums. Moreover, the algorithm takes the weight values into account, which distinguishes the contributions of different well logs to lithology identification. For the same accuracy, the modified K-means cluster

algorithm would take less iteration times and time than the traditional K-means cluster algorithm. Therefore, the identification result showed that the modified K-means cluster algorithm has high identification accuracy and improve practicability.

### Conclusions

Reservoir lithology is one of the most important data to evaluate rock formation; it mainly carried out by core recovery in laboratory which is very expensive, and its interpretation is time consuming. Accurate identification of





**Fig. 4** Well logging graphic of CCSD-MH lithology analysis in the depth Sect. 1700.0 ~ 1750.0 m. 1 Rutile eclogite, 2 phengite eclogite, 3 retrograded eclogite, 4 paragneiss, 5 orthogneiss, 6 amphibolites, 7 serpentinite, 8 chlorite amphibolites, 9 moyite and 10 fracture zone

**Table 4** Performance comparison of clustering analysis

Algorithm name	Iteration times	Accuracy rate (%)	Time (ms)
Traditional K-means cluster	1250	77.78	553
Modified K-means cluster	1250	88.89	428

lithology from geophysical log data plays a significant role in reservoir evaluation.

In this study, a fast and practical K-means clustering algorithm was proposed based on the shortcoming of the traditional K-means clustering algorithm. Euclidean distance was replaced by Mahalanobis distance, and the initial cluster centers are acquired from the average of characteristic values in matrix but not selected randomly, in

addition, adding weight value in each characteristic value of the objective function. For the same 45 rock samples of CCSD-MH, the accuracy rate of traditional K-means clustering algorithm is 77.78 %, while modified K-means clustering algorithm is 88.89 %, which shows that our modified algorithm is applicable.

According to the modified K-means cluster algorithm, 45 samples can be classified into 9 classes and get 9 cluster centers. The classes of 27 undetermined samples were identified by analyzing the hamming approach degree curves. The predicted results and core recovery are exactly the same by comparison.

In conclusion, the modified K-means clustering algorithm gives better accuracy of CCSD-MH lithology classification than traditional K-means clustering. The hamming approach degree, as well as the cluster centers identified the whole well of CCSD-MH lithology effectively and accurately. Since its flexibility and capability in identifying the lithology, the model can be served as an effective approach in evaluating rock reservoirs when it lacks of core recovery materials. Further, the model may be useful in other fields.

**Acknowledgments** This work was financially supported by the National Natural Science Foundation of China (No.41074086) and the National Natural Science Foundation of China (No.41104080).

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Bezdek JC, Fordon WA (1979) The application of fuzzy set theory to the medical diagnosis. *Advances of Fuzzy Sets and Theories*. North-Holland, Amsterdam, pp 445–461
- Bosch D, Ledo J, Queralt P (2013) Fuzzy logic determination of lithology from well log data: application to the KTB project data set. *Surv Geophys* 34(4):413–439
- Feng F, Xu SG, Liu JW, Liu D, Wu B (2010) Comprehensive benefit of flood resources utilization through dynamic successive fuzzy evaluation model: a case study. *Sci China Technol Sci* 53(2):529–538
- Gao XB (2004) Fuzzy cluster analysis and its applications. Xi'an University of Electronic Science and Technology Press, China
- Gath I, Geva AB (1989) Fuzzy clustering for the estimation of the parameters of the components of mixtures of normal distributions. *PRL* 9(1):77–86
- Gu JF, Jin ZK, Zhu LF et al (2009) The identification of UHP metamorphic of Chinese Continental Scientific Drilling. *Proc Geophys* 24(4):1252–1256
- Jing JE, Wei WB, Jin S et al (2007) Classification and well-logging identification of eclogite in main hole of Chinese Continent Scientific Drilling Project. *Earth Sci J China Univ Geosci* 32(4):504–510
- Karayiannis NB, Mi GW (1997) Growing radial basis neural network: merging supervised and unsupervised learning with network growth techniques. *IEEE Trans. NN* 8(6):1492–1506
- Liao SY (2013) Fuzzy C-means, K-means clustering algorithm and their parallelization. Taiyuan University of Science and Technology, China, pp 38–40
- Liu FL, Xu ZQ, Xue HM (2004) Tracing the protolith, UHP metamorphism, and exhumation ages of orthogneiss from the SW Sulu terrane (eastern China): SHRIMP U-Pb dating of mineral inclusion-bearing zircons. *Lithos* 78(4):411–429
- Liu YS, Zhang ZM, Lee CT, Gao S (2005) Decoupled high-Ti from low-Nb (Zr) of eclogites from the CCSD: implications for magnetite fractional crystallization in basalt chamber. *Acta Petrologica Sinica* 21(2):339–346
- Luo M, Pan HP (2010) Well logging responses of UHP metamorphic rocks from CCSD main hole in Sulu terrane, eastern central China. *J Earth Sci* 21(3):347–357
- Mukherjee DP (1995) Water quality analysis: a pattern recognition approach. *Pattern Recogn* 2(28):269–281
- Newton SC (1992) Adaptive fuzzy leader clustering of complex data sets in pattern recognition. *IEEE Trans NN* 3(5):794–800
- Ni JL, Liu JL, Tang XL, Yang HB, Xia ZM, Guo QJ (2013) The wulian metamorphic core complex: a newly discovered metamorphic core complex along the Sulu orogenic belt, eastern China. *J Earth Sci* 24(3):297–313
- Runkler TA, Palm RH (1996) Identification of nonlinear system using regular fuzzy c-elliptotype clustering. *FUZZIEEE* 96: 1026–1030
- Ryoke M, Nakamori Y, Suzuki K (1995) Adaptive fuzzy clustering and fuzzy prediction models. *FUZZIEEE'95*: 2215–2220
- Wang XX, Zattin M, Li JJ, Song CH, Chen S, Yang C, Zhang SD, Yang JW (2013) Cenozoic tectonic uplift history of western qinling: evidence from sedimentary and fission-track data. *J Earth Sci* 24(4):491–505
- Windham C (1985) Cluster analysis to improve food classification within commodity groups. *J Amer Assoc* 85(10):1306–1315
- Wu XH, Niu SJ, Wu CO et al (2011) An improvement on estimating covariance matrix during cluster analysis using mahalanobis distance. *J Appl Stat Manage* 30(2):240–245
- Xu ZQ, Yang WC, Zhang ZM, Yang TN (1998) Scientific significance and site-selection researches of the first Chinese Continental Scientific Deep Drillhole. *Cont Dyn* 3(1/2):1–13
- Xu HJ, Jin ZM, Ou XG (2006) Lithology determination of rocks from CCSD 100–2000 m main hole by magnetic susceptibility and density using discriminant function analysis. *Earth Sci J China Univ Geosci* 31(4):513–519
- Ya K, Yao YP, Katayama I, Cong BL, Wang QC, Maruyama S (2000) Large areal extent of ultrahigh-pressure metamorphism in the Sulu ultrahigh-pressure terrane of East China: new implications from coesite and omphacite inclusions in zircon of granitic gneiss. *Lithos* 52(1–4):157–164
- Zhang ZM, Xu ZQ, Xu HF (2000) Petrology of ultrahigh-pressure eclogite from the ZK703 drillhole in the Donghai, eastern China. *Lithos* 52(1):35–50
- Zhang ZM, Xu ZQ, Xu HF (2003) Petrology of the non-mafic UHP metamorphic rocks from a drillhole in the Southern Sulu orogenic belt, eastern-central China. *Acta Geol Sinica* 77(2):173–186
- Zhang ZM, Xiao YL, Hoefs Jochen, Liou JG, Simon K (2006) Ultrahigh pressure metamorphic rocks from the Chinese

Continental Scientific Drilling Project: I. Petrology and geochemistry of the main hole (0–2, 050 m). *Contrib Mineral Petrol* 152(4):421–441

Zhang GB, Pang JH, Chen GH, Ren XL, Ran Y (2010) An evaluation method based on multiple quality characteristics for

CNC machining center using fuzzy matter element. *Proceedings of the 36th International MATADOR Conference* 5–4: 157–160