Send Orders for Reprints to reprints@benthamscience.ae

The Open Petroleum Engineering Journal, 2015, *8*, (Suppl 1: M7) 297-307

An Identified Method for Lacustrine Shale Gas Reservoir Lithofacies Using Logs: A Case Study for No. 7 Section in Yanchang Formation in Ordos Basin

Hongyan Yu[1,2,*], Zhenliang Wang[1], Hao Cheng[1], Qianqian Yin[1], Bojiang Fan[4], Xiaoyan Qin[1], Xiaorong Luo[3], Xiangzeng Wang[4] and Lixia Zhang[4]

[1]State Key Laboratory of Continental Dynamics, Geology Department of Northwest University, Xi'an, 710069, P.R. China; [2]Research Institute of BEG, CNPC, Zhuozhou 072750, China; [3]Institute of Geology and Geophysics & Key Laboratory of Petroleum Resource Research, Chinese Academy of Science, Beijing 100029, China; [4]Shaanxi Yanchang Petroleum(Group) Co., Ltd, Xi'an 710075, China

Abstract: Unconventional reservoirs are keys to oil and gas exploration and development, especially shale gas reservoirs. Discriminated shale gas reservoir lithofacies are, in particular, a primary problem in shale gas reservoir engineering. The mineral composition will affect both absorbed and free gas contents, therefore their identification is important. The mineral composition is one part of lithofacies. The shale content has always been used in previous lithological identifications: this method is effective in sand reservoirs; however, it is not suitable for use in shale gas reservoirs. This paper takes No.7 section in Yanchang formation in Ordos basin as an example. Through a lithological analysis, it was concluded that overlap method and cross-plot method are not also inappropriate for shale gas reservoirs. The Ordos basin shale gas reservoir is divided into seven lithofacies. We form a mathematical method and apply it to shale gas reservoirs using the shale volume and ΔlgR which are available from conventional well logging and reflect organic matter in the processed dataset. Decision tree is used here. However, there were too many parameters to discriminate all lithofacies precisely. Principal component analysis (PCA) is a technique used to reduce multidimensional data sets to lower dimensions for analysis. This technique can be useful in petro-physics and geology as a preliminary method of combining multiple logs into a single entity or two logs without losing information. Combining PCA and a decision tree algorithm, the lithofacies of a shale gas reservoir were accurately discriminated.

Keywords: Lacustrine facies, lithofacies, shale gas, well logging.

1. INTRODUCTION

In the mid-late period of oil and gas field development, unconventional reservoirs are important for increasing hydrocarbon reserves and production. Shale gas reservoirs are an important exploration direction for unconventional reservoir engineering. Shale gas reservoirs are self-generating and self-accumulating. Their stratification comprises fine mixtures composed of organic matter, hydrocarbons, and rock debris, therefore any physical analysis and testing thereof is complicated. This brings significant challenges and opportunities to identify, evaluate, and predict the productivity of shale gas reservoirs. Shale gas reservoirs are characterized by low porosity and low permeability and lithology thereof is quite different from that of conventional reservoirs. So exactly identifying the lithological characteristics of such reservoirs may provide the basis for further research and development of shale gas reservoirs.

Various studies have found that for shale deposits with laminations, sand and shale inter-laminar strata are an important reservoir space [1]. For shale reservoirs, it is not enough to calculate the shale content alone. Firstly, in addition to shale sections with high shale contents, shale reservoirs also include shale inter-laminar sections. These layers cannot be recognised when only calculating the shale content. Shale with laminations contributes more to the reservoirs than mudstone. So exactly identifying the formation lithology, composition, and structure, of internal minerals is a key to productive shale gas exploration. According to research from the US National Petroleum Council, using geochemical analysis combined with logs represents an important direction for future development [2]. Accurately identifying layer lithofacies can greatly improve the success rate of shale gas reservoir developments. Currently, due to the development cost, the amount of drilling logs is limited and is unable to meet the requirements of modern shale gas exploration. Logs contain a wealth of geological information [3-9]. Using this geophysical method to identify lithologies has high accuracy and low costs. So, this research focusses on the indentification of the lithofacies of shale gas reservoirs through the features contained in geophysical logs.

The research takes the Yanchang formation of triassic period in the south-eastern Ordos Basin as an example (Fig. **1**): previous research and field tests show that this area has lacustrine shale gas accumulation conditions. This lacus-

*Address correspondence to this author at the State Key Laboratory of Continental Dynamics, Geology Department of Northwest University, Xi'an, 710069, P.R. China; Tel: +86-187-9299-5250; Fax: +86-029-88302202; E-mail: yuhy@nwu.edu.cn

1874-8341/15 2015 Bentham Open

trine deposit was also multi-level, multi-type, evolutionary, with miscellaneous other phenomena present (Fig. **2**). Some experts suspect that lacustrine deposits do not have wide-spread shale gas development and exploitable shale gas exploration potential [10]. Despite the development potential, marine shale gas reservoirs are not preferred by those experts. However, based on the investigation of shale gas wells in the south-eastern Ordos Basin, Zhang had analysed shale gas accumulation features, shale gas generation and accumulation mechanisms, and its enrichment [11-13]. Xu studied the Ordos Basin shale gas and considered the claystone of the Yanchang formation to be widely distributed [14]. The thickness of the shale, and its organic carbon content, are higher; thermal evolution is moderate and gas measurement shows activity. Shale fractures are well developed, and its shale gas resource potential is significant. So, the south-eastern region of the Yanchang formation in the Ordos Basin is a shale gas range with potential profitability (indeed data suggest it represents an improvement over some marine shale deposits). In April 2011, to review LP177 Well changes, seven sections of the shale strata were pressure-tested (yielding a daily volume of 2350 m^3 of natural gas) thus making the first lacustrine shale gas well in the world. Many wells have been subsequently fractured successfully and industrial gas-flows obtained therefrom. From then on, it formed a prelude to a lacustrine shale gas development revolution in China. Unlike extensive research into marine shale gas deposits, lacustrine shale gas engineering can call upon no previous experience. In the process of exploration and operation, engineers are confronted with multiple challenges. Therefore, the establishment of a set of continental facies identification methods for shale gas was urgently required.

This research analysed the characteristics of a terrestrial shale gas reservoir. Firstly, cross-plot and curve over-play methods were used to identify the lithofacies therein. However, these conditional methods could not evaluate the lithofacies accurately. Using principal component analysis (PCA), the relevant parameters were obtained therefrom. Then, a decision tree algorithm was used to discriminate between the lithofacies using PCA parameters.

2. WELL-LOGGING RESPONSE CHARACTERISTICS OF THE LITHOFACIES

Shale gas reservoirs are not the unique source of black shale. All tight microclastic rocks, which are rich in organic material, and where gas is present in both absorbed and free forms, are effectively shale gas reservoirs. Shale gas reservoirs have complex mineral compositions. The chosen research area contains quartz, feldspar, clay minerals, small amounts of calcite and dolomite, pyrite, and occasionally siderite. The mineral compositions vary greatly between wells and in different layers. Well-logging response characteristics are always affected by lithology and mineral composition. This research is committed to analyzing wells through different lithologies and will be beneficial for identifying future exploitable reservoirs. Lithofacies, a part of the sedimentary facies, are always used in oil geology, especially in stratigraphy and sedimentology. Lithofacies have a direct relationship with mineral and organic contents of strata.

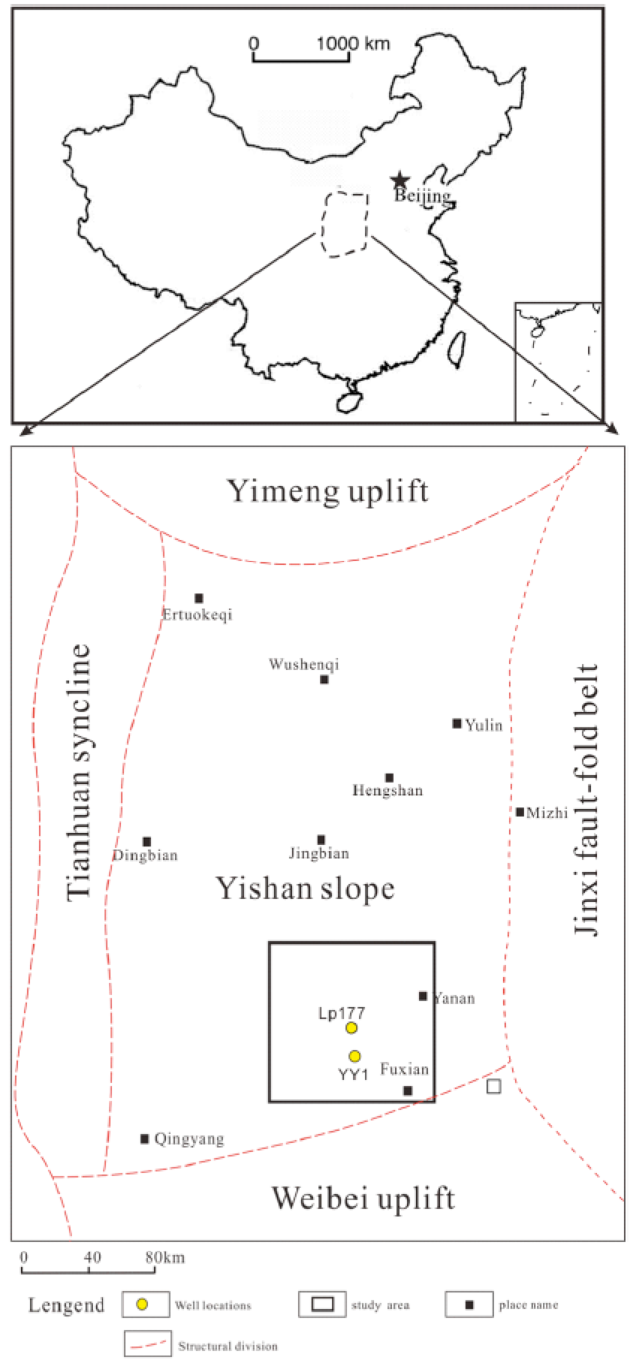


Fig. (1). Research area location map.

Through XRD and organic matter pyrolysis experiments, the sedimentary lithofacies were divided into fine sandstone (F S), siltstone (S S), sand and shale inter-laminae with more shale (S & S(sh)), sand and shale inter-laminae with more sand (S & S(sa)), mudstone (M S), organic-rich black shale(B S), and ash tuff (A S).

Here, box-and-whisker plots were used to represent the lithology and the diversity of the different lithofacies and their logging response characteristics. Box and whisker plot contain six data nodes: a set of data is arranged from large to small and the upper limb, upper quartile, median lower quartile, lower edge, and some other outliers were calculated and shown therewith. From the acoustic time log and neutron porosity log, it may be seen from Fig. (**3**) that there were obvious differences between the black shale and other rocks. Siltstone and fine sandstone were also easily distinguished from other rocks, but the gaps between them were small. Though most of lithologic characteristics of ash tuff were the same as those of other rocks, the radiation signature of the ash tuff was strong. Although a lot of differences were found between S & S(sh) and S & S(sa), there was still an overlapping area in the single logging curve. From the box and whisker plots of deep and shallow resistivity, the resistivity of fine sandstone, siltstone, ash tuff, S & S(sh), and S & S(sa) were all between 30 and 1000 Ω·m. Fig. (**1**) showed that mudstone had a low resistivity, and that shale had a high resistivity here. Because the shale was fine-grained and lacking in organic matter, it formed a well-developed conductive network.

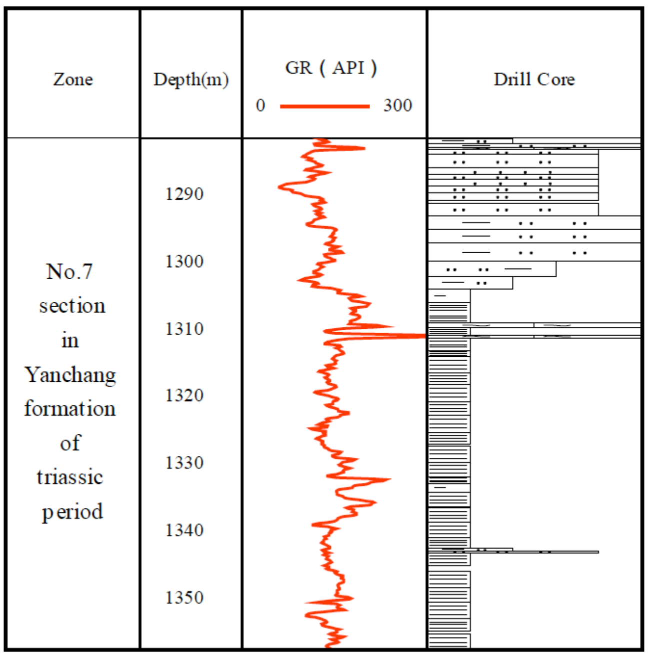


Fig. (2). The stratigraphy of research area.

3. LITHOFACIES IDENTIFICATION

3.1. Traditional Identification Methods

3.1.1. Qualitative Identification

According to the logging curve characteristics of different rock types in the study area, we identified the rock types from well-logging features such as AC, RT, and other logging curves. GR reflected the total shale content, the permeability of rock could be deduced through SP, AC indicated the total porosity, and the resistivity was indicated by RT. According to the aforementioned characteristics, we can identify the rock types in each target layer. Due to the poor permeability, strong adsorption capacity, and relatively high total porosity, the claystone was easily distinguished. If it contained oil and gas, it would have a high resistivity. Compared with claystone, sandstone has a greater permeability, low SP, low GR, and low to medium AC. The characteristics of argillaceous siltstone and silty mudstone lie between claystone and permeable sandstone. This method can qualitatively discriminate between pure sand and pure shale; however, just as shown in Fig. (**4**), S & S(sa) cannot be discriminated accurately. A photo of S & S(sa) is shown in Fig. (**5**).

3.1.2. Curve Overlap Method

Generally speaking, the overlap method is one using different logging response characteristics and their differences to recognise the lithology of each stratum. Corresponding to the different lithologies, sensitivity and logging curve are different. Selecting the most two sensitive curves, confirming the base line, then overlapping the two curves forms the basis for the method. As shown in Fig. (**6**), use of acoustic and neutron overlapping can readily identify shale; but it still suffers from some qualitative identification problems depending upon the local geology.

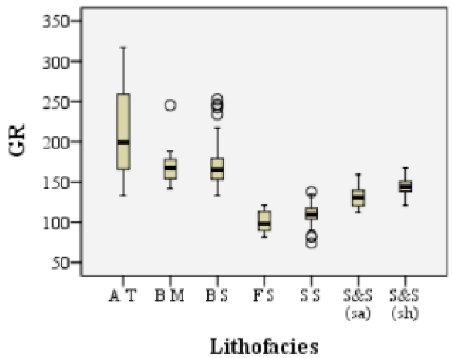
3.1.3. Cross-plot Method

Cross-plotting uses different lithofacies well-logging response characteristics. From Fig. (**7**), we found those seven lithofacies showed up in different regions when cross-plotted. Black shale had a higher -value, a longer acoustic time, and a higher resistivity than other strata. However, there remained the problem of fuzzy interfaces. For now though, it was concluded that the conventional method was not suitable for application to shale gas reservoirs.

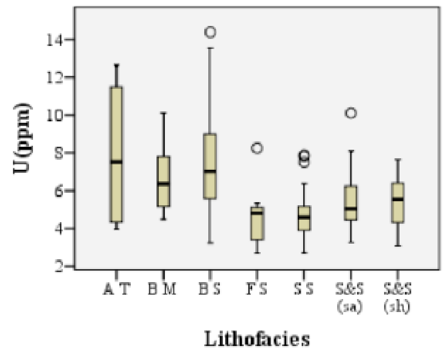
3.2. Mathematical Geological Modelling

3.2.1. The Flow Diagram of Mathematical Geology

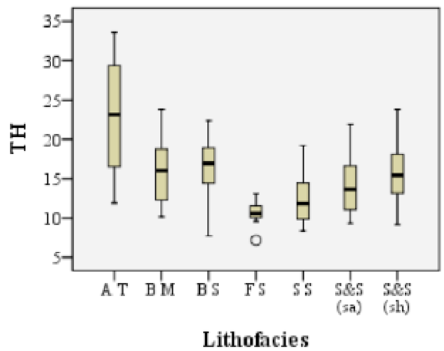
Petroleum exploration has been transformed by the application of artificial intelligence (AI) [15-17]. In the middle, and later, periods of petroleum exploration and development, the total quantity of data has accumulated considerably and AI has become more and more important [18, 19]. Much valuable information can be extracted from the large volumes of petroleum-exploration data. As one of the most significant AI methods, data analysis plays an essential role in petroleum exploration and development [20, 21]. The traditional methods, cross-plot identification or ordinary linear regression, have not resolved the present problem. The dilemma of "more data, less knowledge" arises. So, machine learning, data processing, and a training model are helpful when trouble-shooting [22]. Many results may be obtained from data analysis methods, such as lithology identification, porosity and permeability distributions, flow unit types, sedimentary types, oil and gas and water reservoir identification, and so on. Data analysis technology has accomplished much within the petroleum industry. In future, data analysis will accelerate the development of petroleum exploration. The basic steps for data analysis in petroleum exploration and development are shown in Fig. (**8**). ① Data mining task definition: marking the label attributes. ② Data collection: we should collect more data as far as possible to avoid over-fitting and under-fitting after task definition. ③ Pre-treatment: a high data quality is necessary for data mining. It is important to carry out data cleaning, noise reduction, and missing value processing. ④ Data processing: to make the data more suitable for data mining, attribute selection, feature subset selection, discretisation, binarisation, and other processing steps are needed. ⑤ Machine learning: the input



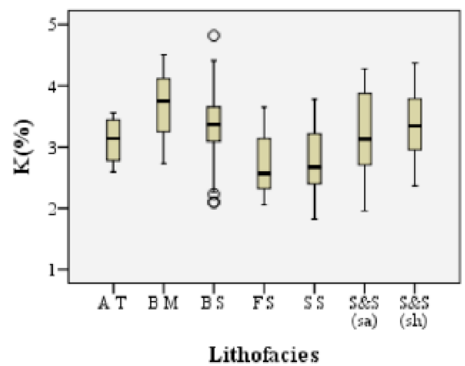
The GR distribution characteristics of lithofacies



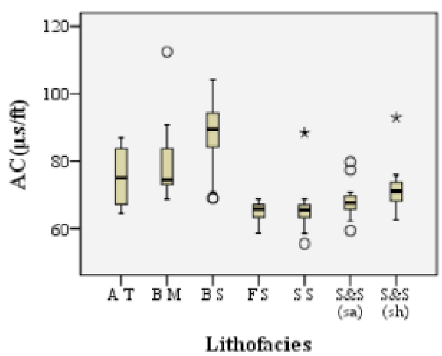
The Uranium distribution characteristics of lithofacies



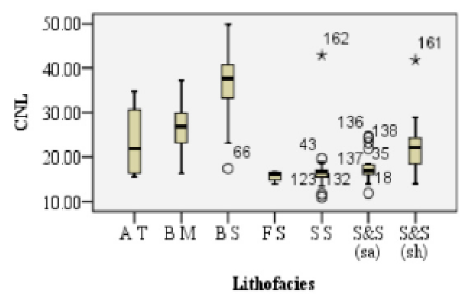
The Thorium distribution characteristics of lithofacies



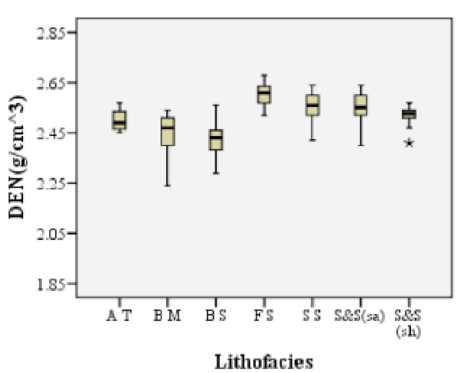
The Kalium distribution characteristics of lithofacies



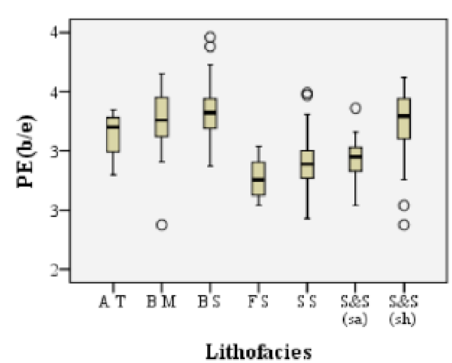
The acoustic time(AC) log distribution characteristics of lithofacies



The neutron porosity(CNL) log distribution characteristics of lithofacies

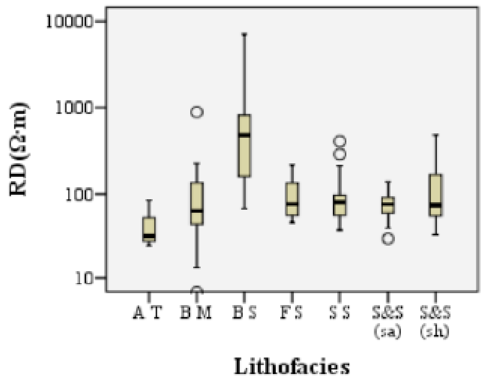


The litho-density(DEN) log distribution characteristics of lithofacies

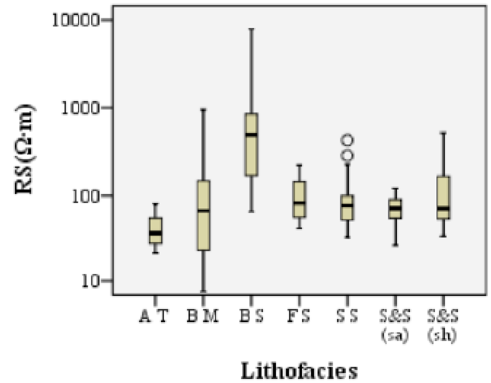


The litho-density(PE) log distribution characteristics of litho-facies

Fig. (3) contd….



The deep resistivity(RD) log distribution characteristics of lithofacies



The shallow investigation resistivity distribution characteristics of lithofacies

Fig. (3). Well logging response characteristics of Terrestrial deposit shale gas reservoir lithofacies.

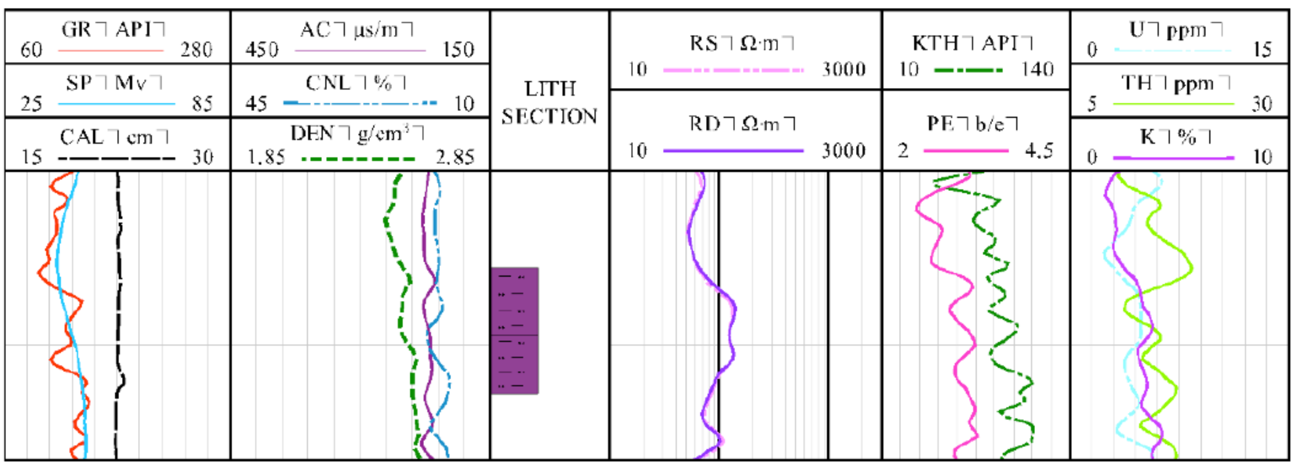


Fig. (4). The well logging characteriscs of S&S(sa).

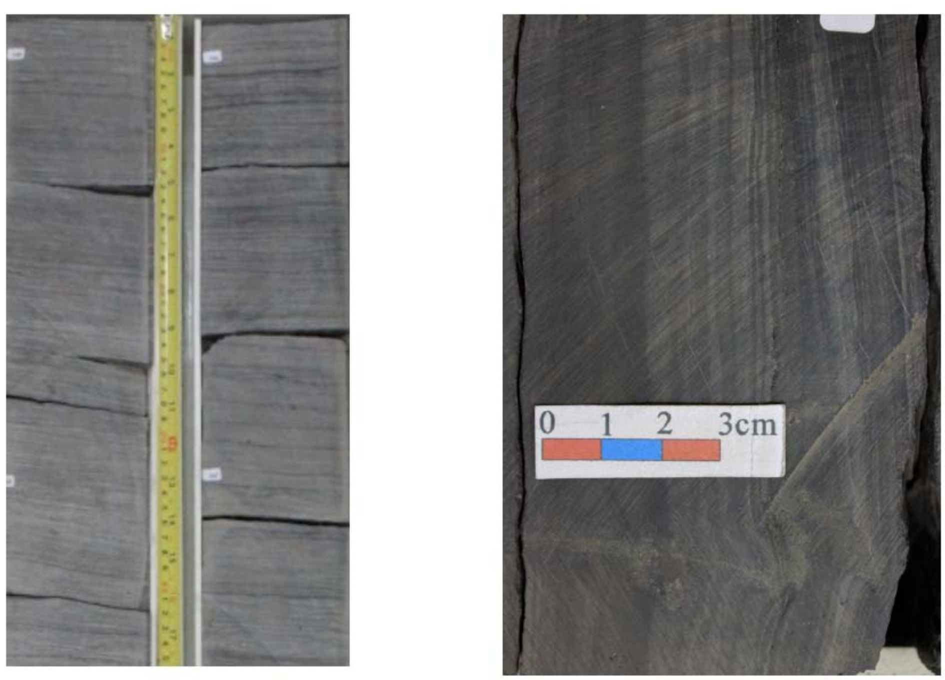


Fig. (5). The photo of S&S(sa).

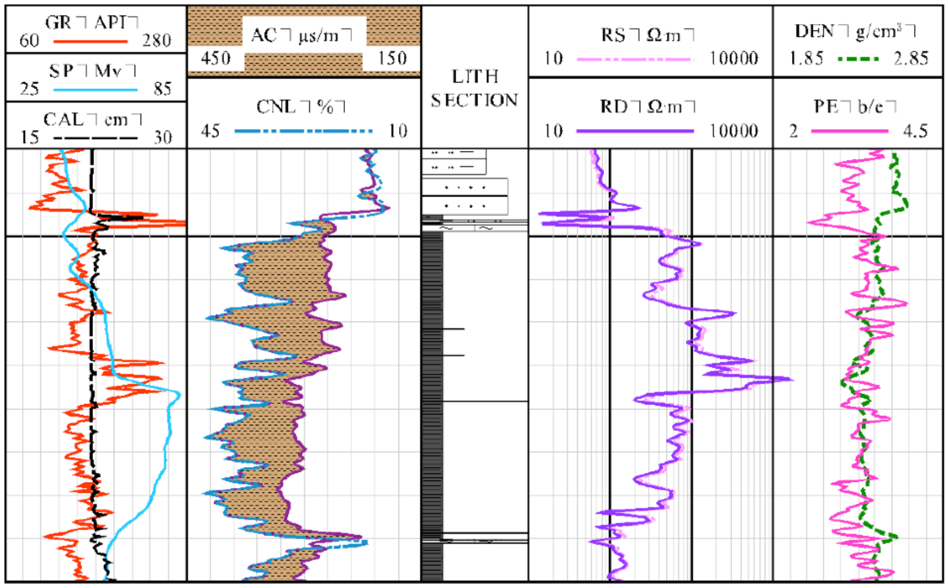


Fig. (6). The curves overlap method of Black shale.

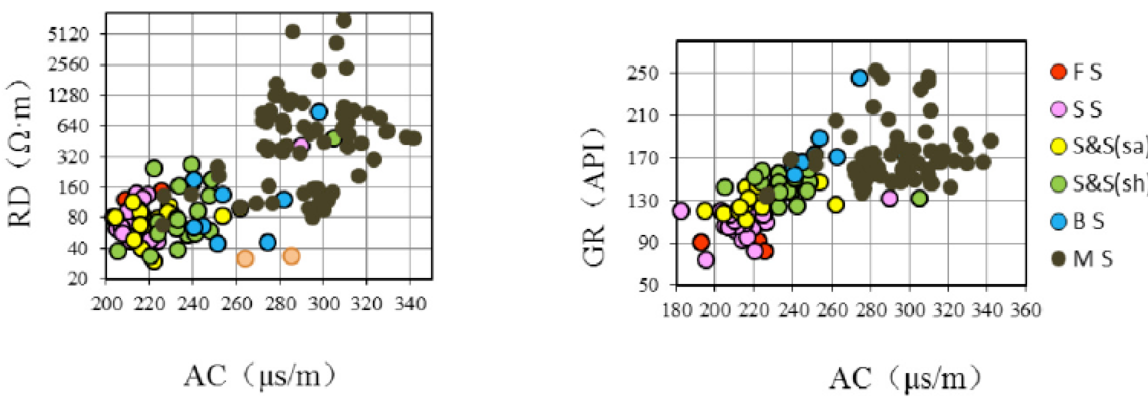


Fig. (7). Cross-plot figure.

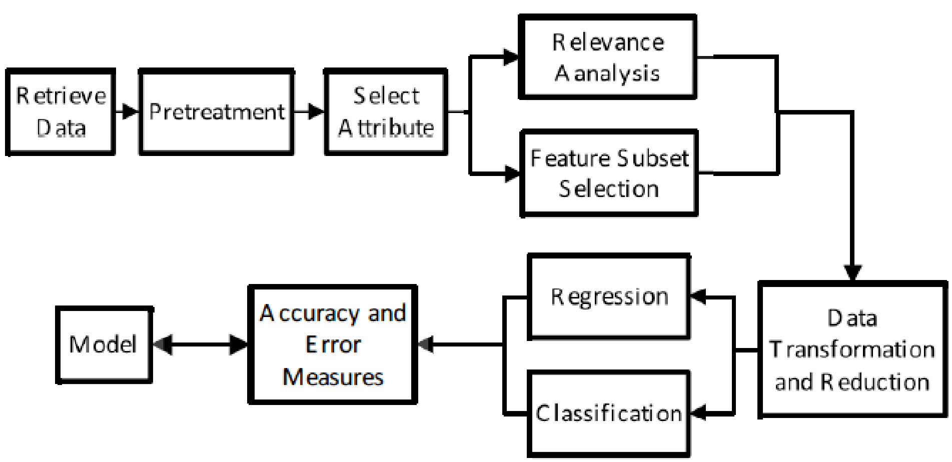


Fig. (8). The flow diagram of mathematical geology.

dataset is vast and varied. Machines need to learn more information from the data. ⑥ Model generation: taking advantage of different data mining algorithms, different models may be obtained by judicious setting of parameters. ⑦ Performance measurement: different models produced by different data mining algorithms should be evaluated for their accuracy. ⑧ Knowledge: that model which has offered the best performance measurement will form part of the body of new knowledge.

3.2.2. Shale Gas Reservoir Dataset

The common attributes which this topic selected include lithology curves (GR NGR SP), porosity logs (AC, DEN, CNL), resistivity curves (RD, RS), as well as PEF, V_{sh} and $\Delta \lg R$ which were calculated from well-logging data.

(1) V_{sh} calculation method

V_{sh}, which is the the clay volume, was mainly calculated from the γ-curve according to the experiential formula shown below:

$$SH = (GR - GMAX)/(GMAX - GMIN)$$
$$V_{sh} = \left(2^{GCUR*SH} - 1\right)/\left(2^{GCUR} - 1\right) \quad (1)$$

In particular : *GMAX*: the log value of clean sandstone

GMIN: the log value for pure shale

GCUR: empirical coefficient, related to the tertiary strata, adopted value here: 3.7.

(2) $\Delta \lg R$ calculation method

$\Delta \lg R$ can reflect the organic matter content and its maturity. Through the log curve overlapping method, overlapping the porosity and resistivity curves in accordance with the appropriate scale, we can determine the source rock types according to the difference between the two curves. Under normal circumstances, the acoustic porosity curve and *r* deep resistivity curve are selected. If the formation is full of water and lacking in organic matter, the two curves are parallel and overlap together. However, in oil and gas reservoirs, and non-reservoirs rich in organic matter, the two curves differ. In application, the acoustic transit time, AC 50 μs/ft (168 μs/m) is equal to a logarithmic unit of resistivity RT:

$$\Delta \lg R = \lg(RT/RT_{base}) + 0.02*(AC - AC_{base}) \quad (2)$$

This research selected different baselines according to the different wells and used (2) to calculate $\Delta \lg R$ which showed a linear correlation with total organic carbon (TOC) content. The organic matter content in the formation can be used to judge differences in lithofacies.

3.2.3. Decision Tree Method

Decision tree (DT) is the most simple and widely used technique. The result from DT analysis looks like a tree (hence the name): it is intuitive, concise, quantitative, and more logical to human interpretation. DT uses a hyperplane with a single attribute to cut the input space repeatedly: each class can thus be divided. The DT analysis is used on a processed dataset to build a predictive model. From the model it was found that the lithofacies in this shale gas reservoir were complex. If the model had accuracy as high as that suggested in Table **1**, the model would over-fit (see Fig. **9**). Over-fitting occurs because the model describes a random error or noise instead of the underlying relationship. The model was also excessively complex. A model that has been over-fitted will generally return poor predictive performance, as it can exaggerate minor fluctuations in the data. Otherwise, if the model were applicable, its accuracy would be poor.

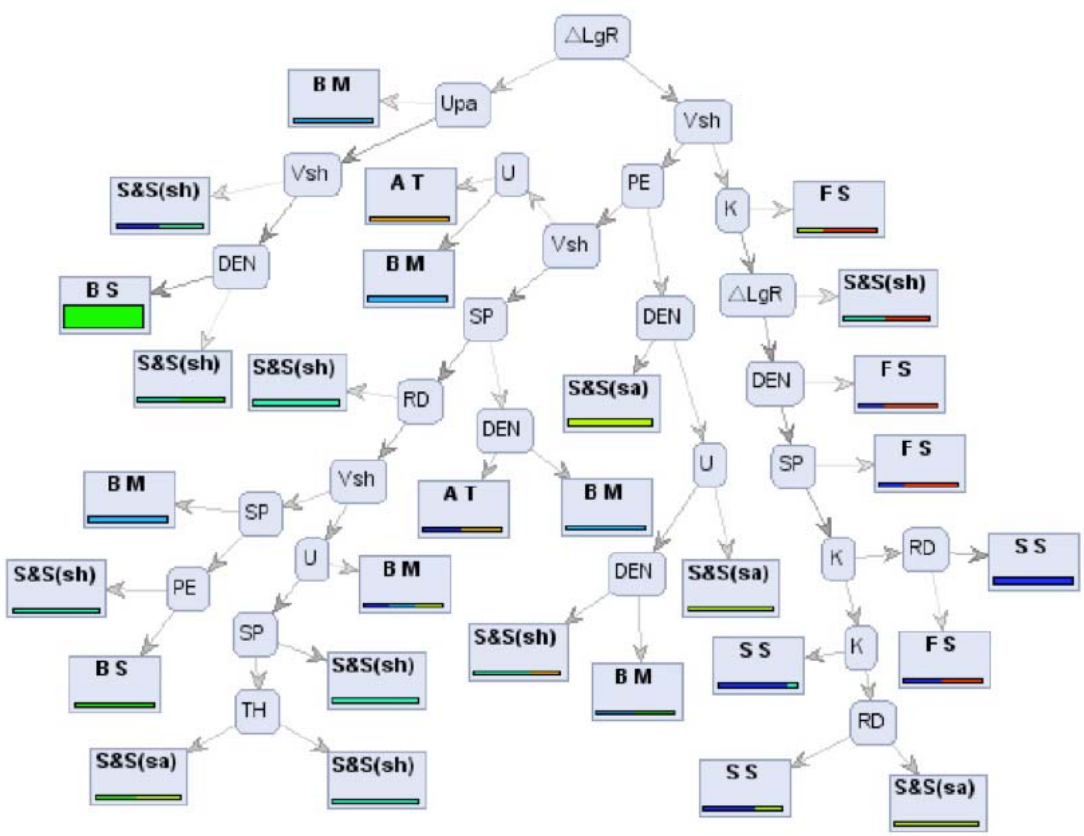


Fig. (9). Decision tree model.

Table 1. Decision tree model accuracy.

	True S S	True B M	True S&S(sh)	True B S	True S&S(sa)	True A T	True F S	Class Precision
Pred. S S	23	0	1	0	1	0	0	92.00%
Pred. B M	1	16	0	1	1	0	0	84.21%
Pred. S&S(sh)	1	0	23	1	0	1	1	85.19%
Pred. B S	0	1	0	64	0	0	0	98.46%
Pred. S&S(sa)	0	0	0	1	19	0	0	95.00%
Pred. A T	1	0	0	0	0	3	0	75.00%
Pred. F S	3	0	0	0	1	0	7	63.64%
Class Recall	79.31%	94.12%	95.83%	95.52%	86.36%	75.00%	87.50%	

Table 2. Principal component variables proportion of variance.

Component	Standard Deviation	Proportion of Variance	Cumulative Variance
PC 1	2.643	0.411	0.411
PC 2	1.633	0.157	0.568
PC 3	1.452	0.124	0.692
PC 4	1.187	0.083	0.774
PC 5	1.073	0.068	0.842
PC 6	0.918	0.05	0.892
PC 7	0.841	0.042	0.933
PC 8	0.637	0.024	0.957
PC 9	0.584	0.02	0.977
PC 10	0.42	0.01	0.988
PC 11	0.35	0.007	0.995
PC 12	0.276	0.004	0.999
PC 13	0.084	0	1
PC 14	0.052	0	1
PC 15	0.04	0	1
PC 16	0.02	0	1
PC 17	0	0	1

3.2.4. Principal Component Analysis (PCA) Method

The former analysis methods failed to grasp key information since the dataset had too many cross-correlated parameters relative to the number of observations. PCA is a statistical procedure which uses an orthogonal transformation to convert a set of attributes of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. The number of principal components is less than or equal to the number of original variables (original attributes). That is to say that many original well-logging attributes were converted to several independent attributes by using the aforementioned orthogonal transformation. These principal component variables can represent a transformation relationship between many well-logging attributes (Table **2**). This will reduce information loss. From the analysis, principal component variable values, and their proportional variance can obtained (Table **2**). The variables which have the greatest proportion of the overall variance are chosen. The input data were normalized for each curve. This

Table 3. Dataset range after normalization.

Attribute	Range
CNL	[-1.529 ; 2.289]
DEN	[-8.510 ; 1.767]
SP	[-2.176 ; 1.897]
GR	[-2.095 ; 4.666]
RD	[-0.444 ; 8.314]
RS	[-0.432 ; 8.746]
LgRD	[-2.647 ; 3.220]
LgRS	[-2.467 ; 3.206]
U	[-1.557 ; 3.554]
K	[-2.227 ; 2.564]
TH	[-1.968 ; 4.486]
PE	[-2.503 ; 2.619]
THK	[-2.030 ; 4.393]
Upa	[-4.043 ; 2.317]
Vsh	[-0.402 ; 11.781]
AC	[-1.823 ; 2.931]
ΔLgR	[-1.538 ; 2.562]

Table 4. The coefficient of principal component variables.

Attribute	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13	PC14	PC15	PC16	PC17
CNL	0.327	0.017	-0.067	-0.321	0.12	-0.014	-0.137	0.027	0.071	0.217	0.361	0.754	-0.011	-0.023	0.018	0.001	0
DEN	-0.244	-0.024	-0.17	0.365	0.392	0.11	-0.246	-0.14	-0.327	0.576	0.084	-0.029	-0.005	-0.02	0.019	0.29	0
SP	0.109	-0.037	0.214	-0.171	0.722	-0.278	-0.236	0.37	0.142	-0.139	-0.149	-0.184	-0.004	-0.087	0.131	-0.017	0
GR	0.295	-0.265	0.109	0.151	-0.112	-0.209	0.08	-0.039	-0.043	0.291	-0.765	0.269	0	0.014	0.004	0.002	0
RD	0.266	0.258	0.103	0.386	0.038	0.133	0.231	0.244	0.2	0.102	0.107	-0.015	-0.375	0.519	0.314	0.031	0
RS	0.263	0.26	0.098	0.392	0.025	0.128	0.223	0.25	0.225	0.141	0.096	-0.024	0.378	-0.501	-0.308	-0.031	0
LgRD	0.306	0.255	-0.099	0.036	0.203	0.192	-0.092	-0.29	-0.174	-0.276	-0.154	0	-0.557	-0.163	-0.437	0.015	0
LgRS	0.307	0.255	-0.108	0.031	0.184	0.18	-0.104	-0.322	-0.163	-0.258	-0.131	0.001	0.623	0.359	0.139	0.011	0
U	0.248	0.017	0.184	0.219	-0.27	-0.376	-0.238	0.29	-0.65	-0.159	0.231	-0.019	0.008	-0.008	-0.002	0.001	0
K	0.124	-0.235	-0.259	-0.018	0.297	-0.324	0.733	-0.156	-0.229	-0.048	0.195	-0.052	0.008	-0.011	-0.001	0	0.105
TH	0.149	-0.47	0.165	0.017	0.026	0.471	-0.046	0.105	-0.08	-0.076	0.066	-0.041	0.01	0.002	-0.005	0	0.689
PE	0.181	-0.158	-0.538	0.052	-0.152	-0.077	-0.134	0.217	0.227	-0.217	-0.037	-0.054	0.008	-0.05	0.032	0.67	0
THK	0.162	-0.486	0.121	0.013	0.069	0.406	0.063	0.078	-0.11	-0.08	0.092	-0.047	0.011	0	-0.005	0	-0.717
Upa	0.071	-0.17	-0.606	0.209	0.014	-0.031	-0.237	0.157	0.083	0.034	0.002	-0.058	-0.015	0.065	-0.039	-0.675	0
Vsh	0.148	-0.289	0.248	0.356	-0.006	-0.327	-0.239	-0.542	0.398	-0.087	0.279	-0.06	-0.02	-0.011	0	-0.001	0
AC	0.321	-0.013	0.026	-0.377	-0.107	-0.091	-0.06	0.01	0.064	0.449	0.101	-0.456	0.056	0.328	-0.44	0.048	0
ΔLgR	0.345	0.112	-0.078	-0.205	-0.139	0.091	-0.037	-0.199	-0.063	0.216	0.014	-0.315	-0.115	-0.449	0.618	-0.073	0

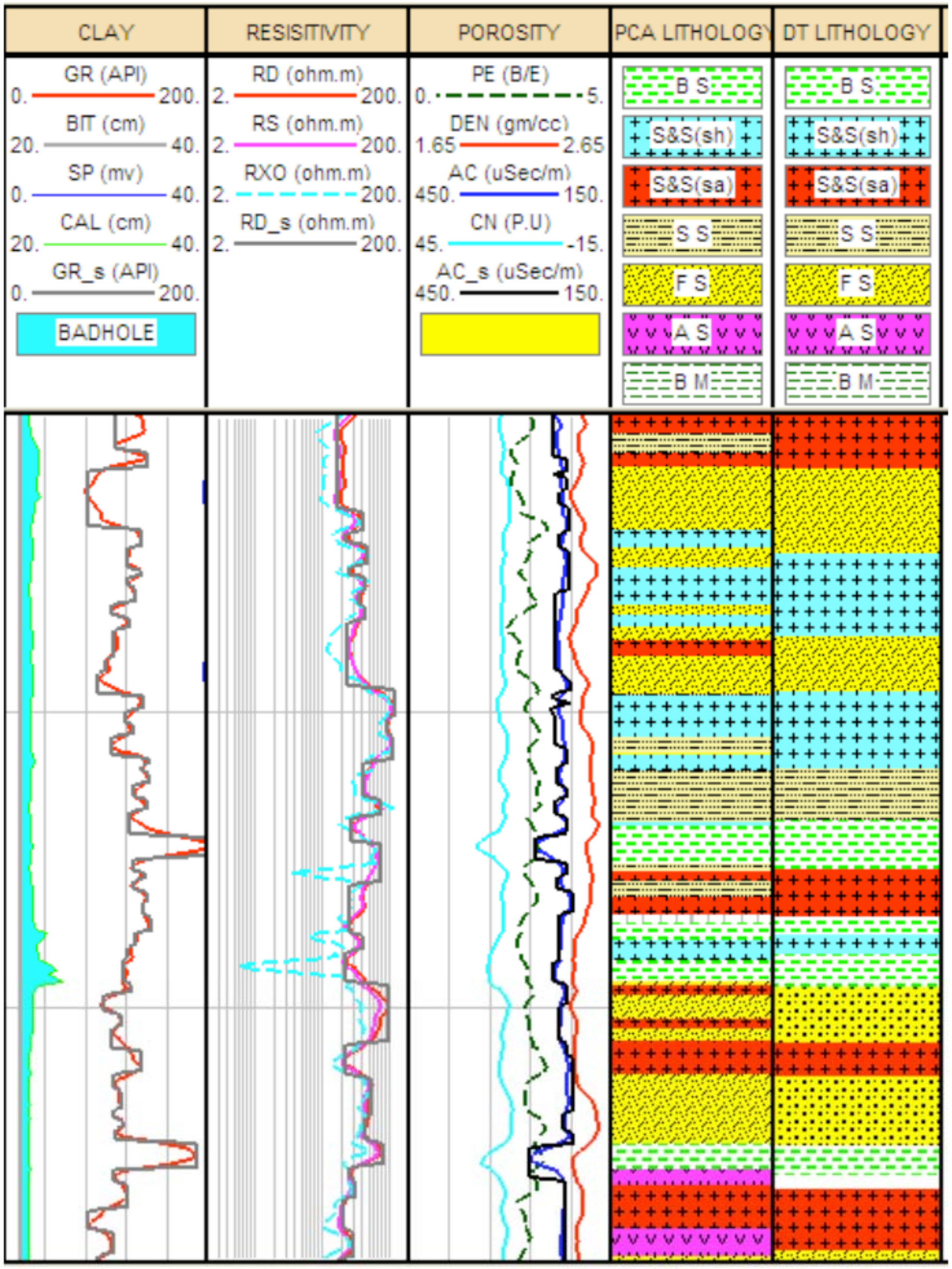


Fig. (10). The lithofacis identification result with PCA and Decision Tree.

was done for each data point by subtracting the curve mean value and dividing by the curve standard deviation. The normalized data were then used to create the principal component curves. The resulting PC curves were calculated from the eigenvectors by taking an input data level in a given well. Using the normalisation data, and then multiplying the normalized curves by the corresponding eigenvalue for the curve, the results are summed. The normalized curve range is shown in Table **3**. The dataset after pre-processing by normalisation had nearly the same range. The proportion of variance of each variable formed the basis for the principal component curves as shown in Table **2**. In the above example 41.1% of the total variability in the data can be seen in the PC 1 curve, the PC 2 curve explained 15.7% thereof, and PC 9 only explained 0.02%. Hence the first eight curves explained 95.7% of the variability. Hence we have practically reduced the information in the 17 curve input to eight curves. The coefficient of the PC curves against each raw curve is shown in Table **4**: PCA values were calculated using these values. Then, DT analysis will be re-used to classify the PC variables dataset. After PCA, the accuracy rose to 90%, and the method promised wide applicability.

4. FIELD EXAMPLE

PCA and DT were applied to data from the Ordos Basin lacustrine deposit shale gas well and the results were displayed in Fig. (**10**). First, the well logs should be inverted before processing to get precise boundary conditions and reduce the effects of instrument resolution limits. The first track was the clay index curves, the second track was the resistivity curves, the third track was porosity, the fourth track was the PCA lithofacies identified result, and the fifth track was the DT-identified result. From the lithofacies identification result, it may be seen that PCA was better than DT and other methods in this shale gas deposit. At the bottom of the well, the DT method showed a dead zone (indicated in white): the PCA was more precise and of higher accuracy than the DT method.

CONCLUSION

With improvements in shale gas reservoir development, engineers need more accurate techniques to overcome the limitations of relying on a large analytical laboratory: logging data therefore play an important role. Using the diversity of logging data from different lithofacies, the proposed method can accurately identify formation rock facies and may pave the way for future shale gas reservoir studies.

Using conventional well-logging curves the method can calculate the formation parameters. For example, reflecting the organic matter content of the formation of $\Delta \lg R$, and the shale content were possible. By comparing multiple parameters, it was found that the principal component analysis method for lithological facies recognition was more effective. Removing the logging curves that reflected the characteristics of the formation of repeatability and obtaining the principal component vectors reflecting the characteristics of the formation, the shale gas reservoir lithology was eventually determined.

CONFLICT OF INTEREST

The authors confirm that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

This study was supported by MOST Special Fund from the State Key Laboratory of Continental Dynamics at Northwest University in China, and NCFC (Natural Science Foundation Of China) (41172122) and acknowledges the Donors of the Shaanxi Province Specialized Research Fund of Higher Education (grant No. 14JK1740), and National High Technology Research and Development Program 863 (2012AA050103). We would also like to thank Dr. Hua Wang for providing critical reviews of the manuscript.

REFERENCES

[1] C.F. Jiang, Y.Q. Cheng, and B.J. Fan, “Progress in and challenges to geologic research of terrestrial shale in China: A case study from the 7th member of the Upper Triassic Yanchang Fm in the Yanchang exploration block, Ordos Basin”, *Natural Gas Industry*, vol. 34, pp. 27-33, 2014.

[2] C.A. Popielski, Z. Heidari, and C. Torres-Verdín, “Rock Classification from Conventional Well Logs in Hydrocarbon-Bearing Shale”, *Paper SPE 159255 Presented at the SPE ATCE*, San Antonio, Texas, USA, 2012.

[3] A. Aranibar, M. Saneifar, and Z. Heidari, “Petrophysical Rock Typing in Organic-Rich Source Rocks Using Well Logs”, *Paper URTeC 1619574 Presented at the Unconventional Resources Technology Conference*, Denver, Colorado, USA, pp. 14, 2013.

[4] J. Quirein, J. Witkowsky, J. Truax, and J. Galford, “Integrating Core Data and Wireline Geochemical Data for Formation Evaluation and Characterization of Shale Gas Reservoirs”, *Paper SPE 134559 Presented at the SPE ATCE*, Florence, Italy, pp. 19-22, 2010.

[5] M. Rezaei, and B. Movahed, “Lithofacies prediction and permeability values estimation from conventional well-logs applying fuzzy logi case study: Alwyn NorthField”, In: *Proceedings of 19th World Petroleum Congress*, Spain, 2008.

[6] E.A. El-sebakhy, O. Asparouhov, and A. Abdulraheem, “Data mining in identifying carbonate litho-facies from well logs based from extreme learning and support vector machines”, In: *Proceeding of AAPG GEO 2010 Middle East Geoscience Conference & Exhibition*, 2010, pp. 1-17.

[7] V. Berteig, J. Helgeland, and E. Mohn, “Lithofacies prediction from welldata”, In: *Proceedings of SPWLA Twenty-Sixth Annual Logging Symposium*, 1985.

[8] R.L. Liu, C.D. Zhou, and Z.W. Jin, “Lithofacies sequence recognition from well logs using time-delay neural networks”, In: *Proceedings of SPWLA 33rd Annual Logging Symposium*, 1992.

[9] S. Q. Wang, “Shale gas exploration and appraisal in China: Problems and discussion”, *Natural Gas Industry*, vol. 33, pp. 13-29, 2013.

[10] S.L. Xu, S.J. Bao, “Preliminary analysis of shale gas resource potential and favorable areas in ordos basin”, *Natural Gas Geoscience*, vol. 20, pp. 460-465, 2009.

[11] J. O. Amaefule, M. Altunbay, and D. Tiab, “Enhanced Reservoir Description: Using Core and Log Data to Identify Hydraulic (Flow) Units and Predict Permeability in Uncored Intervals/Wells”, *Paper SPE 26436 Presented at the SPE ATCE*, Houston, Texas, USA, pp. 3-6, 1993.

[12] J.C. Zhang, R. K. Bian, and T .Y. Jing, “Fundamental significance of gas shale theoretical research”, *Geological Bulletin of China*, vol. 30, pp. 318-323, 2011.

[13] J.C. Zhang, Z.J. Jin, and M. S. Yuan, “Reservoiring mechanism of shale gas and its distribution”, *Natural Gas Industry*, vol. 24, pp. 15-18, 2009.

[14] J.C. Zhang, S.L. Jiang, and X. Tang, “Accumulation types and resources characteristics of shale gas in China”, *Natural Gas Industry*, vol. 29, pp. 109-114, 2004.

[15] P. Tan, M. Steinbach, and V. Kumar, *Introduction to Data Mining*, Addison Wesley: USA, 2005.

[16] J.W. Han, and M. Kamber, *Data Mining Concepts and Techniques*, 2nd ed., China Machine Press: Beijing, 2006.

[17] I.H. Witten, and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, 2nd ed., Morgan Kaufmann: USA 2005.

[18] F. Aminzadeh, “Applications of AI and soft computing for challenging problems in the oil industry”, *Journal of Petroleum Science and Engineering*, vol. 47, pp. 5-14, 2005.

[19] H.G. Liu, L. Yuan, and J.W. Han, “Methods for mining frequent items in data streams: an overview”, *Knowledge and Information Systems*, vol. 26, pp. 1-30, 2011.

[20] B.E. Shie, H.F. Hsiao, and V.S. Tseng, “Mining high utility mobile sequential patterns in mobile commerce environments”, In: *Conference: Database Systems for Advanced Applications - DASFAA*, 2011, pp. 224-23.

[21] K. Bhaduri, R. Wolff, and C. Giannella, “Distributed decision-tree induction in peer-to-peer systems”, *Statistical Analysis and Data Mining*. vol. 1, pp. 85-103, 2008.

[22] M. Kantarcioglu, B. Xi, and C. Clifton, “Classifier evaluation and attribute selection against active adversaries”, *Data Mining and Knowledge Discovery DATAMINE*, vol. 22, pp. 291-335, 2011.

Received: October 18, 2014 Revised: January 15, 2015 Accepted: June 23, 2015