Petroleum Science 22 (2025) 42–54

Contents lists available at ScienceDirect

Petroleum Science

journal homepage: www.keaipublishing.com/en/journals/petroleum-science

Original Paper

Machine learning-based grayscale analyses for lithofacies identification of the Shahejie formation, Bohai Bay Basin, China

Yu-Fan Wang [a], Shang Xu [a, *], Fang Hao [a], Hui-Min Liu [b], Qin-Hong Hu [a], Ke-Lai Xi [a], Dong Yang [a]

[a] National Key Laboratory of Deep Oil and Gas, China University of Petroleum (East China), Qingdao, 266580, Shandong, China
[b] Shengli Oilfield Branch Company, SINOPEC, Dongying, 257015, Shandong, China

ARTICLE INFO

Article history:
Received 23 February 2024
Received in revised form
4 June 2024
Accepted 23 July 2024
Available online 24 July 2024

Edited by Jie Hao and Meng-Jiao Zhou

Keywords:
Shale
Machine learning
Absolute grayscale
Relative amplitude
Grayscale phase model
Lithofacies identification

ABSTRACT

It is of great significance to accurately and rapidly identify shale lithofacies in relation to the evaluation and prediction of sweet spots for shale oil and gas reservoirs. To address the problem of low resolution in logging curves, this study establishes a grayscale-phase model based on high-resolution grayscale curves using clustering analysis algorithms for shale lithofacies identification, working with the Shahejie Formation, Bohai Bay Basin, China. The grayscale phase is defined as the sum of absolute grayscale and relative amplitude as well as their features. The absolute grayscale is the absolute magnitude of the gray values and is utilized for evaluating the material composition (mineral composition + total organic carbon) of shale, while the relative amplitude is the difference between adjacent gray values and is used to identify the shale structure type. The research results show that the grayscale phase model can identify shale lithofacies well, and the accuracy and applicability of this model were verified by the fitting relationship between absolute grayscale and shale mineral composition, as well as corresponding relationships between relative amplitudes and laminae development in shales. Four lithofacies are identified in the target layer of the study area: massive mixed shale, laminated mixed shale, massive calcareous shale and laminated calcareous shale. This method can not only effectively characterize the material composition of shale, but also numerically characterize the development degree of shale laminae, and solve the problem that difficult to identify millimeter-scale laminae based on logging curves, which can provide technical support for shale lithofacies identification, sweet spot evaluation and prediction of complex continental lacustrine basins.

1. Introduction

With the successful development of shale oil and gas in the United States, the shale oil and gas field has received great attention from geologists around the world (Davies, 2014; Wang et al., 2022). China is rich in resources of terrestrial shale oil, and its exploration and development is of great significance for ensuring the national energy security (Gou et al., 2023). While paying attention to the research achievements and exploration and development results of shale oil in the world, Chinese geologists have initially discussed the formation conditions, occurrence forms, accumulation mechanisms, resource potential and favorable development areas of continental shale oil in China based on its geological conditions (Tang et al., 2021; Hou et al., 2022; Xu et al., 2023). With the deepening of shale oil exploration and development, domestic geologists have found that shale lithofacies has an important impact on the reservoir property, oil content, fracturing characteristics and shale oil fluidity (Wang and Garr, 2013; Zhou et al., 2023; Cui et al., 2022; Gao et al., 2022). Therefore, the classification and identification of shale lithofacies has become one of the important research frontiers in shale oil exploration and evaluation.

To reflect the characteristics of the sedimentary environment and hydrocarbon production capacity of different types of lithofacies, organic carbon content, rock structure and mineral composition are generally used as the basis for the division of shale lithofacies (Iqbal et al., 2021; Williams et al., 2022; Ibad and Padmanabhan, 2022; Pang et al., 2023). However, due to the limited samples and cost of coring, it is impossible to evaluate the

* Corresponding author.
E-mail address: xushang0222@163.com (S. Xu).

https://doi.org/10.1016/j.petsci.2024.07.021

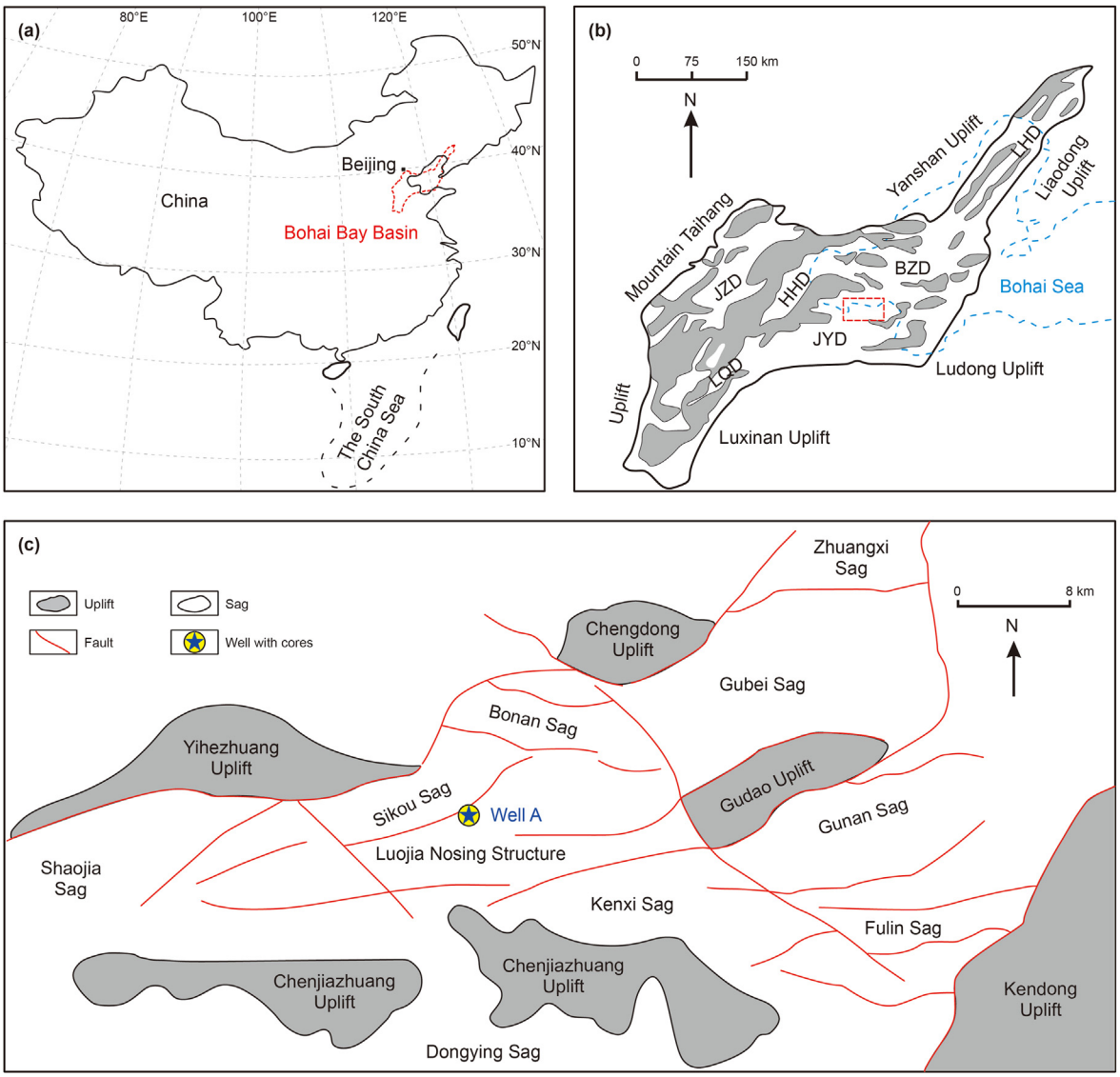

Fig. 1. Map of the study area in the Zhanhua Sag, Bohai Bay Basin. **(a)** Location of Bohai Bay Basin (After Hou et al., 2022). **(b)**–**(c)** Regional structure in Bohai Bay Basin and Zhanhua Sag (After Shi et al., 2019; Li et al., 2017). JZD: Jizhong Depression, LQD: Linqing Depression, HHD: Huanghua Depression, BZD: Bozhong Depression, JYD: Jiyang Depression, LHD: Liaohe Depression.

whole well in a wide production area. Conventional logging data contain rich petrophysical information, which can indirectly classify lithofacies. Therefore, machine learning algorithms based on logging curves to identify lithofacies have been widely presented and applied by geologists and engineers (Gifford and Agah, 2010; Bhattacharya et al., 2016; Koeshidayatullah et al., 2022; Tian et al., 2023). For example, Al-Mudhafar et al. (2019) used the K-means clustering algorithm to successfully identify five carbonate lithofacies developed in Mishrif Formation in West Khulna Oilfield, southern Iraq, based on 11 logging curves such as gamma ray (GR) and spontaneous potential (SP). Xue et al. (2023) predicted shale lithofacies by conventional logging curves based on the XGBoost machine learning model and achieved good application in the Ordos Basin. Recently, Liu et al. (2023) established four machine learning models to identify shale lithofacies in the Luzhou block, southern Sichuan province based on mineral content regression, and concluded that the MT-LightGBM algorithm has good predictive performance for shale lithofacies.

However, most of the existing machine learning algorithms only focus on the identification of mineral composition for shale lithofacies, without a sufficient consideration of shale structure. Because the resolution of the conventional logging curves (0.125 m) limits the accurate identification of millimeter-scale laminae. In this paper, the shale of the lower third member (Es_3L) of the Eocene Shahejie Formation in the Bohai Bay Basin is taken as the research object. Based on the high-resolution (at less than 1 mm) grayscale curve analyses of shale for the logging of well A, a model of grayscale phase is established by the machine learning algorithm to rapidly and efficiently identify shale lithofacies.

2. Geological setting

Located on the east coast of China, the Bohai Bay Basin is divided into six depressions: Linqing Depression, Jiyang Depression, Bozhong Depression, Jizhong Depression, Huanghua Depression and Liaohe Depression (Fig. 1(a) and (b)) (Qi and Yang, 2010; Zhang

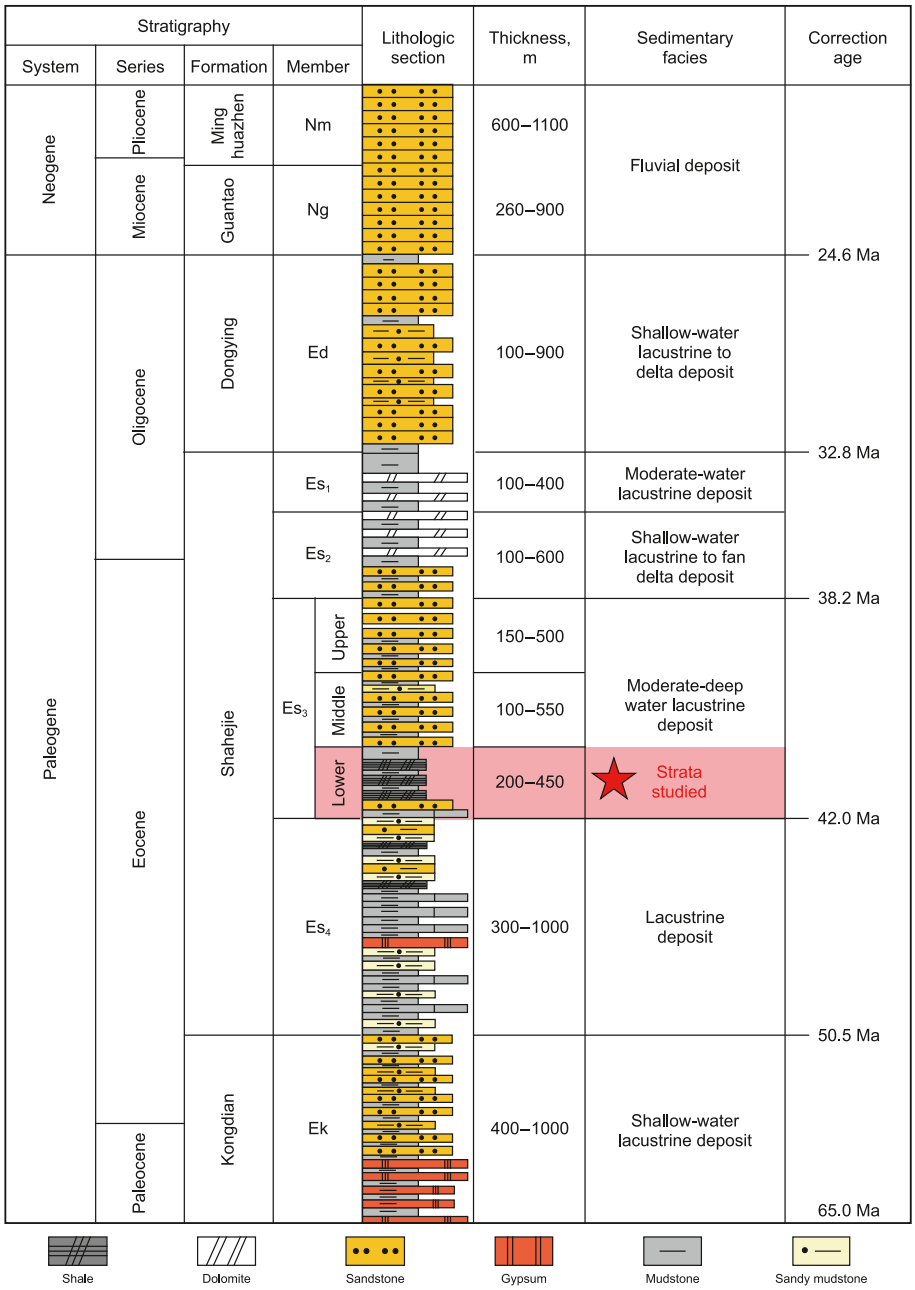

Fig. 2. Generalized stratigraphic column of the Zhanhua Sag (After Li et al., 2017). The study interval is the third member of the Shahejie Formation.

et al., 2020). Zhanhua Sag, located in the northeast of Jiyang Depression, is a third-order tectonic unit in Bohai Bay Basin with an area of about 3610 km^2 (Fig. 1(b)). The sag is bounded by Chengdong and Chengbei Uplift to the north, Chenjiazhuang Uplift to the south, Yihezhuang Uplift to the west, and Kendong Uplift to the southeast (Su et al., 2009; Liu et al., 2016). The research area is mainly concentrated in the Luojia nose structure area of the sag, and the target well is well A (Fig. 1(c)).

Kongdian Formation, Shahejie Formation, Dongying Formation, Guantao Formation and Minghuazhen Formation developed successively from the Paleogene to Neogene strata in Zhanhua Sag (Fig. 2). The Shahejie Formation is widely distributed in graben and half-graben within the sag, and is mostly absent on the tectonic uplift zones separating grabens (Ma et al., 2016). The formation is further divided into four Members from the bottom up: Shahejie Member 4, Shahejie Member 3, Shahejie Member 2 and Shahejie Member 1. The Shahejie Member 3 are subdivided into the upper member, the middle member and the lower member. Among them, the lower member of Shahejie Member 3 is the research object of this paper. During its deposition period, it was in the deep fault depression period of the sag, belonging to a stable and persistent deep-lake to semi-deep lake environment, and was deposited with relatively thick gray and dark gray shale with a thickness between 200 and 450 m, which showed an increasing trend from south to north.

3. Samples and methods

A total of 421 mineralogy analyses, 205 total organic content (TOC, wt%) analyses, and a 194 m length shale grayscale analyses

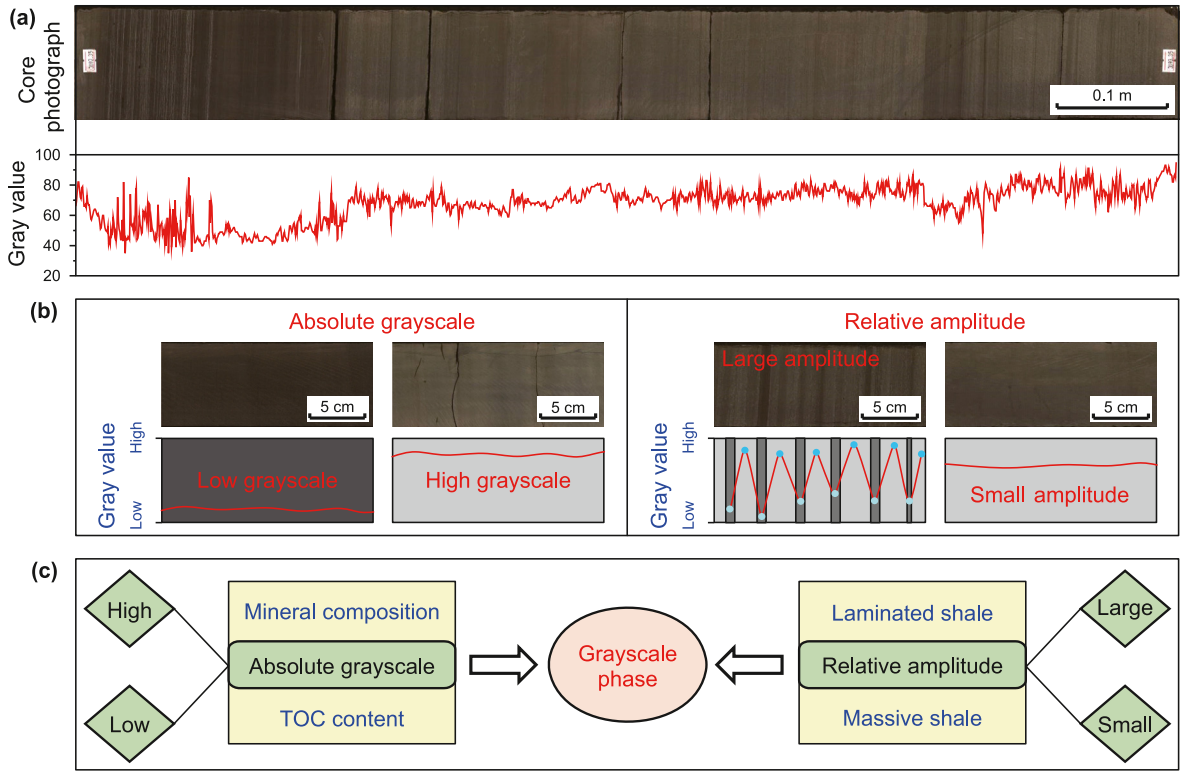

Fig. 3. The model of shale grayscale phase. **(a)** Shale core photos and corresponding grayscale curve characteristics. **(b)** Comparison of grayscale curve characteristics of shale in different intervals, the absolute grayscale represents the extracted gray value, and the relative amplitude represents the difference between adjacent gray values. **(c)** The establishment scheme and connotation of the shale grayscale phase. The sum of absolute grayscale and relative amplitude and their features is defined as the grayscale phase.

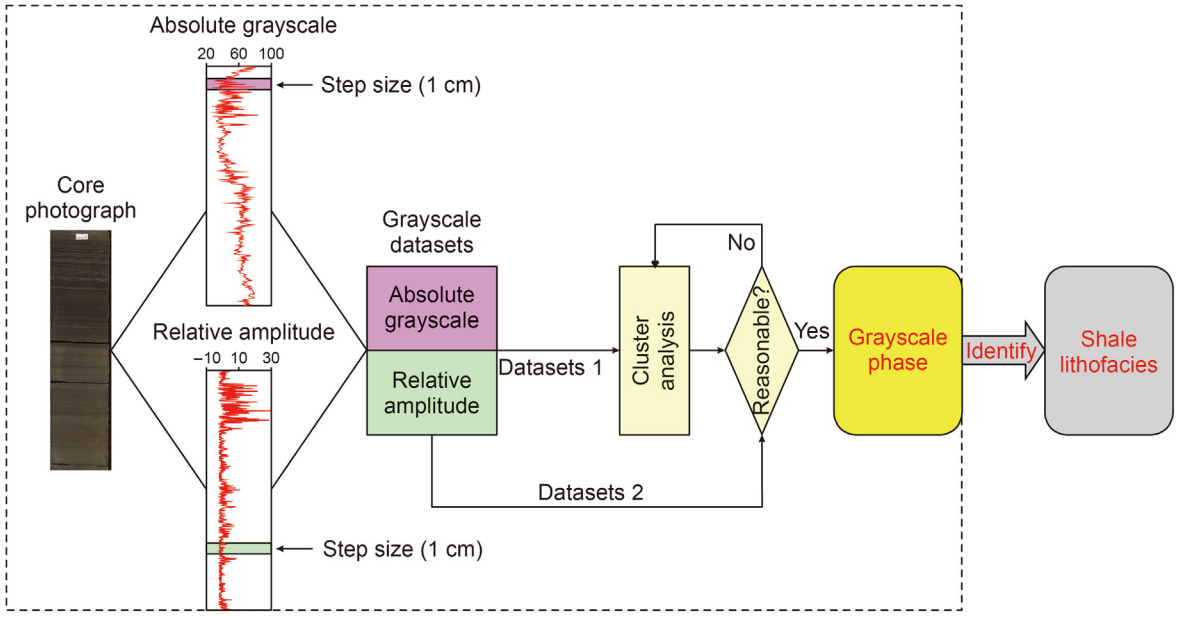

Fig. 4. Evaluation of shale lithofacies implementation process based on machine learning. The difference between gray value and adjacent gray value is defined as absolute grayscale and relative amplitude respectively, and the grayscale phase is the sum of absolute grayscale and relative amplitude and their feature.

were used for this study.

XRD experiments can provide a quantitative estimate of the composition of various minerals in shale (Chen et al., 2011). The mineral composition of the sample was determined by a D/MAX-2600X-ray diffraction instrument. The powdered shale sample of about 100 mg was weighed and mixed with ethanol, ground by hand, and then smearing on the slide for XRD analyses. Finally, the area under the main peak curve of each mineral was calculated and the Lorentz polarization was corrected to determine relative mineral percentages (Gou et al., 2021). This experiment process is

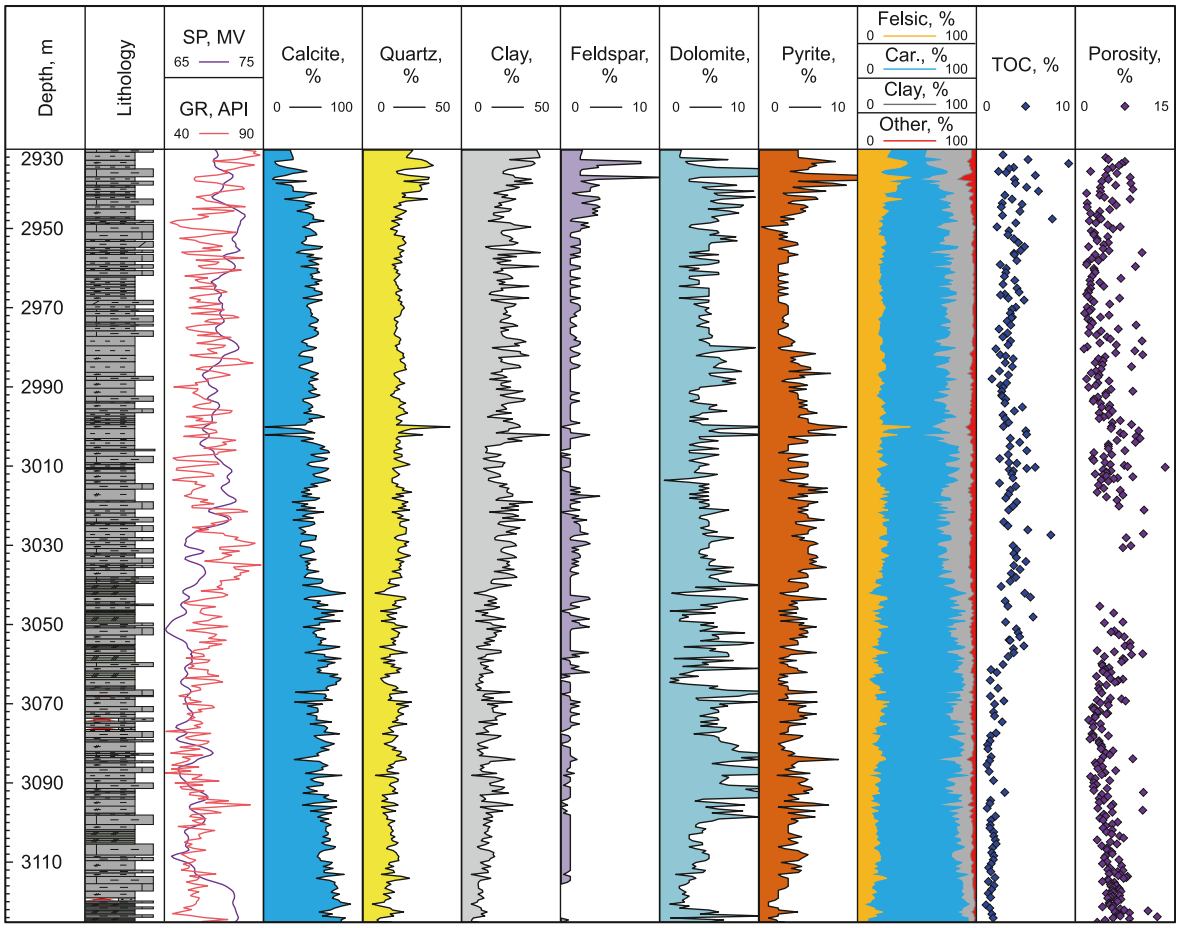

Fig. 5. Comprehensive histogram of well A. Car. = Carbonate.

carried out according to China's oil and gas industry standard SY/T6201-1996.

An organic carbon analyzer (CS-600) was used to determine TOC content. Before the formal test, the shale sample was first crushed into powder, and then the inorganic carbon was removed with diluted hydrochloric acid. The remaining sample was put into the instrument for high-temperature combustion, and the final experimental results were quantified by the analyzer, which is carried out according to the national standard GB/T19145-2003.

From 2011 to 2013, to initially investigate the reservoir performance and production of shale oil and gas in the Jiyang Depression, China Petroleum and Chemical Corporation arranged four systematic core wells to conduct continuous core drilling of the Es_4U member and the Es_3L member. Among them, the core length of well A located in Zhanhua Sag is 221 m, and Shengli Oilfield took high-definition photos of these cores (Li et al., 2017). Through the core photos, we first observed the color, structure and mineral distribution of shale in different layers, and then extracted nearly 500,000 sets of gray value data in Image-J software, and processed the gray data in MATLAB software.

4. The principle of identifying shale lithofacies by machine learning

4.1. Grayscale characteristics of different shale

The gray value reflects the reflected light intensity of a unit pixel, and its value is between 0 (black) and 256 (white). The greater the brightness, the higher the gray level. Areas with different "colors" on the shale core and whether the laminae are developed show completely different responses on the grayscale curve (Fig. 3(a)). The darker the shale color, the smaller the gray value; the lighter the color, the larger the gray value (Fig. 3(b)). At the same time, the laminae with dark and light colors corresponds to the minimum and maximum values of the gray value in a certain interval respectively. When the dark and light laminae alternate, the grayscale curve will reverse abruptly, showing large amplitude fluctuation and obvious zigzag; on the contrary, while the laminae are not developed, the fluctuation amplitude of the grayscale curve is small (Fig. 3(b)). Based on the grayscale curve analyses of shale, this paper defines the difference between gray value and adjacent gray value as absolute grayscale and relative amplitude, respectively. The absolute grayscale is used to measure the absolute size of gray value, while the relative amplitude is employed to analyze the difference between adjacent gray values.

4.2. Model of shale grayscale phase

A grayscale phase is defined as the sum of absolute grayscale and relative amplitude and their features, and both of these two terms have the same or similar features in the same grayscale phase (Fig. 3(c)). In the process of building the grayscale phase model, the relationships between absolute grayscale and shale material composition (inorganic minerals and organic matter), as well as relative amplitude and shale structure, are clarified. Therefore, the grayscale phase model can effectively identify shale lithofacies. In

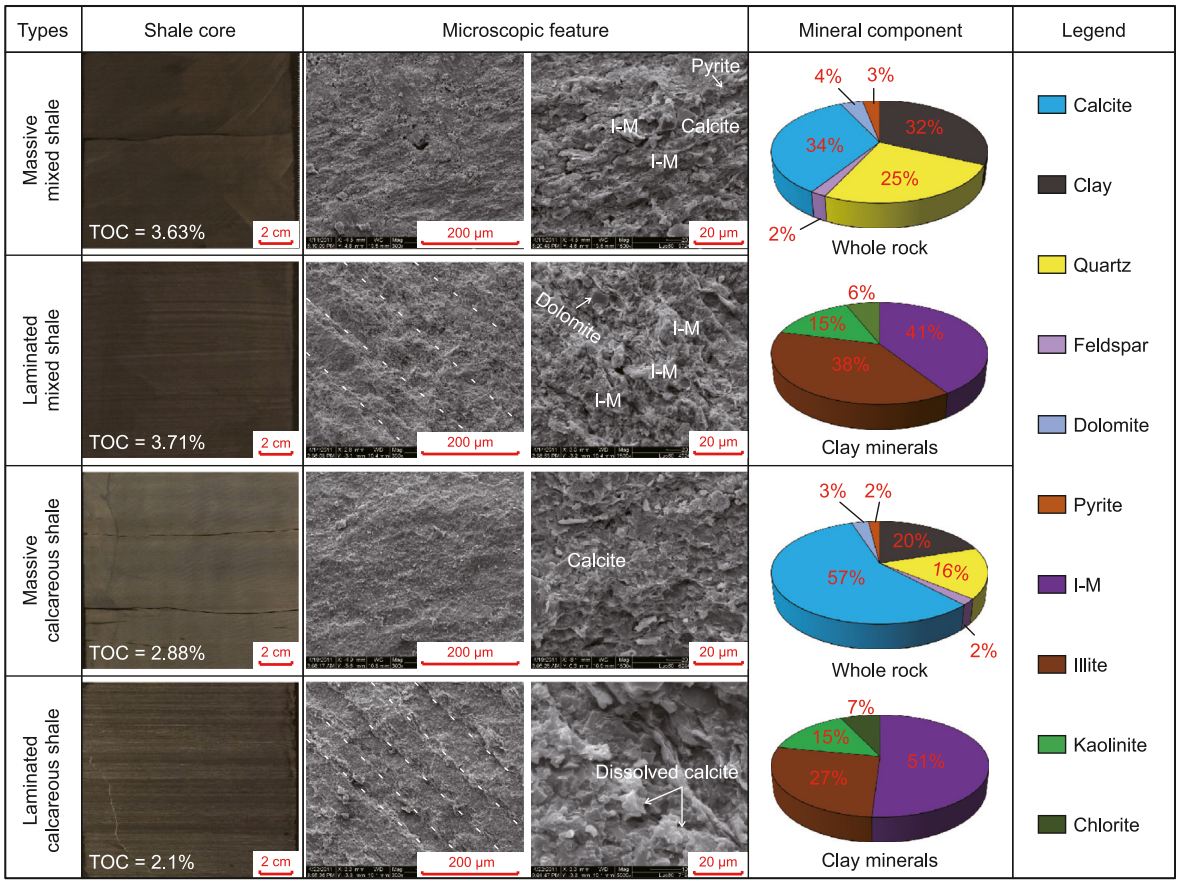

Fig. 6. Basic characteristics of shale structure and mineral composition in the study area. I–M: the mixed layer of illite and the montmorillonite.

terms of mineral composition, the shale in the study area mainly develops calcareous and mixed shales. In terms of shale structure, the shale mainly develops laminated and massive shales. Therefore, we take a boundary value to divide absolute grayscale into high and low grayscales, and take another boundary value to divide relative amplitude value into large and small amplitudes, to obtain a total of four kinds of grayscale phases: low grayscale small amplitude phase, low grayscale large amplitude phase, high grayscale small amplitude phase, and high grayscale large amplitude phase.

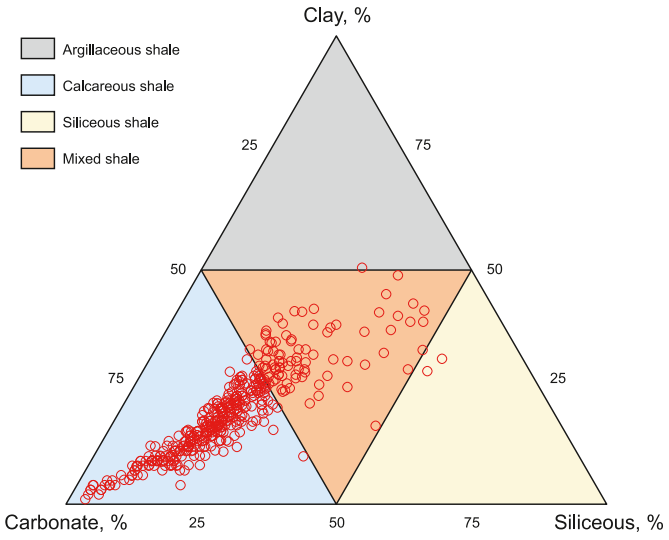

Fig. 7. Ternary mineralogy the Es_3L shale in well A; the target layer of the study area is mainly dominated by calcareous shale and mixed shale.

4.3. Implementation process of shale lithofacies identification

To reduce the error caused by the data themselves, we set the step size for the extraction of absolute grayscale and relative amplitude data series and calculated the average value of all data inside step sizes as the absolute grayscale and relative amplitude of a point, and this process was completed by programming in MATLAB software. The boundary value of absolute grayscale and relative amplitude was determined by a combination of manual and machine learning algorithms. Because the error of boundary values determined manually is too large, we have chosen to use the cluster analysis algorithm to cluster the absolute grayscale and relative amplitude data, and then obtain the boundary values. However, it should be noted that the clustering analysis algorithm is essentially based on the size of the numerical value to classify, and whether the clustering results are reasonable needs to be further verified. Therefore, representative core photos were selected to analyze the distribution characteristics of absolute grayscale and relative amplitude of different shales, and the range of boundary values was initially defined to constrain the results of cluster analyses (Fig. 4). According to the above steps, the grayscale phase model of the Es_3L shale in well A was established, and then the shale lithofacies was identified.

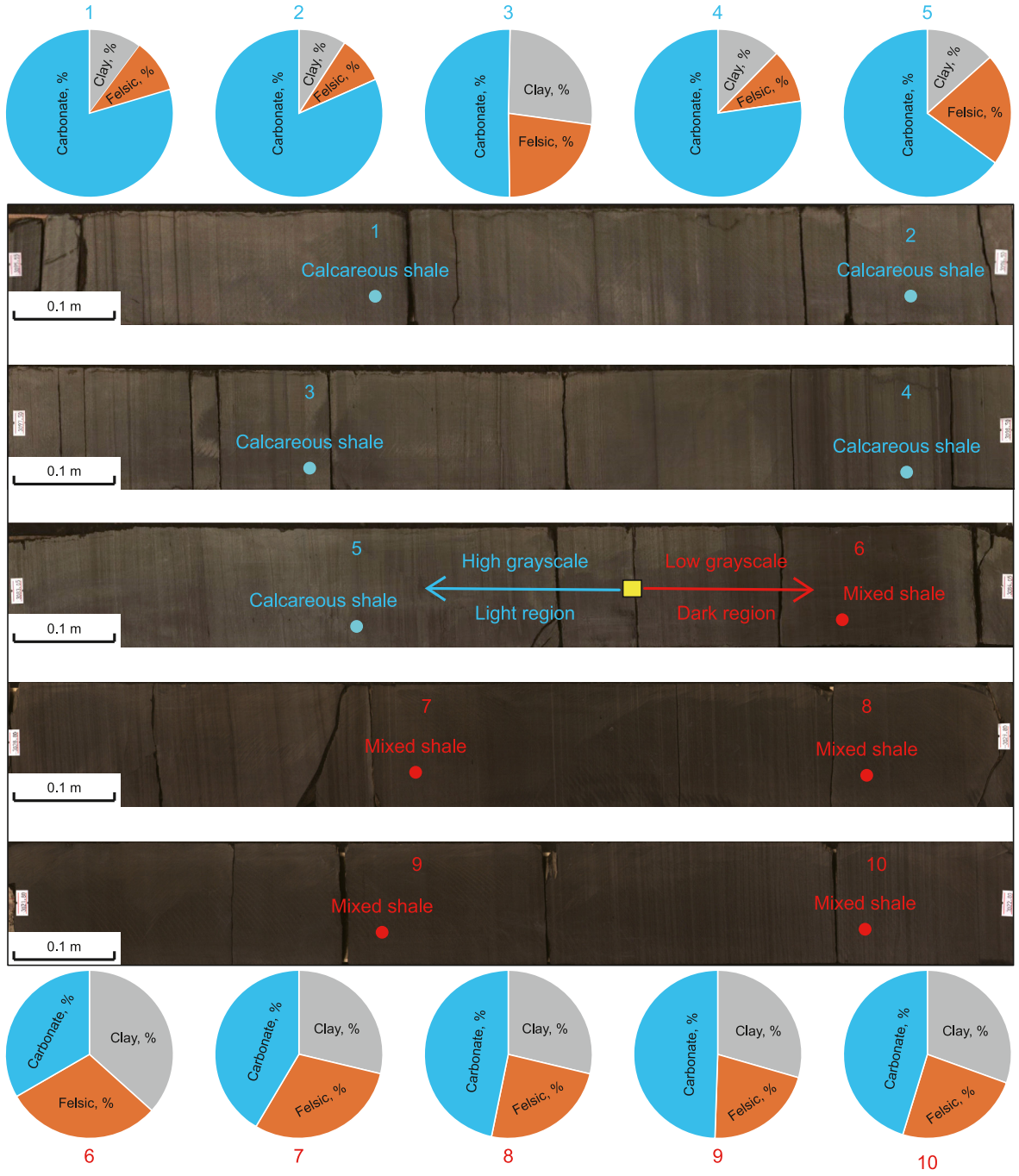

Fig. 8. Mineral composition characteristics in different regions of 1–10; the light-colored area is dominated by calcareous shale with carbonate minerals accounting for more than 50%, while the dark-colored area is dominated by mixed shale with carbonate minerals, clay minerals and felsic minerals accounting for less than 50%.

5. Results and discussion

5.1. Shale characteristics in the study area

X-ray diffraction (XRD) results for 421 samples show that the shale minerals are mainly composed of calcite (averaged at 51.8%), clay minerals (18.9%) and quartz (18%), followed by dolomite (5.8%), pyrite (3.8%) and plagioclase (1.1%), and a few of potassium feldspar (0.3%) and siderite (0.2%) (Fig. 5). According to the previous ternary classification method, shale minerals can be divided into three categories: clay minerals, carbonate minerals (calcite + dolomite + siderite), and felsic terrigenous clastic minerals (quartz + plagioclase + potassium feldspar) (Xu et al., 2020; Cai et al., 2022). The content of carbonate minerals is the highest and varied from 11% to 94% (averaged at 58%), and the contents of clay minerals and felsic terrigenous clastic minerals are 1%–48% (18.9%) and 3%–50% (19.4%), respectively (Fig. 5). The TOC of shale ranges from 0.71 wt% to 9.32 wt% (3.08 wt%) (Fig. 5). In addition, the contents of felsic minerals, clay minerals and TOC in the deep black shale in the study area are higher than those in the light gray shale (Fig. 6).

The subdivision principle of shale structure depends on whether

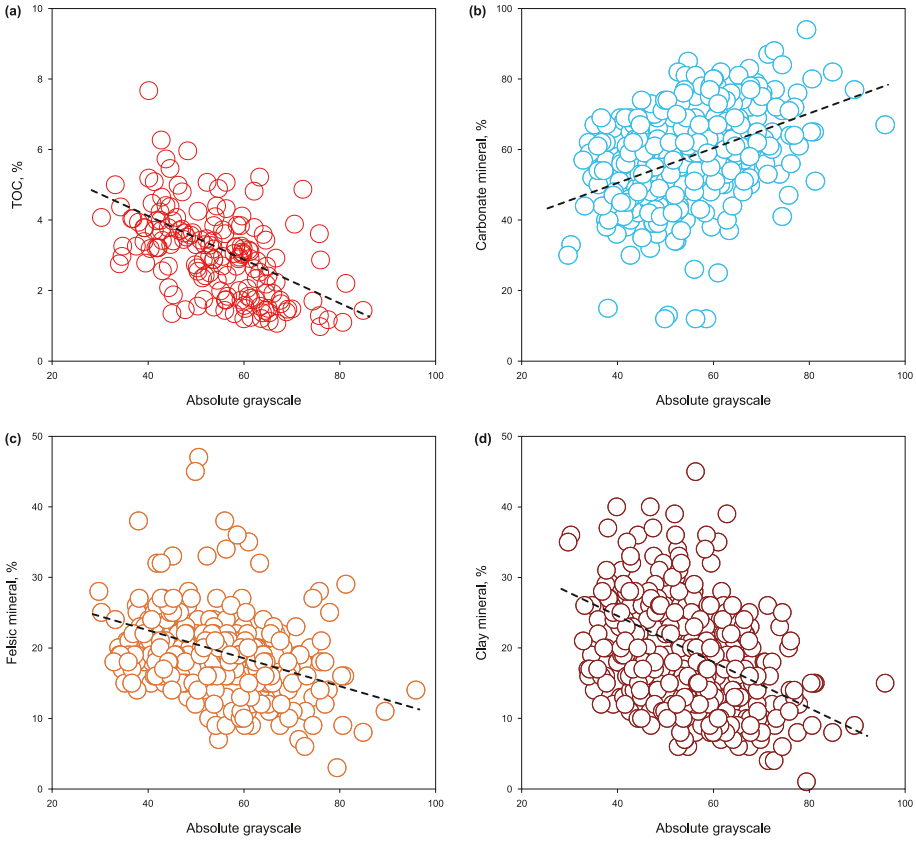

Fig. 9. Relationships between absolute grayscale and shale material composition (TOC, carbonate mineral, felsic mineral and clay mineral).

a bedding is developed. The shale with beddings non-developed is generally named massive shale, and the shale with beddings developed is named laminated shale or layered shale. The main difference between laminated and layered shales is the thickness of the single layer. If the thickness of a single layer is greater than a certain boundary value, it is layered shale, otherwise, it is laminated shale. However, the single-layer thickness boundary that distinguishes laminated from layered shales is not uniform in the literature. For example, Shi et al. (2022) believed that the laminar thickness can reach 3 mm, while Sun et al. (2023) believed that the laminar thickness is less than 1 mm. However, it has also been suggested that the thickness of a lamina varies from a few millimeters to a few centimeters (De Geer et al., 1908; Campbell et al., 1967; Breckenridge et al., 2012; Lazar et al., 2015). Lacustrine shale oil exploration and development practice in China shows that shale with a bedding structure is considered to be the most favorable shale (Su et al., 2018; Li et al., 2021, 2023). Therefore, this paper classifies shale structures in the study area into two categories based on whether beddings are developed: massive structure and laminated structure (Fig. 6). The dark and light-colored laminae are characterized by alternating deposition, mostly in flat shape, and the thickness is generally no more than 10 mm, which is similar to the laminae thickness defined by Campbell et al. (1967).

5.2. Absolute grayscale and mineral composition

The ternary presentation of shale mineral composition shows that the target layer of the study area mainly develops calcareous shale and mixed shale (Fig. 7). We selected representative core photos (core colors ranging from light to dark), and positioned sampling depth points on the core photos to obtain the distribution positions of different types of shale. The results show a clear pattern: the regions with lighter colors mainly develop calcareous shale, while the regions with darker colors mainly exhibit mixed shale (Fig. 8). Previous studies have shown that more clay minerals (with larger specific surface areas) can adsorb more organic matter and form darker organic-rich shales while light-colored shale usually has a low abundance of organic matter (Chen et al., 2016; Blattmann et al., 2019). Our results show that an absolute grayscale is negatively correlated with TOC, clay minerals and felsic mineral content, while positively correlated with carbonate mineral content (Fig. 9), which is consistent with previous understandings. According to the mineral composition characteristics of the study area, the high absolute grayscale area is mainly composed of calcite minerals (with little dolomite content), which indicates calcareous shale, while the low absolute grayscale area has relatively high organic matter, clay minerals and felsic minerals content, indicating mixed shale.

5.3. Relative amplitude and shale structure

Dark and light laminae are usually deposited alternately, which makes it feasible to analyze shale laminae with grayscale. The dark and light laminae correspond to the minimum and maximum of the gray value of a certain interval, respectively. Therefore, the fluctuation amplitude of the grayscale curve is usually large in the laminae-developed area, while the fluctuation amplitude is relatively small in the non-developed area (Fig. 10). Considering the frequent alternating deposition of dark and light-colored laminae and possible contamination stains in the process of core preservation, we took the sampling point as the center and calculated the average relative amplitude within three kinds of step sizes (0.6 cm,

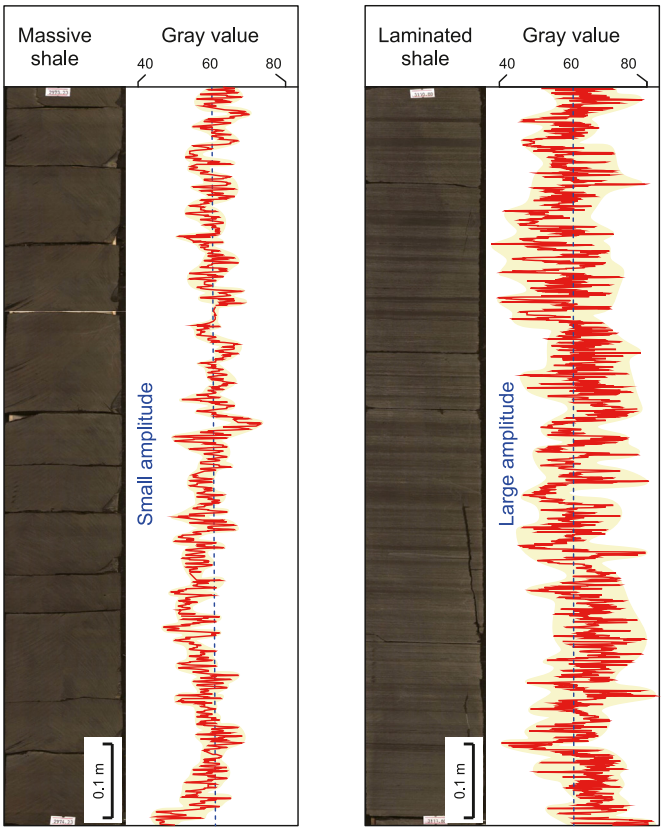

Fig. 10. Grayscale curve characteristics of shale with different structures; on the grayscale curve, the fluctuation amplitude of laminated shale is large, while that of massive shale is small. The width of the yellow shadow represents the amplitude of the fluctuation of the grayscale curve.

1 cm, 2 cm) for comprehensive analyses, so as to reduce the errors that may be caused. The calculation results of the relative amplitude of these three kinds of step sizes are basically the same, but the laminae thickness in the study area does not exceed 10 mm; therefore, we chose the calculation results of the "1 cm step size" scheme (Fig. 11).

Fig. 11 shows the whole interval's relative amplitude distribution characteristics and the corresponding part of the shale core, and the size of the relative amplitude and whether the laminae are developed has a good corresponding relation. The large relative amplitude value is mainly concentrated in the lower part of well A, which indicates that the lower part is extensively developed with laminated shale, while the upper part is mainly developed with massive shale, which is consistent with the understanding of Ma et al. (2016). It is worth noting that the presence of "scratches" during the coring process can lead to human error that the "dark and light" spacing are distributed and the laminae do not develop. Therefore, theoretically speaking, as long as the core surface is clean, the relative amplitude has a good result in laminae identification; therefore, this parameter has a certain reference significance for the identification and characterization of laminae.

5.4. Determination of boundary values

Through the above analyses of absolute grayscale and shale mineral composition, relative amplitude and shale structure, it can be seen that high absolute grayscale indicates calcareous shale, low absolute grayscale suggests mixed shale, large relative amplitude links with laminated shale, and small relative amplitude points to massive shale. However, the boundary values used to distinguish different shales are not clear. Considering the large error of manually setting, we use cluster analysis algorithm to determine the boundary values. The cluster analysis results show that the cluster centers of absolute grayscale data sequences are 47 and 65, respectively (Fig. 12(a)). After the data series are reordered, the boundary value is 56, between 50 and 60 (Fig. 12(b)), which is judged to be reasonable. Therefore, the value of 56 is used as the boundary value to distinguish mixed and calcareous shales; that is, the absolute grayscale less than 56 indicates mixed shale, and the absolute grayscale greater than 56 indicates calcareous shale. The clustering centers of the relative amplitude data series are 1.39 and

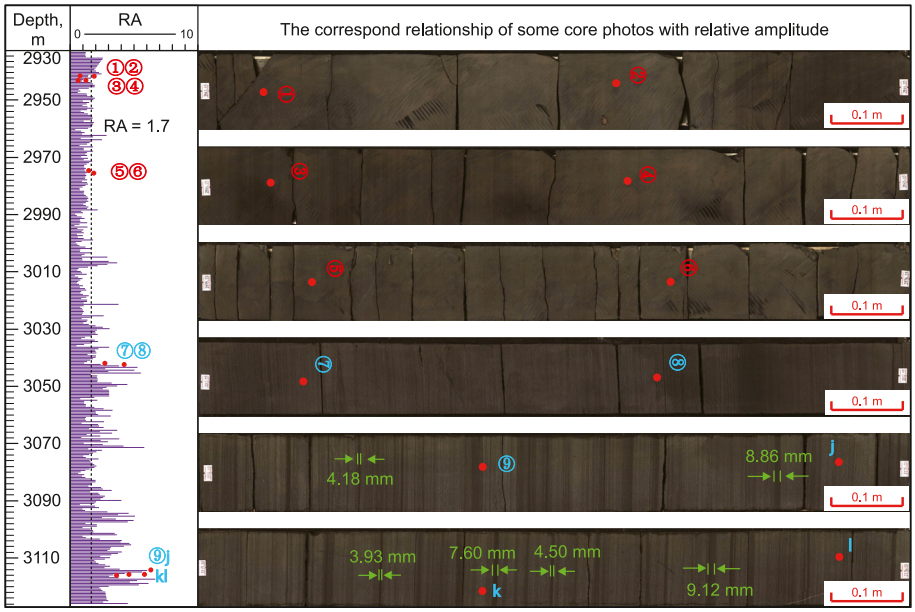

Fig. 11. Relationship between relative amplitude and shale laminae. RA is the relative amplitude, and when the relative amplitude is greater than 1.7, laminae develop on the core; if the relative amplitude is less than 1.7, the laminae on the core is not developed.

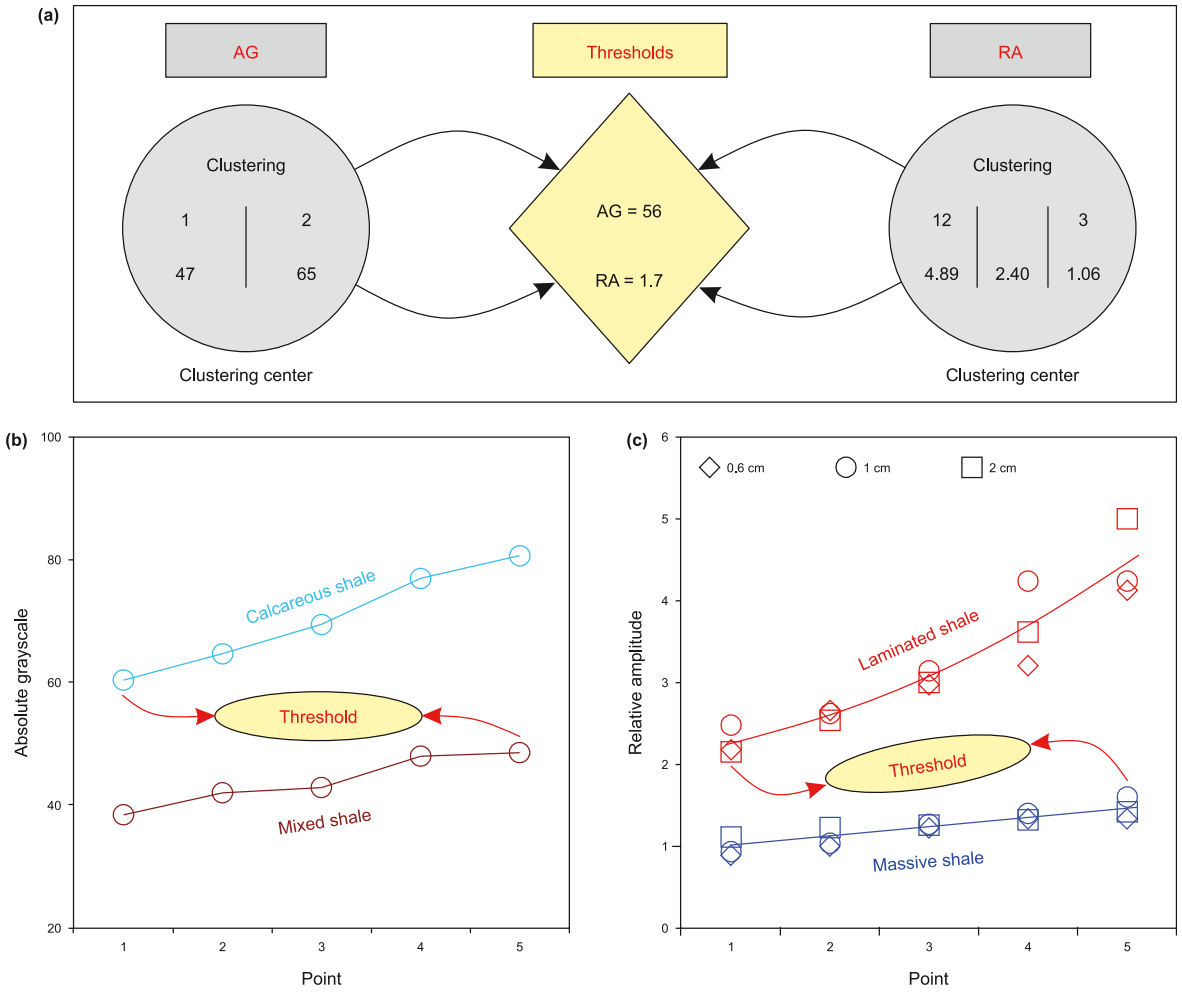

Fig. 12. Determination of absolute grayscale and relative amplitude thresholds. **(a)** The output of absolute grayscale and relative amplitude data sequence cluster analysis. AG = absolute grayscale, RA = relative amplitude. **(b)**–**(c)** Distribution characteristics of absolute grayscale and relative amplitude of different shales. Three kinds of step sizes (0.6 cm, 1 cm, 2 cm) are used to calculate the relative amplitude.

3.99, respectively, and the boundary value is 2.3 after the data series are reordered. However, the value 2.3 does not belong to [1.5, 2], so it is judged that the boundary value is unreasonable. The relative amplitude data series were re-clustered into three categories with the boundary values of 1.7 and 3.4, respectively, with 1.7 within the interval of the boundary values of the relative amplitude (Fig. 12(a)–(c)). Finally, a value of 1.7 is interpreted as the boundary value of relative amplitude, that is, a relative amplitude less than 1.7 indicates massive shale, and a relative amplitude greater than 1.7 indicates laminated shale.

5.5. Fast and efficient identification of shale lithofacies

Following Section 4.2, the grayscale phase with absolute grayscale less than 56 and relative amplitude less than 1.7 is defined as low grayscale small amplitude phase. The grayscale phase with absolute grayscale less than 56 and relative amplitude greater than 1.7 is defined as low grayscale large amplitude phase. The grayscale phase with absolute grayscale greater than 56 and relative amplitude less than 1.7 is defined as high grayscale small amplitude phase. The grayscale phase whose absolute grayscale is greater than 56 and relative amplitude is greater than 1.7 is defined as the high gray amplitude phase. The corresponding relationship between grayscale phase and shale lithofacies is as follows: low grayscale small amplitude phase corresponds to massive mixed shale, low grayscale large amplitude phase corresponds to laminated mixed shale, high grayscale small amplitude phase corresponds to massive calcareous shale, and high grayscale large amplitude phase corresponds to laminated calcareous shale (Fig. 13).

Shale lithofacies are continuously identified using the grayscale phase. The results show that the rapid phase transformation of shale lithofacies caused the same lithofacies unable to develop continuously and the thickness was small during the deposition period of the Es_3L shale in well A (Fig. 13). The mineral composition and structure of shales have changed significantly during the deposition period of the Es_3L shale in well A. From the perspective of mineral composition, the proportion of clay minerals and felsic minerals gradually increased and then decreased, so the shale type, in general, changed from mainly calcareous shale to mainly mixed shale and then to mainly calcareous shale. From the perspective of rock structure, the shale structure had undergone a transition from laminated structure to massive structure, which indicates that the water energy has undergone a process from weak to strong during deposition period of the Es_3L shale (Zolitschka et al., 2015).

5.6. Application prospect of the grayscale phase model

At present, the focus in China's oil and gas exploration and

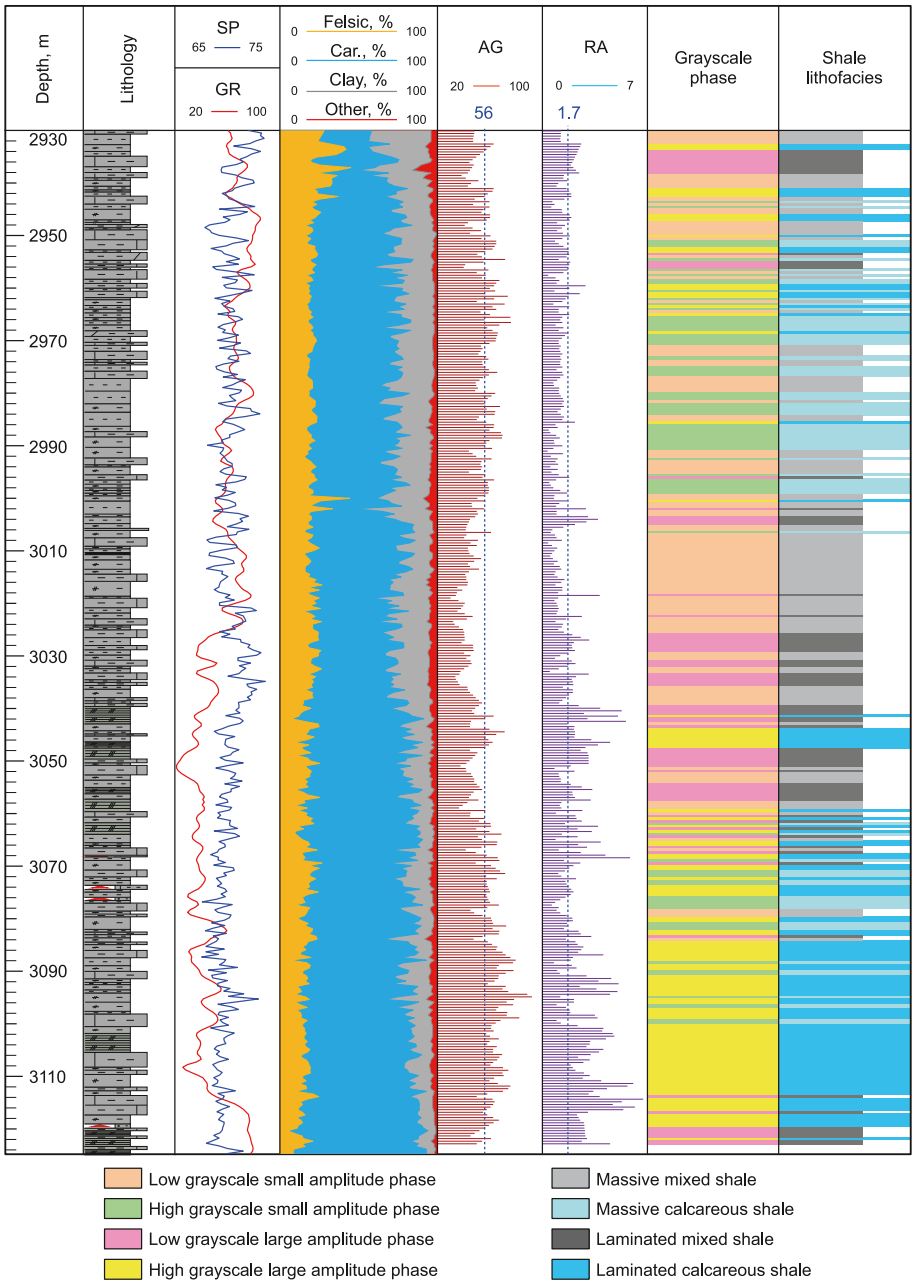

Fig. 13. Characteristics of shale lithofacies association of the Es_3L member in well A. The presence of mixed shale is indicated by an absolute grayscale below 56, while an absolute grayscale (AG) above 56 indicates the presence of calcareous shale; while the relative amplitude (RA) is greater than 1.7, laminae develop on the core; if the relative amplitude is less than 1.7, the laminae on the core is not developed.

development shifts towards unconventional areas, particularly the complex and rapid phase transition of shale lithofacies, necessitates techniques of high resolution and reliability in classifying various lithofacies. In comparison to logging curves, the grayscale curve can achieve a resolution as fine as millimeters (less than 0.1 mm, depending on the clarity of core photos), making it an excellent proxy for studying lacustrine shale formations with rapid phase changes. Additionally, the intrinsic relationship between shale grayscale and TOC content, mineral composition, and rock structure demonstrates that the grayscale curve effectively records these crucial pieces of information for lithofacies classification purposes. Therefore, when dealing with wells that have limited core samples or lack lithofacies data in certain blocks, the grayscale phase model can be utilized to quickly and efficiently classify shale lithofacies. Furthermore, petroleum geologists in China have recognized laminated shale as the most favorable type for exploration and development in lacustrine shale formations. The relative amplitude parameter within the grayscale phase model is capable of accurately identifying laminated intervals within shales; overall, this not only resolves difficulties associated with numerically characterizing millimeter-scale laminae but also enables precise identification of intervals conducive to shale oil development.

6. Conclusion

(1) Based on the grayscale curve analyses of shale, two important parameters are proposed: absolute grayscale and relative amplitude. The absolute grayscale is used to evaluate the

shale material composition (mineral composition + TOC), and the relative amplitude is used to indicate the shale structural type.

(2) Based on the clustering analysis algorithm, the boundary values that can distinguish shale mineral composition and shale structure are clarified with the absolute grayscale of 56 and the relative amplitude of 1.7. Four lithofacies are identified by the model of the grayscale phase in the study area: massive mixed shale, laminated mixed shale, massive calcareous shale and laminated calcareous shale.

(3) The model of grayscale phase can not only effectively characterize the material composition of shale, but also solve the problem that it is difficult to identify millimeter-scale laminae. Therefore, the grayscale phase model is of great significance to the lithofacies identification, sweet spot evaluation and prediction of complex continental lacustrine basins.

CRediT authorship contribution statement

Yu-Fan Wang: Writing – original draft, Software, Methodology, Conceptualization. **Shang Xu:** Validation, Methodology, Data curation. **Fang Hao:** Supervision. **Hui-Min Liu:** Writing – review & editing. **Qin-Hong Hu:** Writing – review & editing. **Ke-Lai Xi:** Validation. **Dong Yang:** Validation.

Declaration of Competing interest

We declare that we have no financial and personal relationships with other people or organizations that can inappropriately influence our work, there is no professional or other personal interest of any nature or kind in any product, service and/or company that could be construed as influencing the position presented in the manuscript entitled.

Acknowledgements

This research was supported by the National Natural Science Foundation of China (42122017, 41821002), and the Independent Innovation Research Program of China University of Petroleum (East China) (21CX06001A).

References

AlMudhafar, W.J., Al Lawe, E.M., Noshi, C.I., 2019. Clustering analysis for improved characterization of carbonate reservoirs in a Southern Iraqi Oil Field. In: Offshore Technology Conference. Society of Petroleum Engineers, Houston, Texas.

Bhattacharya, S., Carr, T.R., Pal, M., 2016. Comparison of supervised and unsupervised approaches for mudstone lithofacies classification: case studies from the Bakken and Mahantango-Marcellus Shale, USA. J. Nat. Gas Sci. Eng. 33, 1119–1133. https://doi.org/10.1016/j.jngse.2016.04.055.

Blattmann, T.M., Liu, Z., Zhang, Y., et al., 2019. Mineralogical control on the fate of continentally derived organic matter in the ocean. Science 366 (6466), 742–745. https://doi.org/10.1126/science.aax5345.

Breckenridge, A., Lowell, T.V., Stroup, J.S., et al., 2012. A review and analysis of varve thickness records from glacial Lake Ojibway (Ontario and Quebec, Canada). Quat. Int. 260, 43–54. https://doi.org/10.1016/j.quaint.2011.09.031.

Cai, Q.S., Hu, M.Y., Zhang, B.M., et al., 2022. Source of silica and its implications for organic matter enrichment in the Upper Ordovician-Lower Silurian black shale in western Hubei Province, China: Insights from geochemical and petrological analysis. Pet. Sci. 19 (1), 74–90. https://doi.org/10.1016/j.petsci.2021.10.012.

Campbell, C.V., 1967. Lamina, laminaset, bed and bedset. Sedimentology 8 (1), 7–26. https://doi.org/10.1111/j.1365-3091.1967.tb01301.x.

Chen, L., Jenkyns, H.C., Xu, G.W., et al., 2016. Preliminary nannofossil and geochemical data from Jurassic black shales from the Qiangtang Basin, northern Tibet. J. Asian Earth Sci. 115, 257–267. https://doi.org/10.1016/j.jseaes.2015.10.004.

Chen, S.B., Zhu, Y.M., Wang, H., et al., 2011. Shale gas reservoir characterisation: a typical case in the southern Sichuan Basin of China. Energy 36 (11), 6609–6616. https://doi.org/10.1016/j.energy.2011.09.001.

Cui, Q.Y., Yang, H.F., Li, X.Q., et al., 2022. Identification of lithofacies and prediction of mineral composition in shales—A case study of the Shahejie Formation in the Bozhong Sag. Unconventional Resources 2, 72–84. https://doi.org/10.1016/j.uncres.2022.09.002.

Davies, R.J., Almond, S., Ward, R.S., et al., 2014. Oil and gas wells and their integrity: implications for shale and unconventional resource exploitation. Mar. Petrol. Geol. 59, 674–675. https://doi.org/10.1016/j.marpetgeo.2014.03.001.

De Geer, G., 1908. On Late Quaternary time and climate. GFF 30 (7), 459–464. https://doi.org/10.1080/11035890809445600.

Gao, P., Xiao, X., Hu, D., et al., 2022. Effect of silica diagenesis on porosity evolution of deep gas shale reservoir of the Lower Paleozoic Wufeng-Longmaxi formations, Sichuan Basin. Mar. Petrol. Geol. 145, 105873. https://doi.org/10.1016/j.marpetgeo.2022.105873.

Gifford, C.M., Agah, A., 2010. Collaborative multi-agent rock facies classification from wireline well log data. Eng. Appl. Artif. Intell. 23 (7), 1158–1172. https://doi.org/10.1016/j.engappai.2010.02.004.

Gou, Q.Y., Xu, S., Hao, F., et al., 2023. Petrography and mineralogy control the nm-μm-scale pore structure of saline lacustrine carbonate-rich shales from the Jianghan Basin. China. Mar. Petrol. Geol. 155, 106399. https://doi.org/10.1016/j.marpetgeo.2023.106399.

Gou, Q.Y., Xu, S., Hao, F., et al., 2021. The effect of tectonic deformation and preservation condition on the shale pore structure using adsorption-based textural quantification and 3D image observation. Energy 219, 119579. https://doi.org/10.1016/j.energy.2020.119579.

Hou, L.H., Luo, X., Lin, S.H., et al., 2022. Assessment of recoverable oil and gas resources by in-situ conversion of shale—case study of extracting the Chang 73 shale in the Ordos Basin. Pet. Sci. 19 (2), 441–458. https://doi.org/10.1016/j.petsci.2021.10.015.

Ibad, S.M., Padmanabhan, E., 2022. Lithofacies, mineralogy, and pore types in Paleozoic gas shales from Western Peninsular Malaysia. J. Pet. Sci. Eng. 212, 110239. https://doi.org/10.1016/j.petrol.2022.110239.

Iqbal, M.A., Rezaee, R., Smith, G., et al., 2021. Shale lithofacies controls on porosity and pore structure: an example from ordovician goldwyer formation, canning basin, western Australia. Gas Sci. Eng. 89, 103888. https://doi.org/10.1016/j.jngse.2021.103888.

Koeshidayatullah, A., Al-Azani, S., Baraboshkin, E., et al., 2022. FaciesViT: vision transformer for an improved core lithofacies prediction. Front. Earth Sci. 10, 992442. https://doi.org/10.3389/feart.2022.992442.

Lazar, O.R., Bohacs, K.M., Macquaker, J.H.S., et al., 2015. Capturing key attributes of fine-grained sedimentary rocks in outcrops, cores, and thin sections: nomenclature and Description guidelines. J. Sediment. Res. 85 (3), 230–246. https://doi.org/10.2110/jsr.2015.11.

Li, Q.Q., Xu, S., Li, J.L., et al., 2023. Effects of astronomical cycles on laminated shales of the Paleogene Shahejie Formation in the dongying sag, Bohai Bay Basin, China. Energies 16 (9), 3624. https://doi.org/10.3390/en16093624.

Li, T.W., Jiang, Z.X., Li, Z., et al., 2017. Continental shale pore structure characteristics and their controlling factors: a case study from the lower third member of the Shahejie Formation, Zhanhua Sag, Eastern China. Gas Sci. Eng. 45, 670–692. https://doi.org/10.1016/j.jngse.2017.06.005.

Li, Y.Y., Zha, M., Song, R.C., et al., 2021. Microstructure and pore systems of shallow-buried fluvial mudstone caprocks in Zhanhua depression, east China inferred from SEM and MICP. Mar. Petrol. Geol. 132, 105189. https://doi.org/10.1016/j.marpetgeo.2021.105189.

Liu, Q.H., Zhu, X.M., Yang, Y., et al., 2016. Sequence stratigraphy and seismic geomorphology application of facies architecture and sediment-dispersal patterns analysis in the third member of Eocene Shahejie Formation, slope system of Zhanhua Sag, Bohai Bay Basin. China. Mar. Petrol. Geol. 78, 766–784. https://doi.org/10.1016/j.marpetgeo.2015.11.015.

Liu, Y., Zhu, R.H., Zhai, S., et al., 2023. Lithofacies identification of shale formation based on mineral content regression using LightGBM algorithm: a case study in the Luzhou block, South Sichuan Basin, China. Energy Sci. Eng. 11 (11), 4256–4272. https://doi.org/10.1002/ese3.1579.

Ma, Y.Q., Fan, M.J., Lu, Y.C., et al., 2016. Climate-driven paleolimnological change controls lacustrine mudstone depositional process and organic matter accumulation: constraints from lithofacies and geochemical studies in the Zhanhua Depression, eastern China. Int. J. Coal Geol. 167, 103–118. https://doi.org/10.1016/j.coal.2016.09.014.

Pang, X.Q., Li, M., Li, B.Y., et al., 2023. Main controlling factors and movability evaluation of continental shale oil. Earth Sci. Rev. 243, 104472. https://doi.org/10.1016/j.earscirev.2023.104472.

Qi, J., Yang, Q., 2010. Cenozoic structural deformation and dynamic processes of the Bohai Bay basin province. China. Mar. Petrol. Geol. 27 (4), 757–771. https://doi.org/10.1016/j.marpetgeo.2009.08.012.

Shi, J.Y., Jin, Z.J., Liu, Q.Y., et al., 2019. Cyclostratigraphy and astronomical tuning of the middle eocene terrestrial successions in the Bohai Bay Basin, Eastern China. Glob. Planet. Change. 174, 115–126. https://doi.org/10.1016/j.gloplacha.2019.01.001.

Shi, J.Y., Jin, Z.J., Liu, Q.Y., et al., 2022. Laminar characteristics of lacustrine organic-rich shales and their significance for shale reservoir formation: a case study of the Paleogene shales in the Dongying Sag, Bohai Bay Basin, China. J. Asian Earth Sci. 223, 104976. https://doi.org/10.1016/j.jseaes.2021.104976.

Su, J.B., Zhu, W.B., Lu, H.F., et al., 2009. Geometry styles and quantification of inversion structures in the Jiyang depression, Bohai Bay Basin, eastern China. Mar. Petrol. Geol. 26 (1), 25–38. https://doi.org/10.1016/

j.marpetgeo.2007.08.003.
Su, S.Y., Jiang, Z.X., Shan, X.L., et al., 2018. Effect of lithofacies on shale reservoir and hydrocarbon bearing capacity in the Shahejie Formation, Zhanhua Sag, eastern China. J. Pet. Sci. Eng. 174, 1303–1308. https://doi.org/10.1016/j.petrol.2018.11.048.
Sun, N.L., Chen, T.Y., Gao, J.B., et al., 2023. Lithofacies and reservoir characteristics of saline lacustrine fine-grained sedimentary rocks in the northern Dongpu Sag, Bohai Bay Basin: implications for shale oil exploration. J. Asian Earth Sci. 252, 105686. https://doi.org/10.1016/j.jseaes.2023.105686.
Tang, Y., Cao, J., He, W.J., et al., 2021. Discovery of shale oil in alkaline lacustrine basins: the late paleozoic fengcheng formation, mahu sag, junggar basin, China. Pet. Sci. 18 (5), 1281–1293. https://doi.org/10.1016/j.petsci.2021.04.001.
Tian, M., Tan, M.J., Wang, M., 2023. Identification of shale lithofacies from FMI images and ECS logs using machine learning with GLCM features. Process 11 (10), 2982. https://doi.org/10.3390/pr11102982.
Wang, G.C., Garr, T.R., 2013. Organic-rich Marcellus Shale lithofacies modeling and distribution pattern analysis in the Appalachian basin. AAPG (Am. Assoc. Pet. Geol.) Bull. 97 (12), 2173–2205. https://doi.org/10.1306/05141312135.
Wang, M., Li, M., Li, J.B., et al., 2022. The key parameter of shale oil resource evaluation: oil content. Pet. Sci. 19 (4), 1443–1459. https://doi.org/10.1016/j.petsci.2022.03.006.
Williams, T.S., Bhattacharya, S., Song, L.S., et al., 2022. Petrophysical analysis and mudstone lithofacies classification of the HRZ shale, North Slope, Alaska. J. Pet. Sci. Eng. 208 (Part: C), 109454. https://doi.org/10.1016/j.petrol.2021.109454.
Xu, J.L., Wang, R.T., Zan, L., 2023. Shale oil occurrence and slit medium coupling based on a molecular dynamics simulation. J. Pet. Sci. Eng. 220 (Part A), 111151. https://doi.org/10.1016/j.petrol.2022.111151.
Xu, S., Hao, F., Shu, Z.G., et al., 2020. Pore structures of different types of shales and shale gas exploration of the Ordovician Wufeng and Silurian Longmaxi successions in the eastern Sichuan Basin, South China. J. Asian Earth Sci. 193, 104271. https://doi.org/10.1016/j.jseaes.2020.104271.
Xue, C.Q., Mcbeck, J.A., Lu, H.J., et al., 2023. Classification of shale lithofacies with minimal data: application to the early Permian shales in the Ordos Basin, China. J. Asian Earth Sci. 259, 105901. https://doi.org/10.1016/j.jseaes.2023.105901.
Zhang, Y., Dilek, Y., Zhang, F.Q., et al., 2020. Structural architecture and tectonic evolution of the cenozoic Zhanhua sag along the tan–Lu fault zone in the eastern North China: reconciliation of tectonic models on the origin of the Bohai Bay Basin. Tectonophysics 775, 228303. https://doi.org/10.1016/j.tecto.2019.228303.
Zhou, J., Wu, G.A., Geng, Y.N., et al., 2023. Laboratory study of the factors affecting hydraulic fracturing effect for inter-salt oil shale layers, Qianjiang Depression, China. Pet. Sci. 20 (3), 1690–1706. https://doi.org/10.1016/j.petsci.2023.01.002.