Oil Shale, 2026, 43, 3, 300–320 https://doi.org/10.3176/oil.2026.3.04

Effect of organic macerals on hydrocarbon generation ability of oil shale and coal and establishment of a contribution model – constraints from machine learning

Duoxiao Sun[a,b], Feng Ma[c,d]*, Hongbiao Wang[a,b], Hongliang Dang[a,e,f], Yanwei Bi[a,b], Pingchang Sun[a,b]

[a] College of Earth Sciences, Jilin University, Changchun, Jilin 130061, China
[b] Key-Lab for Oil Shale and Paragenetic Minerals of Jilin Province, Changchun, Jilin 130061, China
[c] School of Resources and Environment, University of Electronic Science and Technology of China, Chengdu 611731, China
[d] Research Institute of Petroleum Exploration & Development-Northwest (NWGI), PetroChina, Lanzhou 730020, China
[e] Technology Innovation Center for Exploration and Exploitation of Strategic Mineral Resources in Plateau Desert Region, Ministry of Natural Resources, Xining 810000, China
[f] Qinghai Geological Survey, Xining 810000, China

Received 13 November 2025, accepted 1 July 2026, available online 3 August 2026

Abstract. *To quantify the impact of organic macerals on hydrocarbon generation in oil shale and coal, this study analyzed 403 datasets. The results show distinct differences in maceral composition between the two materials. Using a random forest model with aquatic liptinite, terrigenous liptinite, vitrinite, and inertinite as inputs, TOC and HI predictions achieved R^2 values of 0.62 and 0.73, respectively. SHAP analysis reveals that aquatic liptinite is the dominant positive contributor to both parameters, while terrigenous liptinite shows a bidirectional effect. Based on these relationships, a quantitative contribution model for organic macerals in oil shale is established, offering insights for resource evaluation and exploration.*

Keywords: *oil shale, coal, organic macerals, random forest model, SHAP analysis.*

* Corresponding author, mafeng@petrochina.com.cn

1. Introduction

Oil shale and coal are very important combustible organic minerals, playing a significant role in the energy structure [1, 2]. Both coal and oil shale are organic-rich sedimentary rocks. From the perspectives of clean coal utilization and in-situ oil shale conversion, their hydrocarbon generation capacity is a core indicator for assessing the quality and development feasibility of coal and oil shale resources. This capacity is comprehensively constrained by various factors such as organic matter abundance, organic macerals (source of organic matter), and organic matter maturity. Among these, organic macerals, as the basic constituent units of organic matter in coal and oil shale, directly determine the hydrocarbon generation potential, pyrolysis efficiency, and product characteristics of the organic matter through their type, content, and spatial distribution. They are the key intrinsic factors regulating the hydrocarbon generation processes of oil shale and coal [1–5]. Differences in the source of organic matter (organic macerals) and depositional environments directly lead to significant distinctions in the organic maceral composition of coal versus oil shale.

Traditional studies often qualitatively or semi-quantitatively investigate the relationship between organic macerals and hydrocarbon generation capacity through experimental analyses such as optical microscopy and pyrolysis experiments [6–8]. However, these approaches face challenges in precisely quantifying the contribution weights of individual macerals and unveiling the complex non-linear relationships between macerals and key hydrocarbon generation parameters such as total organic carbon (TOC) and hydrogen index (HI). SHAP (SHapley Additive exPlanations) analysis, an explainable machine learning method grounded in game theory, can overcome the limitations of traditional analyses. By calculating the SHAP value of each feature variable relative to the target variable [9], SHAP quantifies the degree of influence of organic macerals on TOC and HI, identifies key controlling macerals, and visualizes interaction effects between feature variables. Therefore, based on a large compiled dataset of organic geochemical parameters and organic macerals from immature to low-maturity coal and oil shale, this paper investigates the impact of organic macerals on the hydrocarbon generation capacity of oil shale and coal. It focuses specifically on utilizing SHAP analysis to decipher the intrinsic relationships between organic maceral data and TOC/HI, aiming to provide theoretical support and technical references for the precise exploration, quality evaluation, and efficient development and utilization of oil shale and coal resources.

2. Data and research methods

2.1. Data collection

In fact, oil shale is rich in immature to low-maturity organic matter. Therefore, in order to facilitate the comparison of coal and oil shale at the same maturity level, the maturity of the oil shale and coal samples collected in this study ranges from immature to low maturity. This study compiled 178 sets of data for immature to low-maturity oil shales ($Ro \leq 0.7\%$) and 225 sets of data for coals from the Songliao Basin and the Junggar Basin in China, as well as from a series of basins in other countries [10–41]. The complete data sources for all samples are listed in Appendices 1 and 2 (see Supplementary online data). It should be noted that although samples with complete datasets – including organic maceral content, TOC, and HI – are relatively limited, the samples collected in this study generally provide both organic maceral and organic geochemical data from the same source. Furthermore, to ensure comparability of data across different basins, the data collected in this study generally adhere to similar testing methods and experimental instruments. For example, the TOC testing instrument was mainly the LECO TS-230, and the data for rock pyrolysis were obtained using the Rock Eval 6. The organic microscopic component counting scheme is based on percentage statistics and has been normalized.

2.2. Methods

Random forest (RF) was proposed by Breiman based on the integration of statistical learning theory with classification and regression methods [27]. The multiple classification regression decision tree (CART) included in the RF algorithm prevents overfitting and can adjust different types of input variables.

In this study, four organic maceral parameters – aquatic liptinite, terrigenous liptinite, vitrinite, and inertinite – were selected as sensitive factors to investigate their influence on the hydrocarbon generation potential (TOC and HI) of coal and oil shale. The RF model was developed using the Python 3.12 programming language, with implementation carried out via the scikit-learn library. The train_test_split method from the scikit-learn library was employed to partition the dataset into training and testing sets, with a ratio of 70% for training and 30% for testing.

Table 1. Optimal hyperparameter combination

Model	random_state	n_estimators	min_samples_leaf	min_samples_split	max_depth
TOC	240	100	3	4	9
HI	211	300	3	2	4

The GridSearchCV function in Python facilitates the identification of parameters that yield the highest model estimation accuracy [28]. By specifying the values and ranges of the hyperparameters for the RF estimator model and applying the GridSearchCV function, the optimal combination of RF hyperparameters was derived (Table 1).

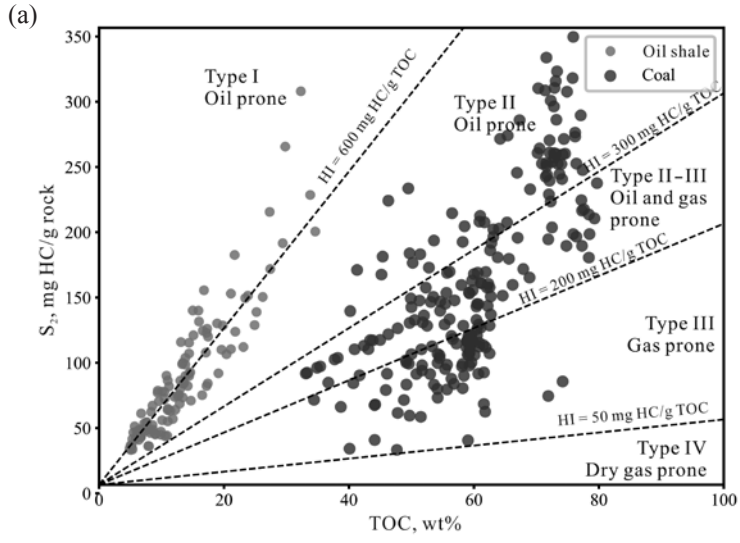
To identify the most influential factors affecting TOC and HI, the SHAP method was selected to evaluate the contribution value of each feature in predicting TOC and HI. SHAP is a game-theory-based approach designed to explain the output of any machine learning model [9, 29]. In this study, both the global features of multiple linear regression and the global/local features of the RF model, along with the contributions of individual samples, were analyzed.

3. Results

3.1. Organic geochemical characteristics

The collected data (Fig. 1) show that the TOC of oil shale samples is predominantly distributed between 5 and 25 wt%, with a maximum value of 34.6 wt% and an average of 13.36 wt% (Fig. 1a, b). The HI values are concentrated in the range of 435–780 mg HC/g TOC, with a few samples approaching 1000 mg HC/g TOC, and an average as high as 668.2 mg HC/g TOC (Fig. 1b, c). The S_2 values range from 35 to 301.4 mg HC/g rock, with an average of 81.6 mg HC/g rock (Fig. 1a, c). A strong positive correlation is observed between TOC and S_2, while moderate positive correlations exist between TOC and HI, and between S_2 and HI. The organic matter in oil shale samples is mainly type I and type II_1 (Fig. 1).

(a)



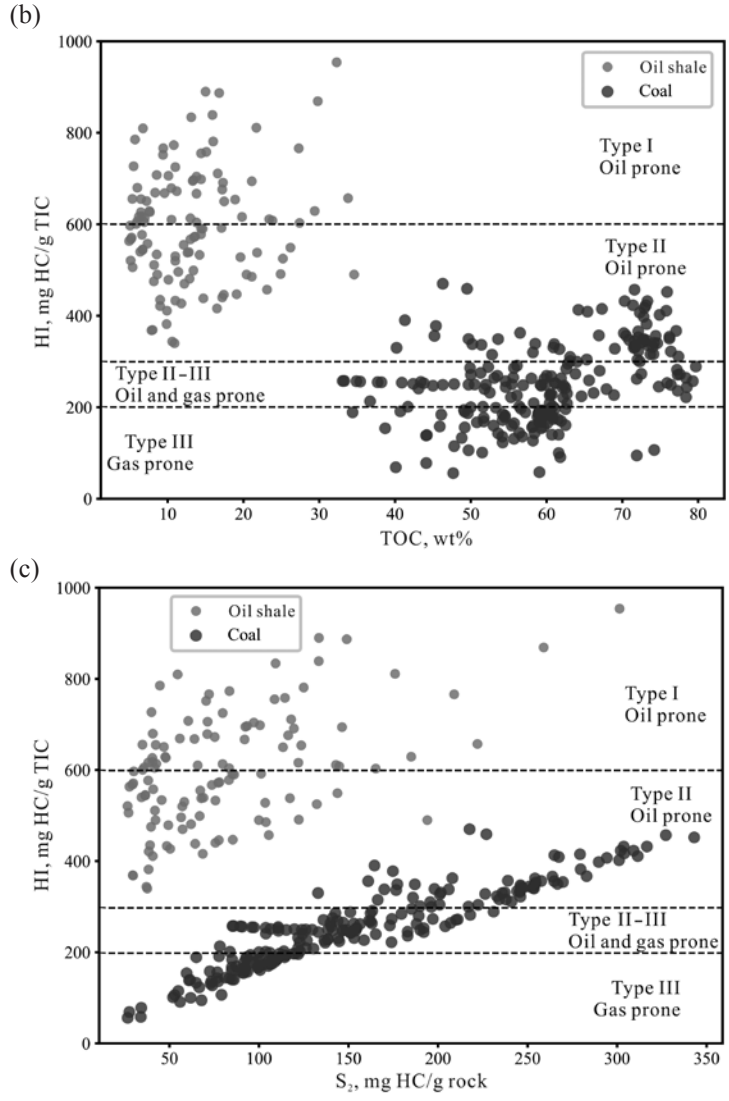


Fig. 1. Scatter plots of the organic geochemical characteristics of oil shale and coal samples: TOC–S_2 scatter plots, showing organic matter types (a); TOC–HI scatter plots (b); S_2–HI scatter plots (c).

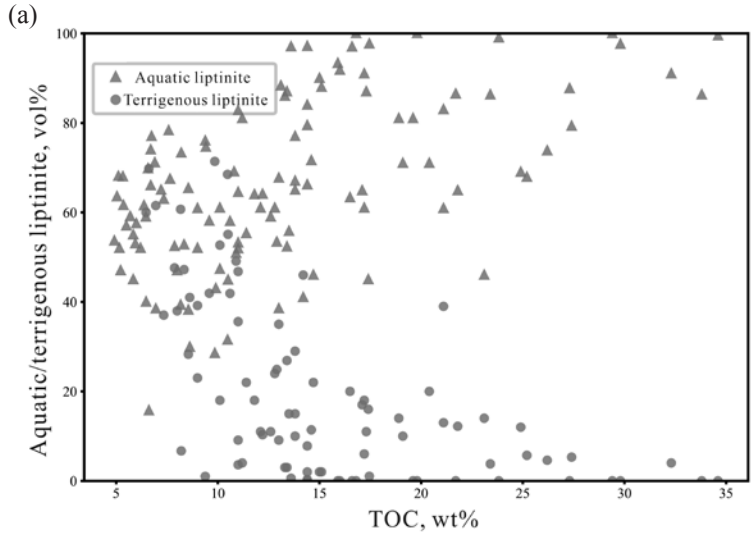
For coal samples, TOC ranges from 43.8 to 79.7 wt%, with an average of 59.68 wt% (Fig. 1a, b). HI mainly falls between 138.2 and 366 mg HC/g TOC, averaging 251.26 mg HC/g TOC (Fig. 1b, c). S_2 is predominantly distributed from 73.9 to 265 mg HC/g rock, reaching a maximum of 343.1 mg HC/g rock, with a high average value of 167.73 mg HC/g rock (Fig. 1a, c). Strong positive

correlations are observed among TOC, S_2, and HI in coal samples. The organic matter in coal samples is mainly type II–III and type III (Fig. 1).

Although the TOC of oil shale is only one-third to one-fifth that of coal, its hydrocarbon generation efficiency per unit mass of TOC (HI, S_2 /TOC) is two to three times higher than that of coal [30, 31]. This indicates that TOC in oil shale per unit mass typically generates more hydrocarbons than that in coal. Oil shale generally contains higher proportions of hydrogen, oxygen, and sulfur, whereas coal is primarily composed of carbon. These compositional differences lead to variations in combustion characteristics and hydrocarbon yields. One ton of TOC from oil shale can yield 200–300 kg of shale oil, whereas one ton of TOC from coal produces only 50–100 kg of tar (and this is feasible only for low-rank coal). In summary, although oil shale has lower TOC, its hydrocarbon generation potential is higher, exhibiting greater hydrocarbon generation efficiency than coal, which positions oil shale as a potentially efficient oil-generating resource [32–34].

3.2. Organic macerals

The collected data reveal that in oil shale samples, the content of aquatic liptinite is predominantly distributed between 45% and 97%, reaching a maximum of 100%, with a high average value of 66.83%. Terrigenous liptinite content primarily ranges from 5% to 30%, with a small number of samples approaching 70%, while the average is only 15.98% (Fig. 2). Vitrinite is mainly distributed between 5% and 30%, with an average content of 12.24%. Inertinite content ranges from 0% to 10%, with a notably low average of 4.93% (Fig. 3).



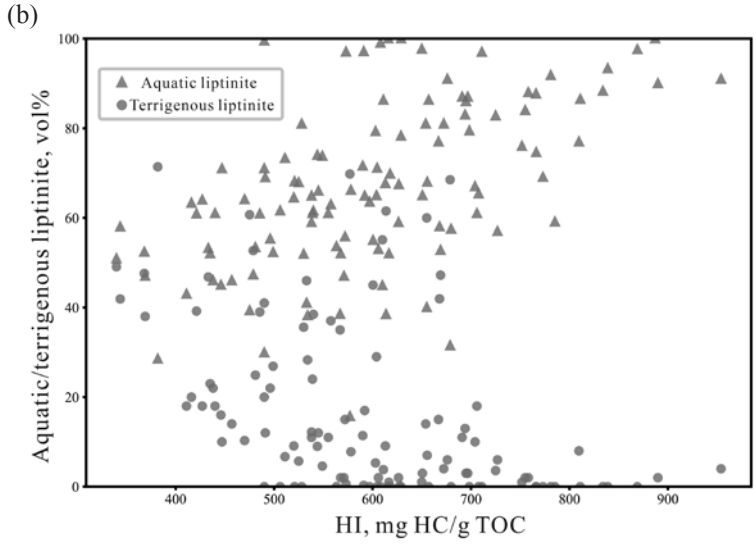
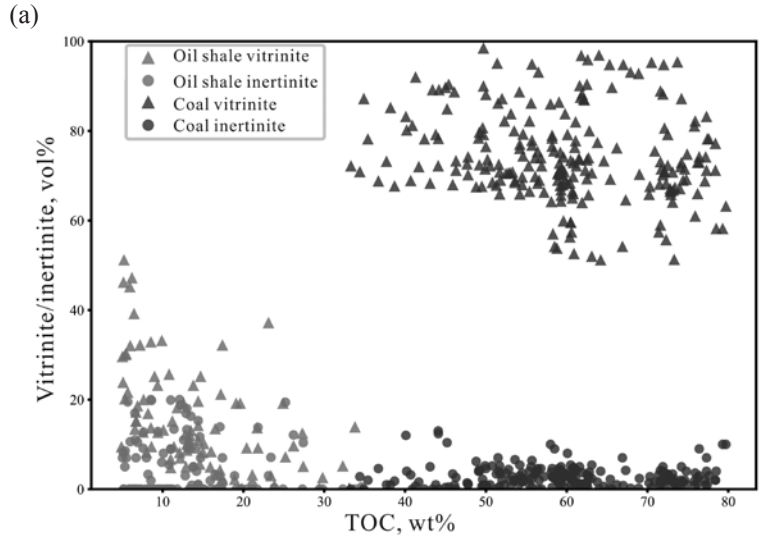


Fig. 2. Scatter plots of TOC vs. aquatic liptinite/terrigenous liptinite (a) and HI vs. aquatic liptinite/terrigenous liptinite (b).

In coal samples, the content of aquatic liptinite is extremely low (generally < 5%, only slightly higher in humic-sapropelic coal). Terrigenous liptinite content in coal is relatively low (1–10%). Vitrinite is predominantly distributed between 40% and 90%, reaching a maximum of 98%, with a high average value of 74.09%. Inertinite content mainly ranges from 0.4% to 5%, with a maximum of 13%, and an average of only 2.73% (Fig. 3).



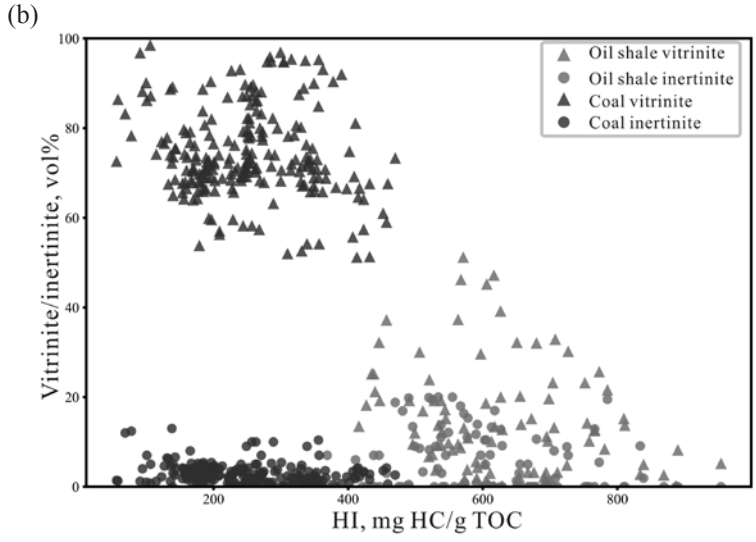
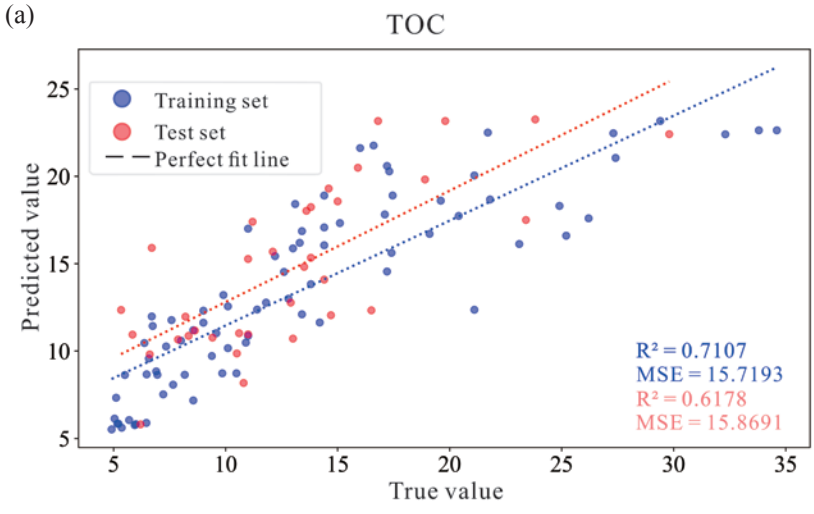


Fig. 3. Scatter plots of TOC vs. vitrinite/inertinite (a) and HI vs. vitrinite/inertinite (b) in oil shale and coal.

3.3. Random forest model and interpretation

An RF model was constructed using four organic maceral parameters to predict TOC and HI in oil shale. The red dots in Figure 4 indicate the predictive performance of the model on the test set, with R^2 (coefficient of determination) values of 0.62 and 0.73, respectively, alongside low root mean square error values, demonstrating satisfactory performance of the model. Therefore, the RF model shows adequate accuracy and adaptability, as well as good predictive ability, and can effectively reveal the complex relationship between organic microscopic components and TOC and HI.



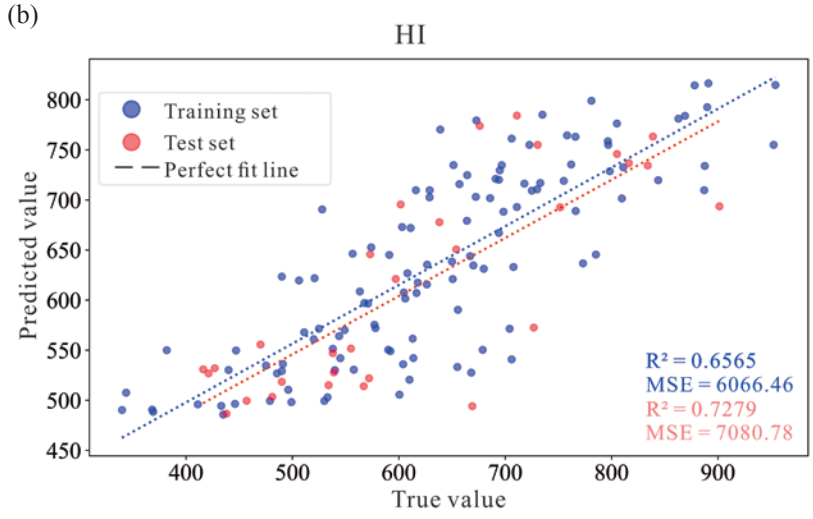
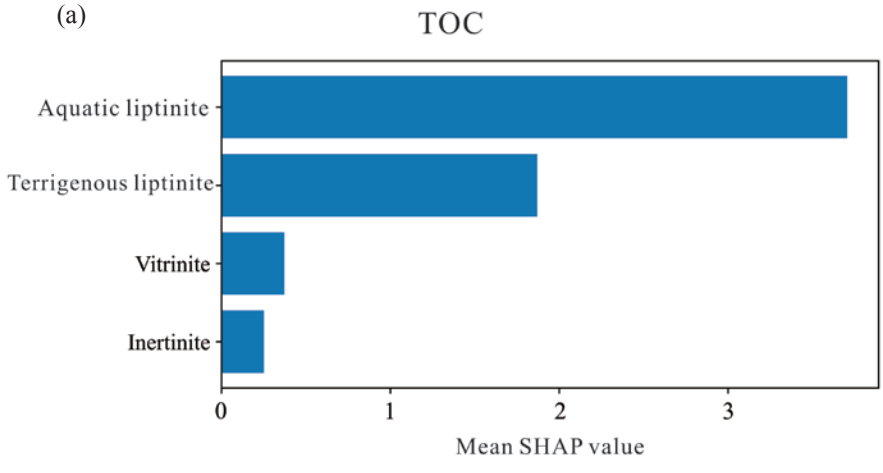


Fig. 4. Comparison of TOC real test tag and integrated output results (a) and HI real test tag and integrated output results.

Based on SHAP analysis, the mean SHAP values were obtained. The global importance of input variables (Fig. 5) represents the average of the absolute SHAP values for each feature across the entire dataset. The input variables are ranked by importance, where a higher mean SHAP value indicates greater importance of the variable [9]. The results demonstrate that aquatic liptinite contributes the most to the model, followed by terrigenous liptinite (Fig. 5). The findings indicate that aquatic liptinite is strongly correlated with TOC and HI, suggesting that it plays a dominant role in hydrocarbon generation potential.



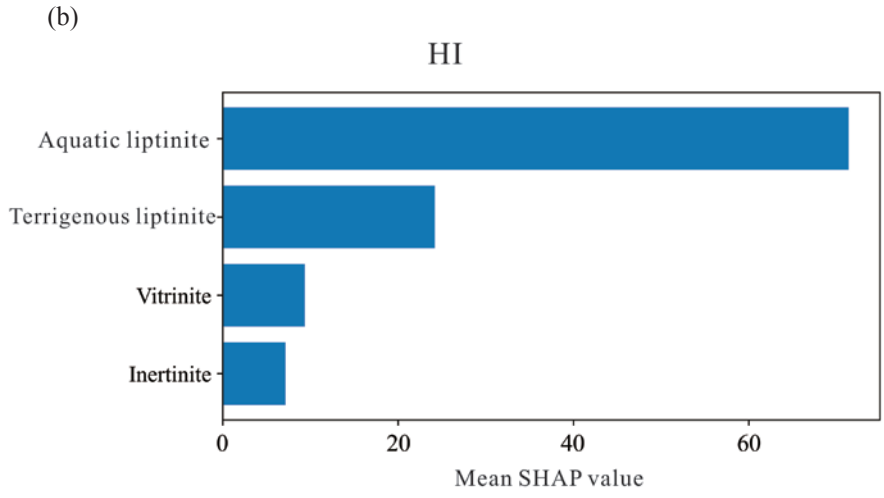


Fig. 5. SHAP values and feature-importance plots of the TOC model (a) and the HI model (b).

By analyzing the factor dependence plots (Fig. 6), the marginal effects of individual input variables on the model predictions in the RF model can be observed. Single dependency analysis illustrates the influence of individual factors on the content. The x-axis represents the feature values of the factor, while the y-axis represents the SHAP values of samples associated with the feature values. The analysis reveals the following:

1. Aquatic liptinite exhibits a strong linear positive correlation with both TOC and HI. In TOC prediction, when aquatic liptinite exceeds 70%, the SHAP values are generally greater than 0, making it more likely to increase TOC content. Conversely, when aquatic liptinite is below 70%, TOC content decreases. The threshold for aquatic liptinite to increase HI content is 70%.
2. The SHAP values of TOC samples show an initial increase followed by a decrease as terrigenous liptinite content increases. When terrigenous liptinite content ranges from 0% to 30%, the SHAP values of TOC samples gradually increase but they begin to decrease when terrigenous liptinite exceeds 30%.
3. Vitrinite has a certain negative influence on both TOC and HI contents.

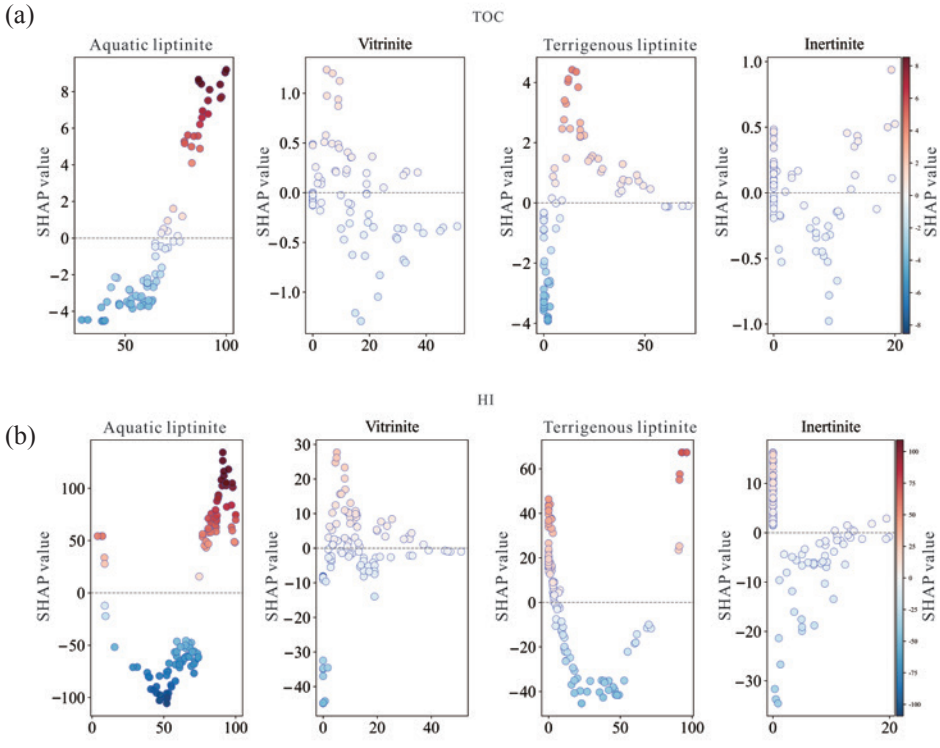


Fig. 6. Dependence plots of influencing factors for TOC (a) and HI (b).

Figure 7 displays four randomly selected SHAP force plots from the actual prediction results of TOC and HI, respectively. The SHAP values decompose the prediction of TOC and HI variation patterns into the sum of influences from each input variable [9].

Figure 7a represents predictions for different TOC contents. Figure 7a1 shows a case of low predicted TOC content. The predicted TOC value is 6.13 wt%, and the true value is 5.04 wt%. Figure 7a2 shows a case of high predicted TOC content. The predicted TOC value is 20.28 wt%, and the true value is 17.30 wt%. Overall, regardless of the increase or decrease in TOC content, aquatic liptinite is the primary influencing factor in TOC content variation. However, when TOC content is low, terrigenous liptinite has a greater influence on TOC, and as TOC content increases, the influence of terrigenous liptinite gradually decreases.

Figure 7b represents predictions for different HI contents. Figure 7b1 shows a case of low predicted HI content. The predicted HI value is 542.01 mg HC/g TOC, and the true value is 545.0 mg HC/g TOC. Figure 7b2 shows a case of high predicted HI content. The predicted HI value is 729.62 mg HC/g TOC, and the true value is 694.85 mg HC/g TOC. Overall, regardless of the increase or decrease in HI content, aquatic liptinite is the

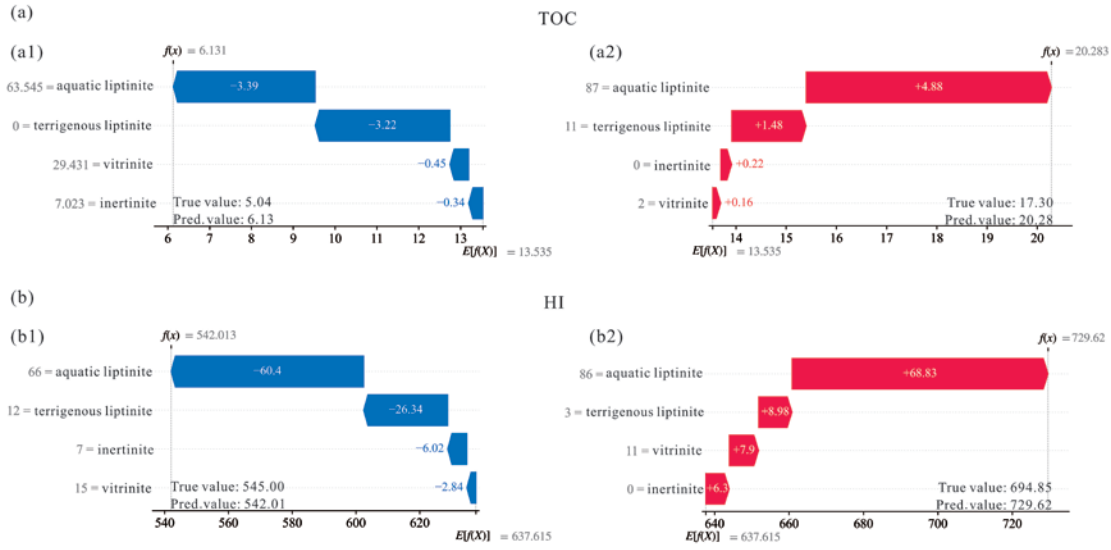


Fig. 7. Explanation of evaluation results based on the SHAP TreeExplainer for TOC (a) and HI (b).

primary influencing factor in HI content variation. However, when HI content is low, terrigenous liptinite has a certain influence on HI, and as HI content increases, the influence of terrigenous liptinite gradually decreases.

4. Discussion

4.1. Contribution of organic macerals to TOC and HI

TOC fundamentally represents the total carbon content of organic matter in oil shale and coal. In oil shale, aquatic liptinite content shows a significant positive correlation with TOC. For every 10% increase in aquatic liptinite content, the TOC of oil shale increases by an average of 5–8 wt% (Fig. 2a). Terrigenous liptinite content in oil shale ranges from 5% to 30%, and its moderate inherent carbon content (72–82%) results in a limited contribution to TOC [35]. However, in low-grade oil shale, a 10% increase in terrigenous liptinite content elevates TOC by 1–3 wt% (Fig. 2a). Based on the scatter plot of TOC/HI versus liptinite, oil shale TOC exhibits a clear positive correlation with liptinite, with a correlation coefficient of $R^2 = 0.563$ (Fig. 8a). Vitrinite content in oil shale varies widely (5–50%), but oil shale with high TOC tends to have lower vitrinite content, showing a minor negative correlation between vitrinite and TOC. Inertinite content in oil shale is < 10%, and its contribution to TOC is much more limited than that of the dominant macerals (Fig. 3a).

In coal, aquatic liptinite content is extremely low, and its contribution to TOC is negligible [15, 17, 21, 36, 37]. Terrigenous liptinite content in coal is relatively low (1–10%) and results in a limited contribution to TOC [35]. Coal TOC shows a poor correlation with liptinite (Fig. 8a). Vitrinite is the most dominant maceral in coal (content 40–90%) (Fig. 3a). Derived from the lignocellulosic tissues of higher plants and with a high inherent carbon content (75–85%), it directly determines the TOC level of coal [44]. Inertinite content in coal is low (< 5%), and its contribution to TOC is much more limited than that of the dominant macerals (Fig. 3a).

The HI fundamentally reflects the macroscopic manifestation of hydrogen enrichment in organic matter. Its value is primarily determined by the hydrogen content of organic macerals. The contribution of different macerals directly dictates the oil-generating capacity of oil shale and coal [39–41]. In oil

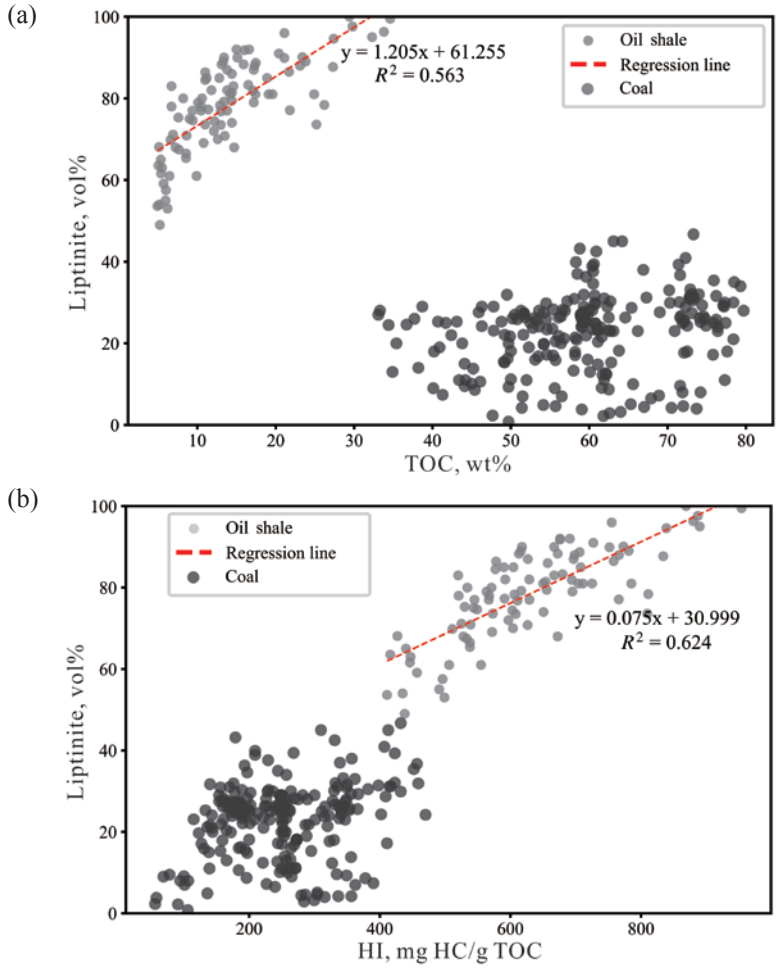


Fig. 8. Scatter plots of TOC vs. liptinite (a) and HI vs. liptinite (b) for oil shale and coal samples.

shale, aquatic liptinite content shows a significant positive correlation with HI. Aquatic liptinite contributes the most to HI, with every 10% increase in its content raising HI by an average of 80–100 mg HC/g TOC (Fig. 2b). Terrigenous liptinite's contribution to HI is secondary to aquatic liptinite; a 10% increase in its content elevates HI by 25–40 mg HC/g TOC, a notably smaller increase than that from aquatic liptinite (Fig. 2b). Based on the relationship between the lipidic group and HI (Fig. 8b), oil shale HI exhibits a clear positive correlation with the lipidic group, with a correlation coefficient of $R^2 = 0.624$. Vitrinite in oil shale exerts a minor inhibitory effect on HI: a 10% increase in its content reduces HI by 60–80 mg HC/g TOC (Fig. 3b). Inertinite content in oil shale is generally below 10%, and its contribution to HI is minor (Fig. 3b).

In coal, aquatic liptinite content is extremely low, making its contribution to HI negligible. Terrigenous liptinite content in coal is relatively low (1–10%). Coal HI shows a poor correlation with the lipidic group (Fig. 10b). Vitrinite is the primary organic maceral in coal (40–90%) (Fig. 3b). Its hydrocarbon generation capacity varies significantly with the degree of coalification [42]. In low-rank coal, vitrinite has a slightly higher hydrogen content, resulting in a more limited inhibitory effect. In high-rank coal, due to intensified aromatization, its inhibitory effect strengthens [43–46]. Inertinite content in coal is typically below 5%, and its contribution to HI is much more limited than that of the dominant macerals (Fig. 3b).

4.2. Key factors controlling oil shale quality

The key organic macerals controlling oil shale quality are primarily aquatic liptinite, followed by minor amounts of terrigenous liptinite and vitrinite. Vitrinite mainly exerts a certain negative influence. Specifically, when TOC < 15 wt%, oil shale quality is jointly controlled by aquatic liptinite and terrigenous liptinite (Fig. 9), indicating that during the depositional period of medium- to low-quality oil shale, influenced by dual supply from aquatic organisms and terrigenous organic matter, the quality of the corresponding oil shale gradually improves with increasing organic matter supply [47–49]. In contrast, when TOC > 15 wt%, the contribution of terrigenous liptinite decreases significantly, and quality is primarily controlled by aquatic liptinite. This suggests that high-quality oil shale forms during periods of minimal terrigenous organic matter influence, where aquatic productivity and effective organic matter preservation are the key factors determining oil shale quality [50–52]. Thus, TOC = 15 wt% can be used as a boundary for distinguishing oil shale quality: high-quality oil shale is primarily controlled by aquatic liptinite, while ordinary oil shale is jointly controlled by aquatic liptinite and terrigenous liptinite. This aligns with the SHAP analysis results, which indicate that when aquatic liptinite content exceeds 70% and terrigenous liptinite content is below 30% in oil shale, both TOC and HI significantly increase (Fig. 6).

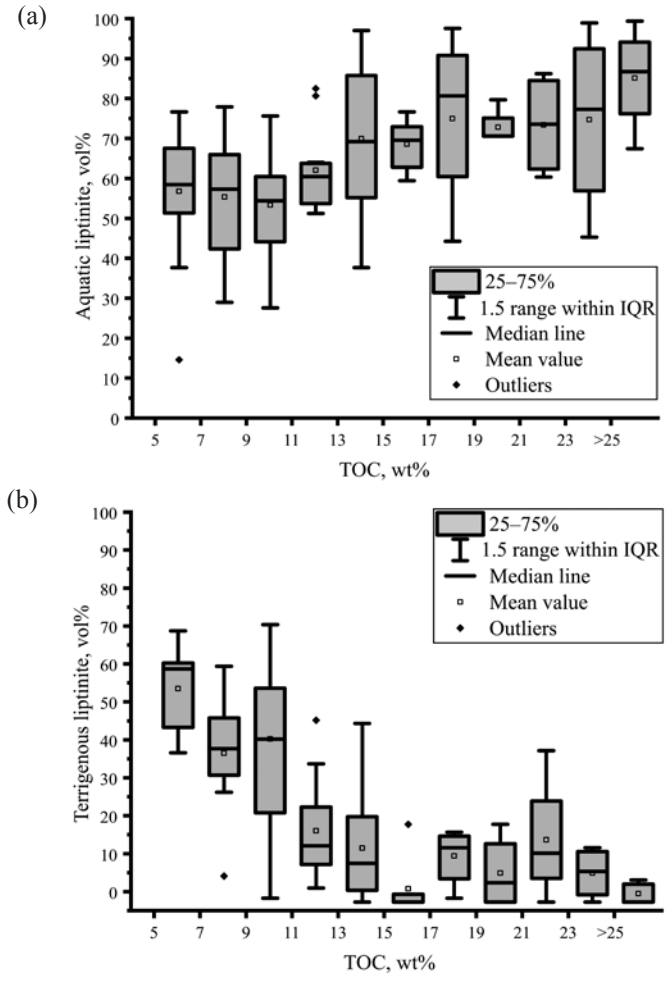


Fig. 9. Box plots of TOC vs. aquatic liptinite (a) and TOC vs. terrigenous liptinite (b) in oil shale.

4.3. Quantitative model of organic maceral contributions to oil shale

The core industrial value of oil shale depends on the oil production potential and economic viability of its organic macerals [53]. From the perspective of hydrocarbon generation [54], only aquatic liptinite, terrigenous liptinite, and vitrinite are considered as organic macerals in oil shale, while inertinite is temporarily excluded from calculations due to its generally negligible content ($\leq 5\%$) and lack of practical contribution to oil yield. The core evaluation metric of the model is the oil production efficiency index of oil shale, which ranges from 0 to 100, with higher values indicating superior oil shale quality.

To quantify the comprehensive contribution of each maceral to oil shale quality, the formula is defined as follows:

$$C = \frac{\mathrm{A}\times P_A + T\times P_T - V\times P_V}{\mathrm{W}\times(\mathrm{A+T+V})}, \quad (1)$$

where C represents the oil production efficiency index of oil shale, W denotes the contribution weight of organic macerals, A is the percentage content of aquatic liptinite, T is the percentage content of terrigenous liptinite, V is the percentage content of vitrinite, P_A is the unit oil production contribution coefficient of aquatic liptinite, P_T is the unit oil production contribution coefficient of terrigenous liptinite, and P_V is the unit loss coefficient of vitrinite. Based on the SHAP analysis results, the values are set as follows: $W = 0.7$, $P_A = 70$, $P_T = 30$, and $P_V = 130$.

The model calculation results indicate that when $C \geq 80$, the oil shale is classified as premium-grade, with aquatic liptinite content $\geq 85\%$ and vitrinite content $\leq 5\%$. When $60 \leq C < 80$, the oil shale is classified as medium-grade, with aquatic liptinite content between 70% and 85% and vitrinite content between 5% and 10%. When $C < 60$, the oil shale is classified as low-grade, with aquatic liptinite content $< 70\%$ and vitrinite content $> 10\%$.

5. Conclusions

Based on data from immature to low-maturity oil shale and coal samples from multiple basins, this study analyzes the contributions of organic macerals to organic matter abundance and hydrocarbon generation potential, identifies key factors controlling the quality of oil shale and coal, and focuses on interpreting the intrinsic relationships between organic maceral data and TOC/HI using SHAP analysis. A quantitative model for the contribution of organic macerals to the hydrocarbon generation capacity of oil shale was established.

1. Machine learning models were developed based on the relationships between TOC/HI and organic macerals in oil shale. SHAP analysis was used to interpret the model results, confirming that aquatic liptinite has the greatest influence on TOC and HI in oil shale, followed by terrigenous liptinite, while vitrinite exhibits a minor negative influence.
2. Oil shale is dominated by aquatic liptinite and terrigenous liptinite, whereas coal is dominated by vitrinite. High-quality oil shale is primarily controlled by aquatic liptinite, while ordinary oil shale is jointly controlled by aquatic liptinite and terrigenous liptinite. The hydrocarbon generation capacity of oil shale significantly increases when aquatic liptinite content exceeds 70% and terrigenous liptinite content remains below 30%. Vitrinite is the most abundant organic maceral in coal, typically accounting for 50–80%, making it the most critical factor influencing coal quality.
3. Based on these relationships, a quantitative contribution model was constructed to evaluate the hydrocarbon generation capacity of oil shale using organic maceral composition.

Data availability statement

All data used in this article are publicly available. No new data were created or analyzed in this study.

Acknowledgments

This study was supported by the PetroChina Company Limited Science and Technology Project, "Research on New Theories of Overseas Oil and Gas Geology, New Resource Evaluation Technologies, and Advanced Region Selection" (project No. 2023ZZ07), and by the National Natural Science Foundation of China (grant No. 42372125). The publication costs of this article were partially covered by the Estonian Academy of Sciences.

Supplementary online data

Supplementary online data to this article can be found at https:/doi.org/... and includes Appendices 1 and 2.

References

1. Liu, Z., Dong, Q., Ye, S., Zhu, J., Guo, W., Li, D. et al. The situation of oil shale resources in China. *Journal of Jilin University (Earth Science Edition)*, 2006, **36**(6), 869–876.
2. Liu, R., Liu, Z. Oil shale resource situation and multi-purpose development potential in China and abroad. *Journal of Jilin University (Earth Science Edition)*, 2006, **36**(6), 892–898.
3. Liu, Z., Meng, Q., Dong, Q., Zhu, J., Guo, W., Ye, S. et al. Characteristics and resource potential of oil shale in China. *Oil Shale*, 2017, **34**(1), 15–41. https://doi.org/10.3176/oil.2017.1.02
4. Wang, Y., Zhai, G., Bao, S.-J., Ren, S., Ge, M., Zhou, Z. Latest progress and trend forecast of China's shale gas exploration and development. *Acta Geologica Sinica (English Edition)*, 2025, **89**(S1) 211–213. https://doi.org/10.1111/1755-6724.12303_25
5. Radwan, A. E., Yin, S., Hakimi, M. H., Li, H. Petroleum geology of conventional and unconventional resources: introduction. *Geological Journal*, 2023, **58**(11), 3965–3969. https://doi.org/10.1002/gj.4898
6. Murchison, D. G. Recent advances in organic petrology and organic geochemistry: an overview with some reference to 'oil from coal'. *Geological Society, London, Special Publications*, 1987, **32**, 257–302. https://doi.org/10.1144/GSL.SP.1987.032.01.15
7. Suárez-Ruiz, I., Flores, D., Mendonça Filho, J. G., Hackley, P. C. Review and update of the applications of organic petrology: Part 1, geological applications. *International Journal of Coal Geology*, 2012, **99**, 54–112. https://doi.org/10.1016/j.coal.2012.02.004

8. Xiao, L., Li, Z., Xu, L., Wang, L., Yang, Y. Characteristics of organic macerals and their influence on hydrocarbon generation and storage: a case study of continental shale of the Yanchang Formation from the Ordos Basin, China. *Geofluids*, 2021, 5537154. https://doi.org/10.1155/2021/5537154
9. Lundberg, S. M., Lee, S.-I. A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems*, 2017, **30**, 4766–4777.
10. Jamaluddin,Wagreich, M.,Schöpfer, K.,Sachsenhofer, R. F.,Maria,Rahmawati, D. Hydrocarbon potential and depositional environment of the Middle Miocene Balikpapan Formation, lower Kutai Basin, Indonesia: sedimentology, calcareous nannofossil, organic geochemistry, and organic petrography integrated approach. *International Journal of Coal Geology*, 2024, **293**, 104591. https://doi.org/10.1016/j.coal.2024.104591
11. Hu, F., Misch, D., Zhang, P., Meng, Q., Sachsenhofer, R. F., Xu, Y. et al. Influence of high-frequency lake level fluctuations on organic matter accumulation in the northern Qaidam Basin, NW China: insights from spectral attribute analysis and geochemistry. *Journal of Asian Earth Sciences*, 2024, **260**, 105937. https://doi.org/10.1016/j.jseaes.2023.105937
12. Cao, T., Liu, H., Xiao, J., Pan, A., Deng, M. Paleoenviromental reconstruction and organic matter accumulation mechanism for Youganwo Formation oil shale in Maoming Basin. *Earth Science*, 2024, **49**(4), 1367–1384. http://dx.doi.org/10.3799/dqkx.2022.260
13. Camacho-Aristizabal, L., Burnaz, L., Castro-Vera, L., Mojica Silva, L., Littke, R. Organic petrology and geochemistry data reveal depositional and thermal history of coal in the Guaduas formation, Colombian Eastern Cordillera. *International Journal of Coal Geology*, 2024, **289**, 104549. https://doi.org/10.1016/j.coal.2024.104549
14. Nie, Y., Fu, X., Liang, J., Wei, H., Chen, Z., Lin, F. et al. The Toarcian Oceanic Anoxic Event in a shelf environment (Eastern Tethys): implications for weathering and redox conditions. *Sedimentary Geology*, 2023, **455**, 106476. https://doi.org/10.1016/j.sedgeo.2023.106476
15. Ajuaba, S., Sachsenhofer, R. F., Meier, V., Gross, D., Schnyder, J., Omodeo-Salé, S. et al. Coaly and lacustrine hydrocarbon source rocks in Permo-Carboniferous graben deposits (Weiach well, northern Switzerland). *Marine and Petroleum Geology*, 2023, **150**, 106147. https://doi.org/10.1016/j.marpetgeo.2023.106147
16. Zhang, P., Misch, D., Meng, Q., Sachsenhofer, R. F., Liu, Z., Jia, J. et al. Lateral changes of organic matter preservation in the lacustrine Qingshankou Formation (Cretaceous Songliao Basin, NE China): evidence for basin segmentation. *International Journal of Coal Geology*, 2022, **254**, 103984. https://doi.org/10.1016/j.coal.2022.103984
17. Fikri, H. N., Sachsenhofer, R. F., Bechtel, A., Gross, D. Coal deposition in the Barito Basin (Southeast Borneo): the Eocene Tanjung Formation compared to the Miocene Warukin Formation. *International Journal of Coal Geology*, 2022, **263**, 104117. https://doi.org/10.1016/j.coal.2022.104117

18. Fikri, H. N., Sachsenhofer, R. F., Bechtel, A., Gross, D. Organic geochemistry and petrography in Miocene coals in the Barito Basin (Tutupan Mine, Indonesia): evidence for astronomic forcing in kerapah type peats. *International Journal of Coal Geology*, 2022, **256**, 103997. https://doi.org/10.1016/j.coal.2022.103997
19. Zhang, P., Misch, D., Hu, F., Kostoglou, N., Sachsenhofer, R. F., Liu, Z. et al. Porosity evolution in organic matter-rich shales (Qingshankou Fm.; Songliao Basin, NE China): implications for shale oil retention. *Marine and Petroleum Geology*, 2021, **130**, 105139. https://doi.org/10.1016/j.marpetgeo.2021.105139
20. Körmös, S., Bechtel, A., Sachsenhofer, R. F., Radovics, B. G., Katalin, M., Félix, S. Petrographic and organic geochemical study of the Eocene Kosd Formation (northern Pannonian Basin): implications for paleoenvironment and hydrocarbon source potential. *International Journal of Coal Geology*, 2020, **228**, 103555. https://doi.org/10.1016/j.coal.2020.103555
21. Song, Y., Liu, Z., Bechtel, A., Sachsenhofer, R. F., Gross, D., Meng, Q. Paleoenvironmental reconstruction of the coal- and oil shale-bearing interval in the lower Cretaceous Muling Formation, Laoheishan Basin, northeast China. *International Journal of Coal Geology*, 2017, **172**, 1–18. https://doi.org/10.1016/j.coal.2017.01.010
22. Song, Y., Bechtel, A., Sachsenhofer, R. F., Gross, D., Liu, Z., Meng, Q. Depositional environment of the Lower Cretaceous Muling Formation of the Laoheishan Basin (NE China): implications from geochemical and petrological analyses. *Organic Geochemistry*, 2017, **104**, 19–34. https://doi.org/10.1016/j.orggeochem.2016.11.008
23. Bai, Y., Liu, Z., Sun, P., Liu, R., Hu, X., Zhou, R. et al. Diverse sedimentary conditions during deposition of coal and oil shale from the Meihe Basin (Eocene, NE China). *Journal of Sedimentary Research*, 2017, **87**(10), 1100–1120. https://doi.org/10.2110/jsr.2017.60
24. Životić, D., Bechtel, A., Sachsenhofer, R. F., Gratzer, R., Radić, D., Obradović, M. et al. Petrological and organic geochemical properties of lignite from the Kolubara and Kostolac basins, Serbia: implication on Grindability Index. *International Journal of Coal Geology*, 2014, **131**, 344–362. https://doi.org/10.1016/j.coal.2014.07.004
25. Strobl, S. A. I., Sachsenhofer, R. F., Bechtel, A., Meng, Q. Paleoenvironment of the Eocene coal seam in the Fushun Basin (NE China): implications from petrography and organic geochemistry. *International Journal of Coal Geology*, 2014, **134–135**, 24–37. https://doi.org/10.1016/j.coal.2014.10.001
26. Li, J. *Study on the Oil Shale Geochemistry of Permian Lucaogou Formation in the Northern Bogda Mountain*. PhD thesis. China University of Geosciences (Beijing), China, 2009.
27. Breiman, L. Random forests. *Machine Learning*, 2001, **45**, 5–32. https://doi.org/10.1023/A:1010933404324
28. Kim, Y., Kim, Y. Explainable heat-related mortality with random forest and SHapley Additive exPlanations (SHAP) models. *Sustainable Cities and Society*, 2022, **79**, 103677. https://doi.org/10.1016/j.scs.2022.103677
29. Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B. et al. From local explanations to global understanding with explainable AI for trees.

Nature Machine Intelligence, 2020, **2**, 56–67. https://doi.org/10.1038/s42256-019-0138-9

30. Jaber, J. O., Amri, A., Ibrahim, K. Experimental investigation of effects of oil shale composition on its calorific value and oil yield. *International Journal of Oil, Gas and Coal Technology*, 2011, **4**(4), 307–321. https://doi.org/10.1504/IJOGCT.2011.043714
31. Mathews, J. P., Krishnamoorthy, V., Louw, E., Tchapda, A. H. N., Castro-Marcano, F., Karri, V. et al. A review of the correlations of coal properties with elemental composition. *Fuel Processing Technology*, 2014, **121**, 104–113. https://doi.org/10.1016/j.fuproc.2014.01.015
32. Gao, B., Wu, X., Zhang, Y., Chen, X., Bian, R., Li, Q. et al. Hydrocarbon generation and evolution characteristics of Triassic Zhangjiatan oil shale in southern Ordos Basin. *Petroleum Geology & Experiment*, 2022, **44**(1), 24–32. https://doi.org/10.11781/sysydz202201024
33. Pollastro, R. M.,Cook, T. A.,Roberts, L. N. R.,Schenk, C. J.,Lewan, M. D. et al. Assessment of undiscovered oil resources in the Devonian-Mississippian Bakken Formation, Williston Basin Province, Montana and North Dakota. U.S. Geological Survey Fact Sheet 2008–3021, 2008. https://doi.org/10.3133/FS20083021
34. Dyni, J. R. Geology and resources of some world oil-shale deposits. *Oil Shale*, 2003, **20**(3), 193–252. https://doi.org/10.3176/oil.2003.3.02
35. Peters, K. E., Walters, C. C., Moldowan, J. M. *The Biomarker Guide: Volume 1, Biomarkers and Isotopes in the Environment and Human History*. Cambridge University Press, 2007.
36. Strobl, S. A. I., Sachsenhofer, R. F., Bechtel, A., Meng, Q., Sun, P. Deposition of coal and oil shale in NE China: the Eocene Huadian Basin compared to the coeval Fushun Basin. *Marine and Petroleum Geology*, 2015, **64**, 347–362. https://doi.org/10.1016/j.marpetgeo.2015.03.014
37. Zdravkov, A., Bechtel, A., Sachsenhofer, R. F., Kortenski, J. Palaeoenvironmental implications of coal formation in Dobrudzha Basin, Bulgaria: insights from organic petrological and geochemical properties. *International Journal of Coal Geology*, 2017, **180**, 1–17. https://doi.org/10.1016/j.coal.2017.07.004
38. Stach, E. *Stach's Textbook of Coal Petrology*. 3rd ed. Gebrüder Borntraeger, Berlin-Stuttgart, 1997.
39. Taylor, G. H., Teichmuller, M., Davis, A., Diessel, C. F. K., Littke, R., Robert, P. *Organic Petrology*. Gebruder Borntraeger, Berlin, Stutgart, 1998.
40. Kalkreuth, W., Holz, M., Mexias, A., Balbinot, M., Levandowski, J., Willett, J. et al. Depositional setting, petrology and chemistry of Permian coals from the Paraná Basin: 2. South Santa Catarina Coalfield, Brazil. *International Journal of Coal Geology*, 2010, **84**(3–4), 213–236. https://doi.org/10.1016/j.coal.2010.08.008
41. Liu, C., Zhao, W., Sun, L., Wang, X., Sun, Y., Zhang, Y. et al. Geochemical assessment of the newly discovered oil-type shale in the Shuangcheng area of the northern Songliao Basin, China. *Journal of Petroleum Science and Engineering*, 2021, **196**, 107755. https://doi.org/10.1016/j.petrol.2020.107755

42. Yin, Z., Xu, H., Chen, Y., Zhao, T., Wu, J. Experimental simulate on hydrogen production of different coals in underground coal gasification. *International Journal of Hydrogen Energy*, 2023, **48**(19), 6975–6985. https://doi.org/10.1016/j.ijhydene.2022.03.205
43. Stanger, R., Xie, W., Wall, T., Lucas, J., Mahoney, M. Dynamic behaviour of coal macerals during pyrolysis – associations between physical, thermal and chemical changes. *Proceedings of the Combustion Institute*, 2013, **34**(2), 2393–2400. https://doi.org/10.1016/j.proci.2012.07.003
44. Liu, Y., Zhu, Y., Liu, S., Chen, S., Li, W., Wang, Y. Molecular structure controls on micropore evolution in coal vitrinite during coalification. *International Journal of Coal Geology*, 2018, **199**, 19–30. https://doi.org/10.1016/j.coal.2018.09.012
45. Xie, K.-C. *Structure and Reactivity of Coal: A Survey of Selected Chinese Coals*. Science Press, Beijing, 2002.
46. Wang, A., Huang, J., Zhao, M., Liu, Y., Cao, D. et al. Effects of functional groups in coal with different vitrinite/inertinite ratios on pyrolysis products. *ACS Omega*, 2023, **8**(20), 18202–18211. https://doi.org/10.1021/acsomega.3c01635
47. Bowker, K. A. Barnett shale gas production, Fort Worth Basin: issues and discussion. *AAPG Bulletin*, 2007, **91**(4), 523–533. https://doi.org/10.1306/06190606018
48. Hutton, A. C., Kantsler, A. J., Cook, A. C., McKirdy, D. M. Organic matter in oil shales. *The APPEA Journal*, 1980, **20**(1), 44–67.
49. Sun, P., Sachsenhofer, R. F., Liu, Z., Strobl, S. A. I., Meng, Q., Liu, R. et al. Organic matter accumulation in the oil shale-and coal-bearing Huadian Basin (Eocene; NE China). *International Journal of Coal Geology*, 2013, **105**, 1–15. https://doi.org/10.1016/j.coal.2012.11.009
50. Jin, J., Wang, J., Meng, Y., Sun, P., Liu, Z., Li, Y. et al. Main controlling factors and development model of the oil shale deposits in the Late Permian Lucaogou Formation, Junggar Basin (NW China). *ACS Earth and Space Chemistry*, 2002, **6**(4), 1080–1094. https://doi.org/10.1021/acsearthspacechem.2c00015
51. Shang, F., Zhou, H., Liu, Y., Zhou, X., Wang, L., Bi, H. et al. A discussion on the organic matter enrichment model of the Nenjiang Formation, Songliao Basin: a case study of oil shale in the 1st and 2nd members of the Nenjiang Formation. *Geology in China*, 2020, **47**(1), 236–248. https://doi.org/10.12029/gc20200119
52. Wang, H., Niu, D., Luan, Z., Dang, H., Pan, X., Sun, P. Kinetic characteristics of secondary hydrocarbon generation from oil shale and coal at different maturation stages: insights from open-system pyrolysis. *International Journal of Coal Geology*, 2025, **308**, 104845. https://doi.org/10.1016/j.coal.2025.104845
53. Liu, B., Teng, J., Mastalerz, M. Maceral control on the hydrocarbon generation potential of lacustrine shales: a case study of the Chang 7 Member of the Triassic Yanchang Formation, Ordos Basin, North China. *Energies*, 2023, **16**(2), 636. https://doi.org/10.3390/en16020636
54. Lin, Z.-Z., Li, J.-Q., Lu, S.-F., Hu, Q.-H., Zhang, P.-F., Wang, J.-J. et al. The occurrence characteristics of oil in shales matrix from organic geochemical screening data and pore structure properties: an experimental study. *Petroleum Science*, 2024, **21**(1), 1–13. https://doi.org/10.1016/j.petsci.2023.09.002