Defining kerogen maturity from orbital hybridization by machine learning

Jun Ma [a,b], Dongliang Kang [a,b], Xiaohe Wang [a,b], Ya-Pu Zhao [a,b,*]

[a] State Key Laboratory of Nonlinear Mechanics, Institute of Mechanics, Chinese Academy of Sciences, Beijing 100190, China
[b] School of Engineering Science, University of Chinese Academy of Sciences, Beijing 100049, China

ABSTRACT

Kerogen is the primary material for oil and gas. Its maturity is used to determine the potential for hydrocarbon generation. Nowadays, kerogen maturity is mainly measured experimentally and characterized by its chemical composition. The fundamental reason for the change in its chemical composition during the maturation is the breaking and recombination of chemical bonds, manifested by the transformation in atomic hybridization based on quantum mechanics. While traditional methods are time-consuming and labor-intensive, machine learning technique has been introduced to clarify the relationship between hybridization and maturity. A kerogen maturity prediction model based on hybridization is constructed. The average error of the predicted values is only 4.91%, and more than 87% of the test samples have an error of less than 10%. The results demonstrate that the model can accurately predict the maturity of kerogen. As the evolution of kerogen maturity increases the proportion of sp^2 hybridized carbons, the orbital hybridization maturity index (OrbHMI) is proposed. The chemical changes in the thermal evolution and pyrolysis mechanism of kerogen can be explained and understood more essentially by OrbHMI. The results provide a basis for guiding artificial maturation and pave a promising path toward studying the kerogen structure and predicting hydrocarbon generating potential.

1. Introduction

Kerogen is the primary organic matter insoluble in organic solvents and gas-generating/oil-forming parent material in sedimentary rocks [1-3]. In addition, kerogen is a complex and heterogeneous material composed of an amorphous porous carbon skeleton [4,5] and its mechanical behavior and properties transform during its evolution [6,7]. The study of the amorphous porous state is conducive to understanding the adsorption and efficient exploitation of shale gas [8]. In general, there are three kerogen types based on van Krevelen diagrams [9] of the ratio of atomic hydrogen/carbon (H/C) to oxygen/carbon (O/C). Type I kerogen has a high initial H/C and a low O/C atomic ratio, whereas type II kerogen has a slightly high H/C ratio and low initial O/C ratio. Type III kerogen has a relatively low initial H/C ratio and high O/C atomic ratio [10]. Kerogen types are associated with the hydrocarbon generation capacity of kerogen, and machine learning (ML) technique has significant advantages for complex problems. Consequently, the ML approach that predicts the components and types of kerogen using nuclear magnetic resonance (NMR) spectroscopy has been developed [11]. However, the hydrocarbon generation capacity of kerogen is directly related to its maturity based on molecular structure analysis. Kerogen's thermal maturity can explain variations in the amount and composition of oil produced during thermal cracking; it is related to the main chemical and physical characteristics of the oil [12]. Research on kerogen is mainly based on experimental methods. In general, pyrolysis experiments are performed to construct kinetic models of kerogen [13,14], and spectroscopy methods are used to study the chemical structure of kerogen [15-17]. Various parameters have been developed to describe the thermal maturity [18,19], and maturity indicators were compared to the extent of oil generation [20]. Thus, the mechanism of maturation evolution is vital from both fundamental and practical aspects.

There are several indicators of maturity characterization. Vitrinite reflectance (VR) is one of the widely used indices for kerogen maturity to partly represent the hydrocarbon generation potential [9]. The measurement of VR is conducted mainly by experimental methods, while simulation methods such as molecular dynamics are used to investigate oil shale pyrolysis [21]. Some VR models are based on kinetics [22-27], and the studies focus on the relationship between the H/C atomic ratio and VR. It was found that the logarithm of VR relates to the H/C atomic ratio [28], and then the O/C atomic ratio was introduced into the relationship for improvement [25]. To analyze the kinetics of kerogen thermal evolution, a kerogen maturity index (molecule-maturity index,

MMI) based on the molecular composition during the evolution of kerogen was proposed [29], and this indicator can help intuitively display the pyrolysis potential of kerogen [30]. These models reveal the change in the carbon skeleton macroscopically and implicitly reflect the alteration of the chemical structure. However, in essence, the change in the chemical structure is caused by the transformation of chemical bonds. Moreover, from the perspective of quantum chemistry, the hybridization of atomic orbitals (AOs) is the intrinsic reason for forming chemical bonds. Therefore, it is crucial to establish an integrated model to combine kerogen maturity with hybridization and account for the underlying physical mechanism.

Hybridization is one of the most significant valence concepts [31]. The sp, sp^2, and sp^3 hybridization have been proposed as a mixture of s and p AOs based on quantum mechanics [32] and the bond is described by a pair of electrons in the AOs [33]. In addition to the formation of chemical bonds, the carbon hybridization state in the bonds is strongly associated with the corresponding bond distances [34]. Thus, this concept was applied to study the molecular structure of organic compounds [35]. The transformation of the chemical structure during the evolution of kerogen was investigated [36], and it was found that chemical reactions cause the change in the molecular composition due to the breaking and generation of chemical bonds [37]. Kerogen is the main organic component of shale with a large molecular weight and very complex structure. Thus, the study of kerogen structure and the mechanism of kerogen evolution upon maturation are vital and challenging. Knowledge of the correlation between the hybridization of AOs and kerogen maturity can help better understand the alteration of chemical bonds during the structural evolution of kerogen. Solid-state

NMR spectroscopy (especially the ^{13}C NMR technique) has become a non-destructive technique for providing genuine chemical structure details of kerogen [3]. The NMR spectra are used to determine the electronic structure of molecules [38], and as one of the chemical structure details, the hybridization state of carbon is indicated by ^{13}C NMR [39]. Chemical shifts are closely related to functional groups, and the correlation between chemical shifts and carbon types has been analyzed [40]. Carbon types can be generally divided into aromatic/aliphatic carbons [41,42] or sp^2/sp^3 carbons [43]. By employing the ^{13}C solid-state NMR technique, thermal maturity increases when kerogen trends toward condensed aromatic structures, and the kerogen types are determined via ^{13}C cross-polarization/magic-angle spinning (CP/MAS) [44-46]. It is suggested that the aliphaticity of kerogen decreases while the aromaticity and carboxyl content increase when the type of kerogen transforms from type I into type III [47]. The change in the chemical structure of kerogen during maturation was also studied via NMR spectroscopy [48], and the results show that the trend is similar to the transformation from type I into type III. The sp^2/sp^3 hybridization ratio increases during the maturation of kerogen [4]. Therefore, exploring the transformation of the hybridization state during maturation is significant both in research and engineering.

In this study, we develop a model for characterizing kerogen maturity based on hybridization and quantum mechanics (Fig. 1). Since the traditional methods are time-consuming and labor-intensive, the ML technique is used to identify kerogen maturity by hybridization owing to its high efficiency. The hybridized atoms are obtained via ^{13}C solid-state NMR spectroscopy, and the ML methods are applied to predict the kerogen maturity compared with the MMI. A new kerogen maturity
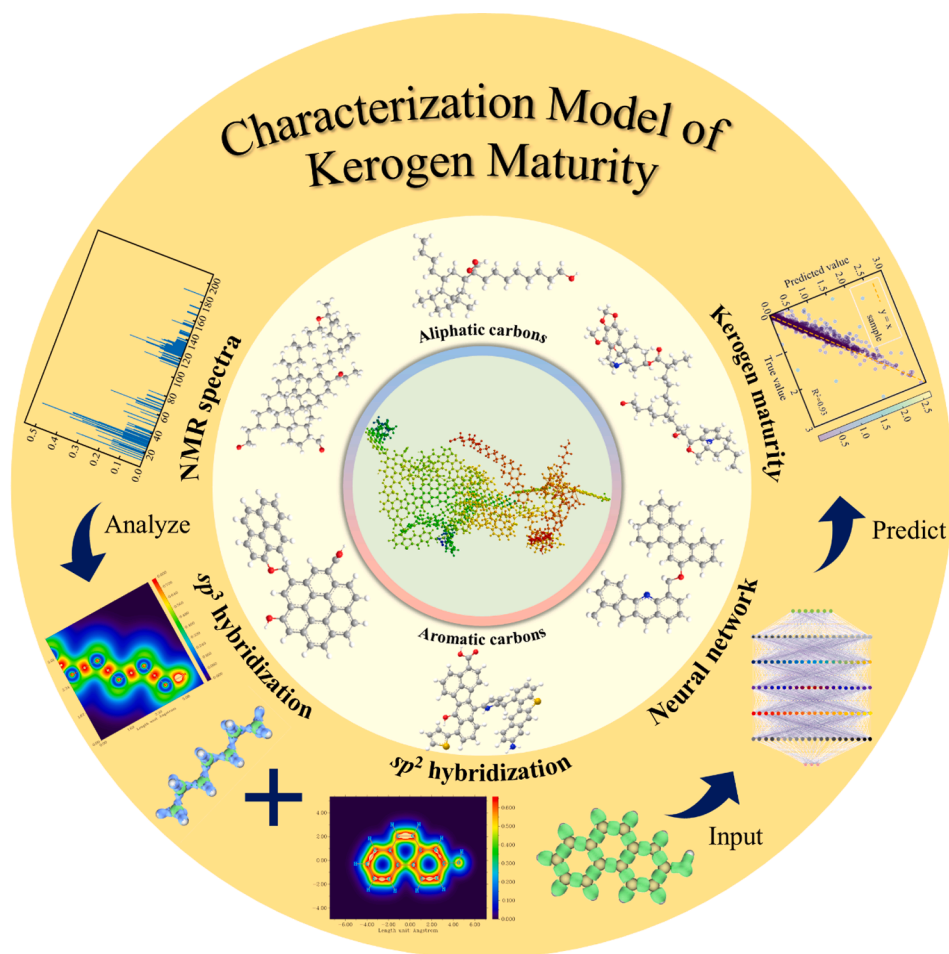

Fig. 1. Schematic of the characterization model. The chemical shift of the functional group containing sp^2 hybridized carbon is generally higher than that containing sp^3 hybridized carbon, and the kerogen with high maturity contained more sp^2 aromatic carbons than the kerogen with low maturity.

index is proposed to quantify the relationship between kerogen maturity and hybridized atoms. The average error of the predicted values of the test set and analyzed samples is quantified to determine model accuracy. The linear relationship of predicted results of the ML model with the index is profiled to determine if the index can be used to characterize kerogen maturity quantitatively. The change in chemical structure and bonds in kerogen thermal evolution is elucidated based on the model and indicator. In addition, the relationship between hybridization, molecular structure, and maturity has been discussed. Therefore, this study may resolve issues concerning artificial maturation. Additionally, this study provides further understanding of the fundamental mechanism of kerogen maturity evolution and promotes the study of the kerogen structure and the prediction of oil/gas generation.

2. Methodology

2.1. Nuclear magnetic resonance (NMR)

NMR was developed in the 1950s and has been widely used to study organic chemical structures [49-53]. The commonly used NMR spectra include ^{1}H and ^{13}C spectra. The ^{13}C spectrum has a more comprehensive chemical shift range than the ^{1}H spectrum, and it is more accurate for the study of kerogen OM structures. The carbon types are generally classified as aromatic and aliphatic carbon. Most aromatic carbon is sp^2 hybridized, while most aliphatic carbon is sp^3 hybridized; thus, the chemical structure of the hybridization information is obtained. ^{13}C NMR is used to obtain information on the sp^2 and sp^3 hybridized carbons in kerogen samples. It can be concluded that the chemical shift of the sp^2 hybridized carbons is greater than that of sp^3 hybridized carbons [43]. This research has great significance for the subsequent study of kerogen molecular structure.

2.2. Hybridization

To describe an atomistic system in detail, information on the charges $(G_1, G_2, ..., G_n)$ and positions $(r_1, r_2, ..., r_n)$ of the nucleus in the Hilbert space is necessary [54]. Wavefunctions are typically used to represent AOs, which can be obtained from the Schrödinger equation as follows:

$$i\hbar \frac{\partial}{\partial t}|\psi\rangle = \widehat{H}|\psi\rangle, \tag{1}$$

where ψ represents the wavefunction, $\widehat{H}$ is the Hamiltonian operator, and $\hbar$ is the reduced Planck's constant. The linear combination of AOs is called a hybrid orbital. As the s and p eigenfunctions are introduced to explain the formation of hybrid orbitals [32], the wavefunctions can be separated into two parts:

$$\psi_{n0}(r, \theta, \varphi) = R_{n0}(r) \cdot s(\theta, \varphi), \tag{2}$$

$$\psi_{n1c}(r, \theta, \varphi) = R_{n1}(r) \cdot p_c(\theta, \varphi), \tag{3}$$

where R is the radial function, s and p eigenfunctions that depend only on θ and φ, and c can take values x, y, and z. The sp^2 hybridization is a linear combination of the s and p_x, p_y orbitals, forming three new hybrid orbitals. The wavefunctions of the hybrid orbitals are orthogonal and equivalent, and the other two wavefunctions can be expressed as angular dependence by rotation operators. Thus, we can only consider ψ_I describing the hybridization of AOs. The origin can be regarded as the position of the sp^2 hybridized atom.

For an atomistic system including n atoms, every wavefunction of the orbital should be considered. In the same hybridization state, the wavefunctions have similar forms, and the expressions are the same in their respective relative coordinate systems. In the global coordinate system, they are generally related to the global coordinates of the central

atom. The wavefunction ψ_I has rotation symmetry relative to the *Ox* axis under the relative coordinates, confirmed by the *Ox* axis and origin. The chemical bond formed between adjacent carbon atoms coincides with the symmetry axis of the hybrid orbital. It allows the displacement vector between the two carbon atoms to act as the *Ox* axis and the starting point carbon atom as the origin of the relative coordinates. Therefore, a set of wavefunctions of the hybrid orbitals can be determined. The number of nuclear charges in the atomistic system determines the type of atom and the spatial position to determine the chemical bond formation. The unit displacement vector $\Delta \widehat{r} = \Delta r / \|\Delta r\|$ is under the global coordinate system, while the wavefunction is under the relative coordinate system. The coordinate transformation is considered $r' = Tr$, with relative coordinates vector $r' \in \mathbb{R}^3$, coordinate transformation matrix, $T \in \mathbb{R}^{3 \times 3}$ and global coordinates $r \in \mathbb{R}^3$. Then, the unit vector $\Delta \widehat{r}' \in \mathbb{R}^3$ of the *Ox* axis under the relative coordinate system is written as

$$\Delta \widehat{r}' = \frac{\Delta r'}{\|\Delta r'\|} = T \frac{\Delta r}{\|\Delta r\|} = T \Delta \widehat{r}, \tag{4}$$

without a change in the vector modulus, by substituting $s = 1/\sqrt{4\pi}$ and $p_x = \sqrt{3/4\pi} \sin\theta \cos\varphi$ into ψ_I, the wavefunctions can be substituted as follows:

$$\psi_I = \frac{1}{\sqrt{2\pi}} \left(\frac{1}{\sqrt{6}} + \|r'\| r'^* \Delta \widehat{r}' \right), \tag{5}$$

where r'^* satisfies the conjugation condition $r'^* r' = 1$. To relate the wavefunction to the coordinates of the spatial position, Eq. (5) is converted into the expression of the global coordinates

$$\psi_I = \frac{1}{\sqrt{2\pi}} \left(\frac{1}{\sqrt{6}} + \|r\| r^* T^{-1} T \Delta \widehat{r} \right) = \frac{1}{\sqrt{2\pi}} \left(\frac{1}{\sqrt{6}} + \|r\| r^* \Delta \widehat{r} \right) \tag{6}$$

Other wavefunctions can be obtained by rotating around the *Oz* axis. Then, the wavefunctions of the sp^2 hybridization in the atomistic system can be regarded as a set of ψ_i, and the sp^3 hybridization is performed analogously. Every carbon atom corresponding to one wavefunction can obtain:

$$\Psi = \left\{ \left(\psi_{sp^2}^1, \psi_{sp^2}^2, ..., \psi_{sp^2}^m \right), \left(\psi_{sp^3}^1, \psi_{sp^3}^2, ..., \psi_{sp^3}^k \right) \right\}, \tag{7}$$

where ψ_{sp^2} represents the wavefunction of the sp^2 hybridization state, and ψ_{sp^3} represents the wavefunction of the sp^3 hybridization state, m is the total number of sp^2 hybridized carbons, and k is the total number of sp^3 hybridized carbons. Since the sp hybridization corresponds to the carbon-carbon triple bond and takes up a very small proportion of kerogen molecules, it is ignored. Only the influence of sp^2 and sp^3 hybridization on kerogen maturity is studied. The hybridization state can be obtained from NMR spectra based on the chemical shift ranges that vary according to the different types of functional groups. The carbon types are divided into sp^2 and sp^3 carbons, which are more suitable than aromatic/aliphatic carbon. Then, with the increase in sp^2 hybridized carbons, the maturity of kerogen increases accordingly. Hence, the ^{13}C NMR spectra with detailed structural information can characterize kerogen maturity through hybridization.

2.3. Model

Two major ML algorithms are applied in engineering, supervised and unsupervised learning. For regression problems, supervised learning is frequently adopted, and features are artificially extracted as sample labels. The dataset is divided into training, validation, and test sets, and the test set is used to evaluate the model's generalization ability. There

are two main challenges in model training; underfitting means that the training error is not sufficiently low, while overfitting implies that the gap between the training error and test error is too large. To solve the underfitting problem, adding new features or applying other activation functions and optimizers can be realized. To avoid overfitting a very important method is to introduce regularization to reduce the test error. In addition, early stopping can effectively avoid overfitting.

The optimization algorithm used in the model is backpropagation, which is simple and effective for ML. The backpropagation algorithm is used to calculate the representation in each layer from the representation in the previous layer to indicate how the machine should change its internal parameters [55].

The high-throughput ML model applied is the multilayer perceptron, which is popular for classification and regression. Linear layers are written as an affine transformation:

$$Z(x) = W^{\mathrm{T}} x + b, \tag{8}$$

where $x \in \mathbb{R}^{D_{in}}$ represents the inputs, $W \in \mathbb{R}^{D_{in} \times D_{out}}$ is the weight matrix, and $b \in \mathbb{R}^{D_{out}}$ is the bias parameter. The mean absolute error (MAE) is applied to define the effects of the neural network model. For nonlinear problems, an activation function is introduced to address them. We use R^2 to describe the goodness of fit for the regression model [56]:

$$R^2 = 1 - \sum_i \left(y_i - y_i'\right)^2 \Big/ \sum_i \left(y_i - M\right)^2, \tag{9}$$

where y_i is the true value, y_i' is the predictive value, and M is the mean value. There are two models in our research: the first one predicts the hybridized atoms and chemical bonds from ^{13}C NMR spectra while the other realizes the mapping between the hybridization of AOs and maturity. A flow chart of the study is shown in Fig. 2.

The optimization algorithm uses backpropagation and stochastic gradient descent (SGD), which is simple and effective. First, define the cost error

$$J(\Theta) = \frac{1}{m} \sum_{i=1}^{m} L\left(x^{(i)}, y^{(i)}, \Theta\right), \tag{10}$$

where m is the total number of samples, x is the input, y is the output, Θ is the parameter, and L is the loss function. Then, the gradient can be calculated according to the chain rule:

$$\begin{aligned}
\nabla_{\Theta} J(\Theta) &= \frac{1}{m} \sum_{i=1}^{m} \frac{\partial L}{\partial \Theta} = \frac{1}{m} \sum_{i=1}^{m} \frac{\partial L}{\partial Z^{l+1}} \frac{\partial Z^{l+1}}{\partial A^l} \frac{\partial A^l}{\partial Z^l} \frac{\partial Z^l}{\partial \Theta} \\
&= \frac{1}{m} \sum_{i=1}^{m} \left(\left(W^{l+1}\right)^{\mathrm{T}} \frac{\partial L}{\partial Z^{l+1}} \right) \odot \sigma'\left(Z^l\right) \frac{\partial Z^l}{\partial \Theta},
\end{aligned} \tag{11}$$

where $\odot$ is the Hadamard product, where each element is the product of the elements of the original two matrices. Then, calculate the gradient of weight W and bias b:
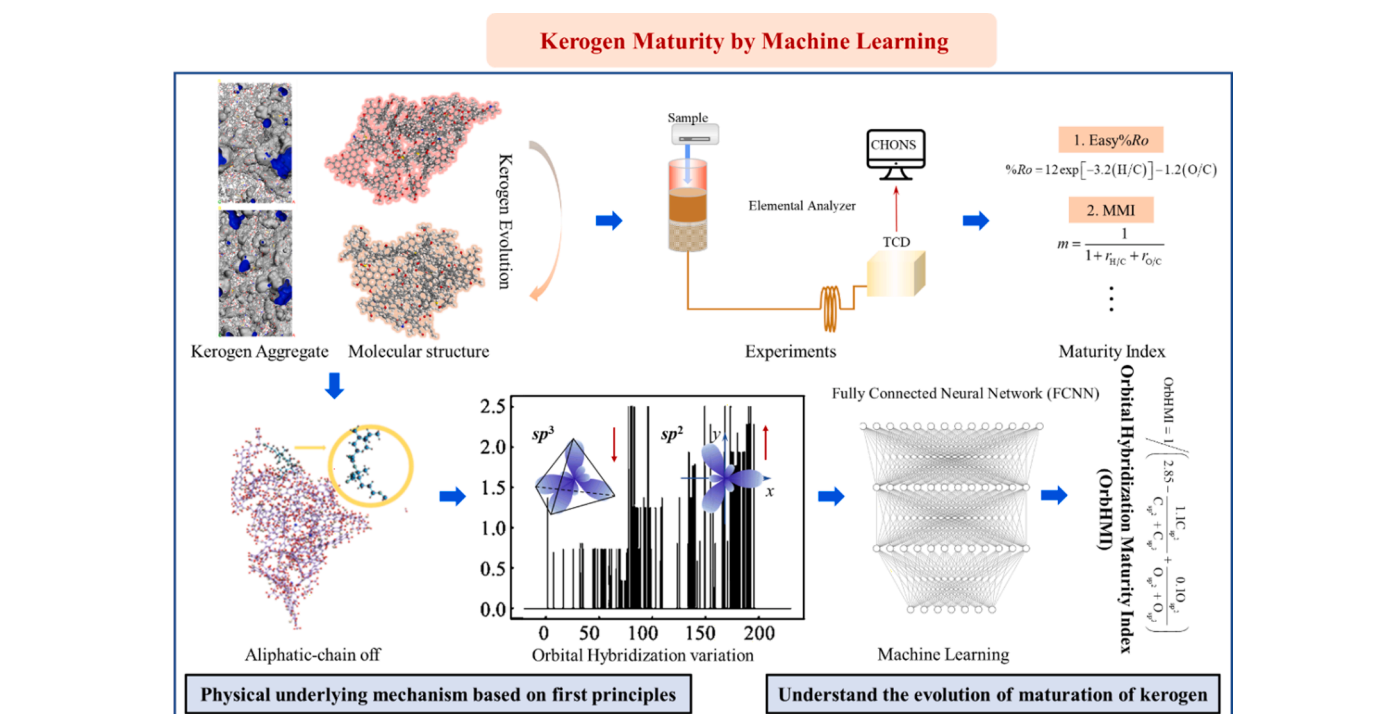
$$\begin{aligned}
\nabla_{W^l} J\left(W^l\right) &= \frac{1}{m} \sum_{i=1}^{m} \frac{\partial L}{\partial W^l} = \frac{1}{m} \sum_{i=1}^{m} \frac{\partial L}{\partial Z^l} \frac{\partial Z^l}{\partial W^l} = \frac{1}{m} \sum_{i=1}^{m} \frac{\partial L}{\partial Z^l} \frac{\partial\left(W^l A^{l-1} + b^l\right)}{\partial W^l} = \frac{1}{m} \sum_{i=1}^{m} \frac{\partial L}{\partial Z^l} \left(A^{l-1}\right)^{\mathrm{T}}, \\
\nabla_{b^l} J\left(b^l\right) &= \frac{1}{m} \sum_{i=1}^{m} \frac{\partial L}{\partial b^l} = \frac{1}{m} \sum_{i=1}^{m} \frac{\partial L}{\partial Z^l} \frac{\partial Z^l}{\partial b^l} = \frac{1}{m} \sum_{i=1}^{m} \frac{\partial L}{\partial Z^l} \frac{\partial\left(W^l A^{l-1} + b^l\right)}{\partial b^l} = \frac{1}{m} \sum_{i=1}^{m} \frac{\partial L}{\partial Z^l},
\end{aligned} \tag{12}$$

Fig. 2. Flow chart of predicting kerogen maturity by ML method.

and use SGD to update

$$W_{t+1}^l = W_t^l - \eta \nabla_{W^l} J(W^l) = W_t^l - \eta \frac{1}{m'} \sum_{i=1}^{m'} \frac{\partial L}{\partial Z^l} (A^{l-1})^{\mathrm{T}},$$

(13)

$$b_{t+1}^l = b_t^l - \eta \nabla_{b^l} J(b^l) = b_t^l - \eta \frac{1}{m'} \sum_{i=1}^{m'} \frac{\partial L}{\partial Z^l},$$

where η is the learning rate, t represents the iterative time step, and m' denotes the number of mini-batches. In this research, the widely applied framework TensorFlow is chosen [57].

2.4. Dataset description

The molecular weight of kerogen is too large, and therefore, it is impractical to utilize a large number of kerogen samples for training directly. However, kerogen molecules are composed of many small functional groups, and their properties are also affected by various functional groups [58]. Furthermore, functional groups contribute to genetic potential [59], which is indicated by maturity. Therefore, the sample set collects small molecules with the same functional groups as the kerogen molecules for research. The basic chemical organization mechanisms between kerogen molecules and these small molecules are same, and machine learning models can learn these mechanisms and use them for prediction. Many of these molecules are the products of the thermal cracking of kerogen or similar substances, such as ethers [60], alkenes [61], alkanes [62], naphthalenes [63], anthracenes [64], and phenanthrenes [65]. Based on this, we establish 90,000 samples for training and testing. According to van Krevelen diagrams [9], the samples here contain types I, II, and III, and part of them have been successfully applied to predict kerogen types and skeleton components [11]. The molecular structure data are obtained from PubChem [66] and laboratories. The ^{13}C NMR spectra are calculated using MestReNova 14 [67], and the information of hybridized atoms and chemical bonds can be obtained using the open-source cheminformatics toolkit RDKit [68]. The Easy%Ro [24] and MMI [29] are directly related to the H/C and O/C, and the MMI is defined as

$$MMI = \frac{1}{1 + r_{H/C} + r_{O/C}},$$

(14)

where 1, $r_{H/C}$, and $r_{O/C}$ represent the atomic ratios of C/C, H/C, and O/C, respectively. When the MMI value is 1, the kerogen is completely carbonized. Since the MMI has a more straightforward expression than Easy%Ro, we choose the MMI as our maturity label. In addition, the MMI has a linear relationship with the VR derived from experimental data [69-72], so this model can also be applied to predict traditional maturity indicators.

3. Results and discussion

3.1. Comparison of ML models

To identify an ML model with the best performance, the ML models with different optimizers and activation functions are compared. The MAE and R^2 as the metrics for the model selection criteria are chosen. The model's inputs are the hybridization state of atoms and the number of chemical bonds and the output is the maturity. We select three popular nonlinear activation functions: rectified linear unit (ReLU), tanh, and sigmoid. From the results of Fig. 3, we can conclude that ReLU has the best performance in the model, compared with the other two activation functions. We choose adaptive moment estimation (Adam), Nadam, Adamax, SGD, and root mean square prop (RMSprop). Adam is an optimization algorithm of stochastic objective functions that are computationally efficient and suitable for big data problems [73], and Nadam and Adamax are variants of Adam. As indicated by Zou et al. [74], the learning rate of the optimizer can significantly influence the performance; hence, we set the learning rate of all optimizers to 0.001. For Adam, Adamax and Nadam, the exponential decay rates for the moment estimates, are 0.9 and 0.999, respectively. The rate of decay of SGD is 0 while that of RMSprop is 0.9. Although SGD performs similarly to Nadam when the activation function is ReLU, Nadam performs better than SGD with other activation functions. Thus, Nadam is more stable. Moreover, Nadam and Adamax perform better than Adam in general. Therefore, Nadam is chosen as the optimizer.

We then consider the influence of different inputs (Table 1). When the inputs are only the hybridization state of atoms or the number of chemical bonds, the model's performance decreases. It suggested that kerogen maturity is closely related to both the hybridization state and the chemical bonds. Therefore, to predict kerogen maturity, we need information on the hybridized atoms and chemical bonds, which can be obtained from the ^{13}C NMR spectra.

3.2. Obtaining the detailed structural information from NMR

The model database for obtaining detailed structural information consists of 90,000 samples, 70,000 of which are used as the training set. By contrast, the residual samples are divided equally into the validation and test sets. The validation data are used to adjust the parameters of the model, whereas the test data are used to evaluate the generalization

Table 1
Different predictive performances of various inputs.

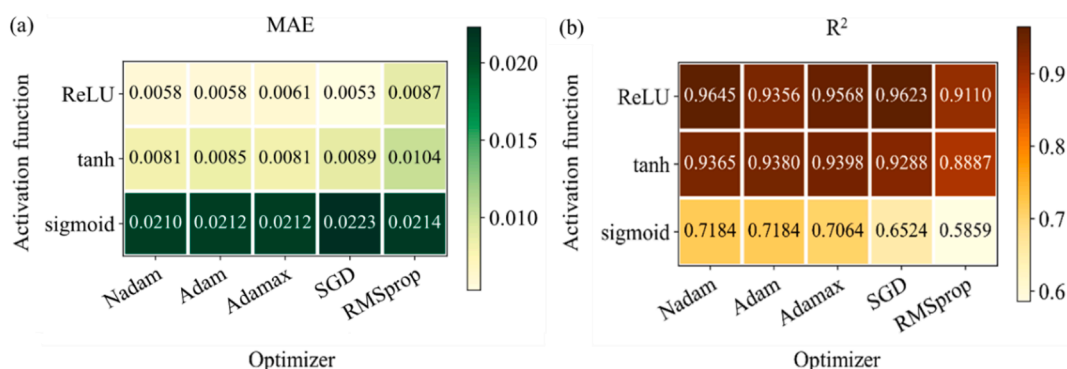| Inputs	Mean absolute error (MAE)	R-squared (R^2)
C_{sp^2}, C_{sp^3}, O_{sp^2}, O_{sp^3}	0.0194	0.7396
B_1, B_2, $B_{1.5}$	0.0241	0.6261
C_{sp^2}, C_{sp^3}, O_{sp^2}, O_{sp^3}, B_1, B_2, $B_{1.5}$	0.0058	0.9645

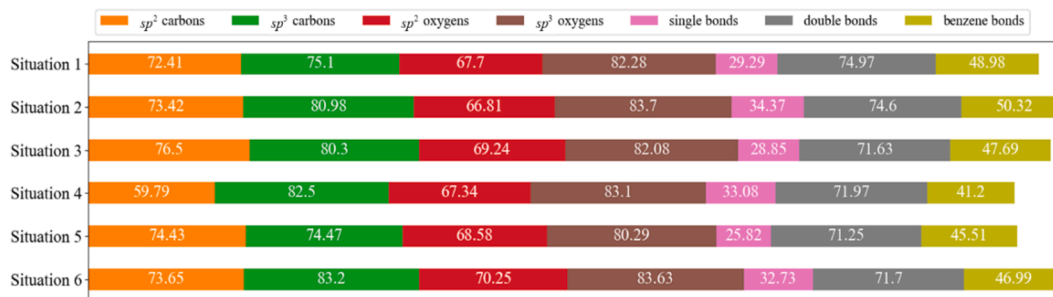**Fig. 3.** Comparison of model performances of different activation functions and optimizers.

Fig. 4. Predictive performances of the hybridization state and chemical bonds (unit: percentage).

ability of the model. The inputs are the ^{13}C NMR spectra results, which contain detailed information about the chemical structure. The outputs are the number of hybridized atoms and chemical bonds, including sp^2/ sp^3 hybridized carbons, sp^2/sp^3 hybridized oxygens, single bonds, double bonds, and benzene C-C bonds.

Six situations with different batch sizes and initial learning rates are considered. The batch sizes are 200, 500, and 1000, while the initial learning rates are 0.001 and 0.0001. For a larger batch size and a lower initial learning rate, the model is difficult to converge. The accuracy of each parameter is counted and displayed as a percentage in Fig. 4. The predictive performance of the hybridization of AOs is better than that of chemical bonds. Therefore, we mainly focus on the hybridization characterization and propose a new index of kerogen maturity based on hybridization wavefunctions.

3.3. Predicting maturity by hybridization

3.3.1. Orbital hybridization maturity index

The ^{13}C NMR spectra contain information on the chemical structure, while the charges and spatial positions of atoms can determine the hybridization state. From Eq. (6), it can be known that how wavefunctions relate to spatial positions. Then, all the sp^2 hybridization wavefunctions are counted and divided into sp^2 hybridized carbons and sp^2 hybridized oxygens based on the charges of the hybrid atoms. The expressions can be written as follows:

$$C_{sp^2} = \sum_m \left(a_m r_m^* \Delta r_m + b_m\right) = \sum_m G_m^C \psi_{sp^2}^{Cm},$$
$$O_{sp^2} = \sum_i \left(c_i r_i^* \Delta r_i + d_i\right) = \sum_i G_i^O \psi_{sp^2}^{Oi}, \quad (15)$$

where a, b, c, d are constants, G^C, G^O represents the coefficient of the wavefunction of the sp^2 hybridized carbon $\psi_{sp^2}^C$ and the wavefunction of the sp^2 hybridized oxygen $\psi_{sp^2}^O$, respectively. Analogously, C_{sp^3} and O_{sp^3} can be written as follows:

$$C_{sp^3} = \sum_k \left(\alpha_k r_k^* \Delta r_k + \beta_k\right) = \sum_k G_k^C \psi_{sp^3}^{Ck},$$
$$O_{sp^3} = \sum_j \left(\gamma_j r_j^* \Delta r_j + \varsigma_j\right) = \sum_j G_j^O \psi_{sp^3}^{Oj}, \quad (16)$$

where α, β, γ, ζ are constants, $\psi_{sp^3}^C$ and $\psi_{sp^3}^O$ represent the wavefunction of the sp^3 hybridized carbon and oxygen, respectively.

To determine the relationship between the hybridization state and kerogen maturity, the H/C atomic ratio is first characterized by the carbon hybridization state. We use Eq. (17) to represent the ratio of the sp^2 hybridized carbons and fit with H/C, and their relation can be written as Eq. (18), which shows a linear relationship

$$r_C = C_{sp^2} / \left(C_{sp^2} + C_{sp^3}\right), \quad (17)$$

$$H/C = -1.1 r_C + 1.85. \quad (18)$$

From the expression, it can be concluded that r_C is proportional to the reciprocal of the H/C atomic ratio, which indicates that with the increase in the sp^2 hybridized carbons, H/C decreases, which is consistent with theory and experiment. Similarly, we define the ratio of the sp^2 hybridized oxygen atoms using Eq. (19), and then obtain an index called the orbital hybridization maturity index (OrbHMI) using Eq. (20)

$$r_O = O_{sp^2} / \left(O_{sp^2} + O_{sp^3}\right), \quad (19)$$

$$\text{OrbHMI} = \frac{1}{2.85 - 1.1 r_C + 0.1 r_O}. \quad (20)$$

As r_O has little impact on maturity, the hybridization state of carbon is dominant when observing the process of kerogen maturation. From Eq. (20), we can conclude that as the content of sp^2 hybridized carbons increases, the maturity increases. This equation reveals the mechanism of kerogen thermal maturation.

3.3.2. Training of the model

The fully connected neural network from Section 2.3 for predicting kerogen maturity has multiple inputs and one output. The output is maturity, in contrast to the maturity indicators mentioned in the literature. The inputs include the number of sp^2 hybridized carbons (C_{sp^2}), sp^3 hybridized carbons (C_{sp^3}), sp^2 hybridized oxygens (O_{sp^2}), sp^3 hybridized oxygen atoms (O_{sp^3}), single bonds (B_1), double bonds (B_2), and benzene C-C bonds ($B_{1.5}$). The output data are calculated using Eq. (14). We set the initial learning rate at 0.001, the batch size at 1000, and an early stopping method is applied to prevent overfitting. The distribution of the dataset based on sp^2 and sp^3 hybridized carbons is shown in Fig. 5. It can be seen that the distributions of the training and validation sets are similar.

The training process of the ML model is shown in Fig. 6a, and both the training and generalization errors are low. Therefore, the developed model can demonstrate the relationship between the hybridization of AOs and kerogen maturity under a deficient error. The inputs of the test set are obtained from the ^{13}C NMR spectra, and the predicted values after training fit well with the labels (Fig. 6b), showing that the model can predict kerogen maturity accurately.

3.3.3. Comparison of OrbHMI and model

To visualize the relationship between OrbHMI and MMI, 300 samples are selected from the test set. The predicted value is obtained from similar samples, and the comparison is shown in Fig. 7. The figure shows that the OrbHMI is equivalent to the MMI, and R^2 indicates the degree of fit of the predicted results and the actual results. The MMI is a continuous representation of maturation and can be compared with traditional maturity indicators. Similarly, the predicted value is equivalent to OrbHMI. The index shows how hybridization impacts maturity, which is from the perspective of quantum. Since hybridization is associated with chemical bonds, the law of transformation from immature to mature kerogen can be obtained. The sp^2 hybridized carbons represent the C-C double bonds and benzene C-C bonds, while the sp^3 carbons represent
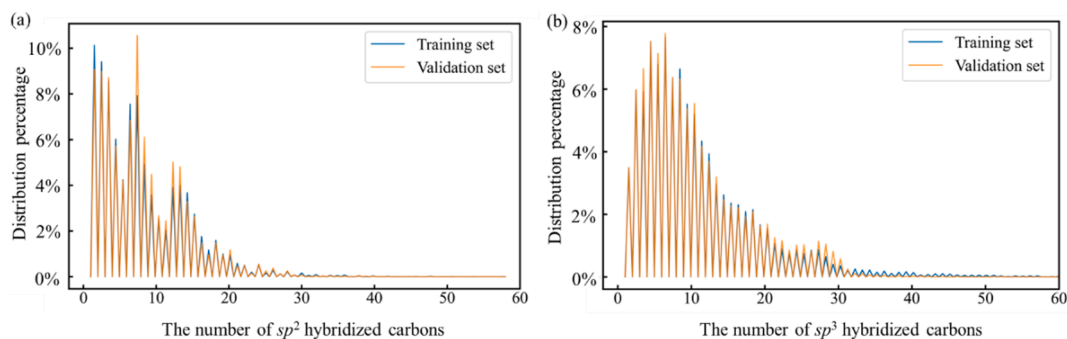
Fig. 5. Data distribution of the database of the training set and validation set. (a) The distribution of sp^2 hybridized carbons. (b) The distribution of sp^3 hybridized carbons.
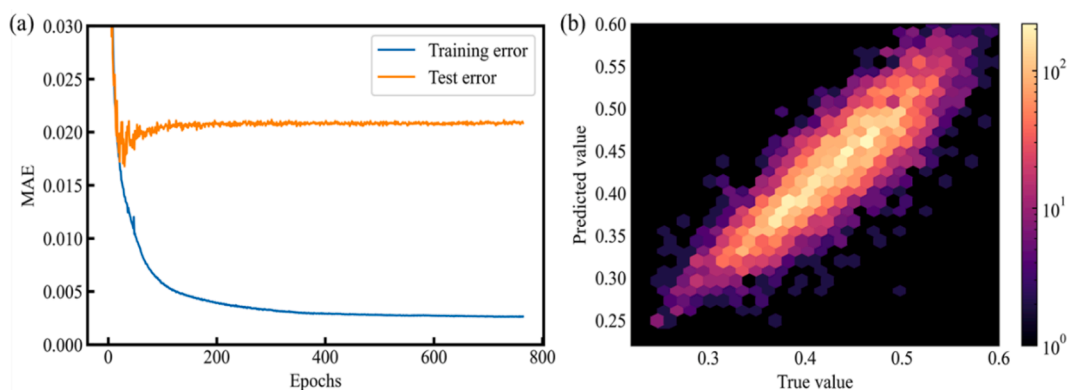

Fig. 6. Performances of the ML model. (a) Training process based on the training error and test error. (b) Comparison of predicted value and true value and the color level represents the density of the samples.
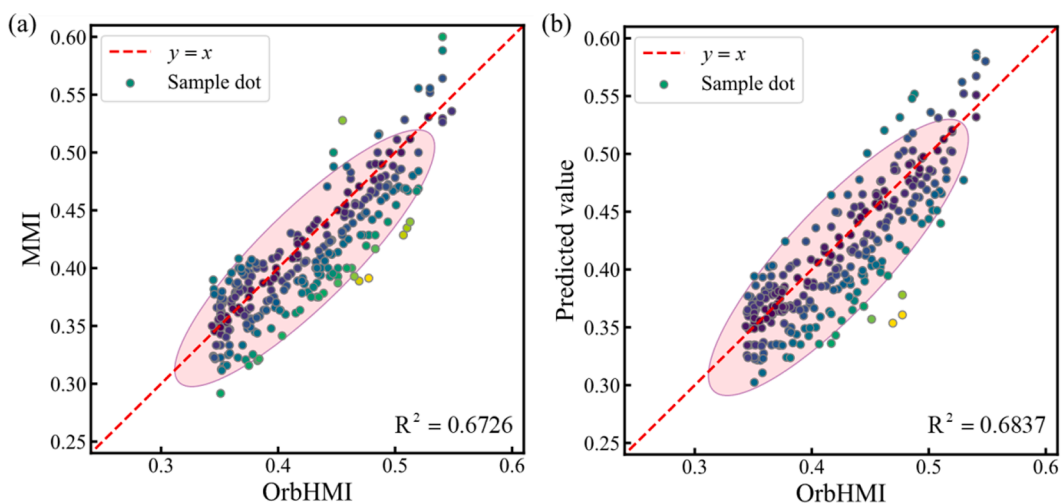

Fig. 7. Comparison of the OrbHMI with the MMI and the predicted value. The gray ovals are the confidence ellipse of the covariance while the number of standard deviations is two.

the C-C single bonds. The sp^2 hybridization implies high H/C, while sp^3 hybridization represents low H/C. As the maturity of kerogen increases, the proportion of the sp^2 hybridized carbons increases, and the sp^3 hybridized carbons break and reform the sp^2 hybridized carbons. Therefore, the H/C atomic ratio decreases. This also indicates that the type of kerogen is from type III to I with the improvement in the hydrocarbon generation capacity.

3.3.4. Error percentage

The average error between the final predicted kerogen maturity and the actual value is less than 5%, and more than 87% of the samples have an error of less than 10%. The samples in the test set are divided into five parts according to the proportion of sp^2 hybridized carbons. The average and standard errors of the predicted values are compared, and the results are shown in Fig. 8. It can be concluded that the average errors under different r_C distributions are not significantly different, and the standard errors are minimal, indicating that most of the errors are concentrated
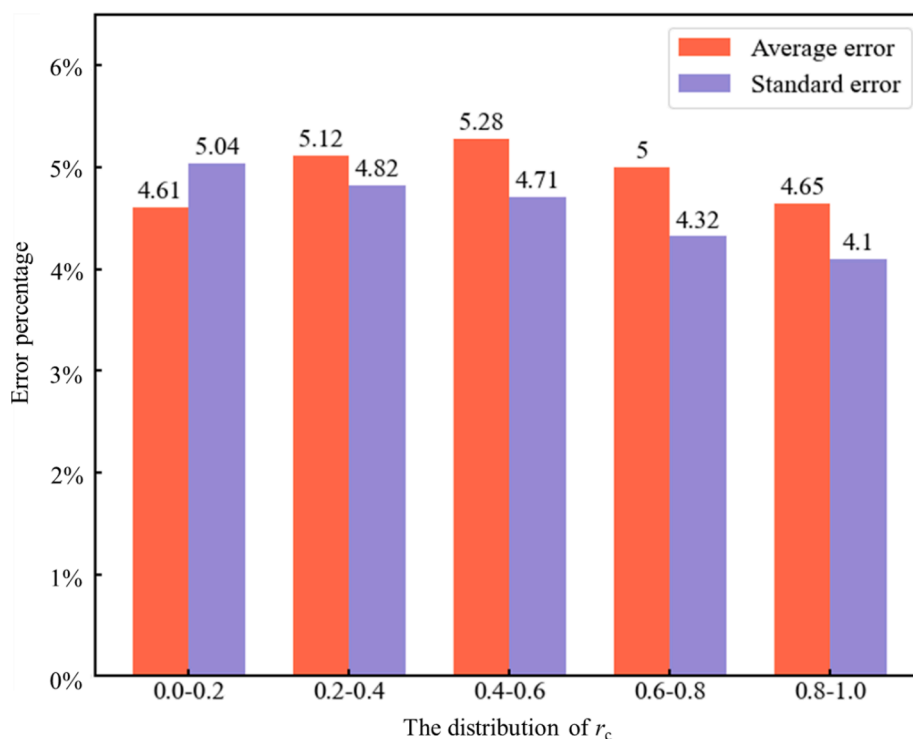
Fig. 8. Percentage of the average and standard error by the distribution of r_C.

around the average errors. Most of the errors are within 20%, indicating that the prediction model has high accuracy. Therefore, the results may provide an accurate ML model for kerogen maturity prediction via ^{13}C NMR spectra.

4. Conclusions

We propose the OrbHMI, a maturity index based on orbital hybridization, to calculate kerogen maturity. With the increase in sp^2 hybridized carbons, OrbHMI increases, and kerogen trends toward aromatic structures. From the perspective of quantum mechanics, chemical changes in the evolution of kerogen are discussed, as well as the breaking and formation of chemical bonds during chemical reactions. This laid the foundation for further understanding of kerogen evolution and hydrocarbon generation mechanisms.

An ML model based on NMR spectra is established to predict the maturity of kerogen. There are a total of 90,000 samples, of which 70,000 samples are used for training, 10,000 for validation, and 10,000 for testing. The hybridized atoms obtained from the NMR spectra are applied to predict kerogen maturity. The average error of the predicted values of the samples used for the test is 4.91%, indicating that the model has good generalization ability and can be directly applied to the prediction of kerogen maturity. The standard error of the prediction error is 4.49%, indicating that most of the errors are close to the average values. More than 99% of the data have an error of less than 20%, and over 87% of the data have an error of less than 10%, indicating that the prediction results of the model are highly accurate, benefited by the ML technique.

Based on the ML model and maturity index OrbHMI, the quantitative characterization of kerogen maturity through hybridized atoms is realized. The chemical changes in the evolution of kerogen are explained from the perspective of the underlying physical mechanism. To further study the potential of kerogen for hydrocarbon production and lay the foundation for the production of shale oil and gas, the relationship between orbital hybridization, molecular structure, and kerogen maturity is clarified. The results contribute to the artificial maturation of kerogen

and an understanding of the mechanism of kerogen oil and gas production.

CRediT authorship contribution statement

Jun Ma: Investigation, Methodology, Software, Formal analysis, Visualization, Writing - original draft. **Dongliang Kang:** Investigation, Methodology, Software, Data curation. **Xiaohe Wang:** Investigation, Methodology. **Ya-Pu Zhao:** Conceptualization, Resources, Funding acquisition, Supervision, Project administration.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was jointly supported by the National Natural Science Foundation of China (NSFC, Grant No. 12032019, 11872363, 51861145314), the Chinese Academy of Sciences (CAS) Key Research Program of Frontier Sciences (Grant No. QYZDJ-SSW-JSC019), and the CAS Strategic Priority Research Program (Grant No. XDB22040401).

References

[1] Wang X, Huang X, Lin K, Zhao Y-P. The Constructions and Pyrolysis of 3D Kerogen Macromolecular Models: Experiments and Simulations. Glob Challenges 2019;3(5): 1900006. https://doi.org/10.1002/gch2.v3.510.1002/gch2.201900006.
[2] Lai D, Zhan JH, Tian Y, Gao S, Xu G. Mechanism of kerogen pyrolysis in terms of chemical structure transformation. Fuel 2017;199:504–11. https://doi.org/10.1016/j.fuel.2017.03.013.
[3] Cao X, Yang J, Mao J. Characterization of kerogen using solid-state nuclear magnetic resonance spectroscopy: A review. Int J Coal Geol 2013;108:83–90. https://doi.org/10.1016/j.coal.2011.12.001.
[4] Bousige C, Ghimbeu CM, Vix-Guterl C, Pomerantz AE, Suleimenova A, Vaughan G, et al. Realistic molecular model of kerogen's nanostructure. Nat Mater 2016;15(5): 576–82. https://doi.org/10.1038/nmat4541.

[5] Psarras P, Holmes R, Vishal V, Wilcox J. Methane and CO2 Adsorption Capacities of Kerogen in the Eagle Ford Shale from Molecular Simulation. Acc Chem Res 2017; 50(8):1818–28. https://doi.org/10.1021/acs.accounts.7b0000310.1021/acs.accounts.7b00003.s001.

[6] Wang X, Huang X, Gao M, Zhao Y-P. Mechanical response of kerogen at high strain rates. Int J Impact Eng 2021;155:103905. https://doi.org/10.1016/j.ijimpeng.2021.103905.

[7] Emmanuel S, Eliyahu M, Day-Stirrat RJ, Hofmann R, Macaulay CI. Impact of thermal maturation on nano-scale elastic properties of organic matter in shales. Mar Pet Geol 2016;70:175–84. https://doi.org/10.1016/j.marpetgeo.2015.12.001.

[8] Huang X, Zhao YP. Characterization of pore structure, gas adsorption, and spontaneous imbibition in shale gas reservoirs. J Pet Sci Eng 2017;159:197–204. https://doi.org/10.1016/j.petrol.2017.09.010.

[9] Krevelen D. Coal: Typology, Chemistry, Physics. ELSEVIER: Constitution; 1961.

[10] Tissot BP, Welte DH. Petroleum formation and Occurance. Saudi Med J 2012;33: 3–8.

[11] Kang D, Wang X, Zheng X, Zhao Y-P. Predicting the components and types of kerogen in shale by combining machine learning with NMR spectra. Fuel 2021; 290:120006. https://doi.org/10.1016/j.fuel.2020.120006.

[12] Vandenbroucke M, Largeau C. Kerogen origin, evolution and structure. Org Geochem 2007;38(5):719–833. https://doi.org/10.1016/j.orggeochem.2007.01.001.

[13] Behar F, Kressmann S, Rudkiewicz JL, Vandenbroucke M. Experimental simulation in a confined system and kinetic modelling of kerogen and oil cracking. Org Geochem 1992;19(1-3):173–89. https://doi.org/10.1016/0146-6380(92)90035-V.

[14] Aboulkas A, El HK. Study of the kinetics and mechanisms of thermal decomposition of Moroccan Tarfaya oil shale and its kerogen. Oil Shale 2008;25:426–43. https://doi.org/10.3176/oil.2008.4.04.

[15] Czaja AD, Kudryavtsev AB, Cody GD, Schopf JW. Characterization of permineralized kerogen from an Eocene fossil fern. Org Geochem 2009;40(3): 353–64. https://doi.org/10.1016/j.orggeochem.2008.12.002.

[16] Wang Q, Hou Y, Wu W, Yu Z, Ren S, Liu Q, et al. A study on the structure of Yilan oil shale kerogen based on its alkali-oxygen oxidation yields of benzene carboxylic acids, 13C NMR and XPS. Fuel Process Technol 2017;166:30–40. https://doi.org/10.1016/j.fuproc.2017.05.024.

[17] Palmer AR, Maciel GE. Relaxation Behavior in the Carbon-13 Nuclear Magnetic Resonance Spectrometric Analysis of Kerogen with Cross Polarization and Magic-Angle Spinning. Anal Chem 1982;54(13):2194–8. https://doi.org/10.1021/ac00250a014.

[18] Arfaoui A, Montacer M, Kamoun F, Rigane A. Comparative study between Rock-Eval pyrolysis and biomarkers parameters: A case study of Ypresian source rocks in central-northern Tunisia. Mar Pet Geol 2007;24(10):566–78. https://doi.org/10.1016/j.marpetgeo.2007.05.002.

[19] Shen JC, Huang WL. Biomarker distributions as maturity indicators in coals, coaly shales, and shales from Taiwan. Terr Atmos Ocean Sci 2007;18:739–55. https://doi.org/10.3319/TAO.2007.18.4.739(TT).

[20] Sweeney J, Talukdar S, Burnham A, Vallejos C. Pyrolysis kinetics applied to prediction of oil generation in the Maracaibo Basin. Venezuela. Org Geochem 1990;16(1-3):189–96. https://doi.org/10.1016/0146-6380(90)90039-3.

[21] Liu X, Zhan J-H, Lai D, Liu X, Zhang Z, Xu G. Initial Pyrolysis Mechanism of Oil Shale Kerogen with Reactive Molecular Dynamics Simulation. Energy Fuels 2015; 29(5):2987–97. https://doi.org/10.1021/acs.energyfuels.5b00084.

[22] Sweeney JJ, Burnham AK. Evaluation of a simple-model of vitrinite reflectance based on chemical-kinetics. Aapg Bull Assoc Pet Geol 1990;74:1559–70.

[23] Nielsen SB, Clausen OR, McGregor E. basin%Ro: A vitrinite reflectance model derived from basin and laboratory data. Basin Res 2017;29:515–36. https://doi.org/10.1111/bre.2017.29.issue-S110.1111/bre.12160.

[24] Burnham AK. Kinetic models of vitrinite, kerogen, and bitumen reflectance. Org Geochem 2019;131:50–9. https://doi.org/10.1016/j.orggeochem.2019.03.007.

[25] Burnham AK, Sweeney JJ. A chemical kinetic-model of vitrinite maturation and reflectance. Geochim Cosmochim Acta 1989;53(10):2649–57. https://doi.org/10.1016/0016-7037(89)90136-1.

[26] Carr AD. A vitrinite reflectance kinetic model incorporating overpressure retardation. Mar Pet Geol 1999;16(4):355–77. https://doi.org/10.1016/S0264-8172(98)00075-0.

[27] Suzuki N, Matsubayashi H, Waples DW. A simple kinetic-model of vitrinite reflectance. Aapg Bull Assoc Pet Geol 1993;77:1502–8.

[28] McCartney JT, Ergun S. Optical properties of graphite and coal. Fuel 1958;37: 272–82.

[29] Wang X, Zhao Y-P. The time-temperature-maturity relationship: A chemical kinetic model of kerogen evolution based on a developed molecule-maturity index. Fuel 2020;278:118264. https://doi.org/10.1016/j.fuel.2020.118264.

[30] Xu H, Yu H, Fan J, Xia J, Wang F, Wu H. Enhanced Gas Recovery in Kerogen Pyrolytic Pore Network: Molecular Simulations and Theoretical Analysis. Energy Fuels 2021;35(3):2253–67. https://doi.org/10.1021/acs.energyfuels.0c04137.

[31] Foster JP, Weinhold F. Natural Hybrid Orbitals. J Am Chem Soc 1980;102(24): 7211–8. https://doi.org/10.1021/ja00544a007.

[32] Pauling L. The nature of the chemical bond. Application of results obatined from the quantum mechanics and from a theory of paramagnetic susceptibility to the structure of molecules. JAmChemSoc 1931;53.

[33] Kirtman B, Chipman DM, Palke WE. Orbital Hybridization 1975;1305–7.

[34] Muller N, Pritchard DE. C13 Splittings in Proton Magnetic Resonance Spectra. I. Hydrocarbons. J Chem Phys 1959;31(3):768–71. https://doi.org/10.1063/1.1730460.

[35] Alabugin IV, Bresch S, Dos Passos GG. Orbital hybridization: A key electronic factor in control of structure and reactivity. J Phys Org Chem 2015;28:147–62. https://doi.org/10.1002/poc.3382.

[36] You Y, Wang X, Han X, Jiang X. Kerogen pyrolysis model based on its chemical structure for predicting product evolution. Fuel 2019;246:149–59. https://doi.org/10.1016/j.fuel.2019.02.075.

[37] Castro-Marcano F, Russo MF, van Duin ACT, Mathews JP. Pyrolysis of a large-scale molecular model for Illinois no. 6 coal using the ReaxFF reactive force field. J Anal Appl Pyrolysis 2014;109:79–89. https://doi.org/10.1016/j.jaap.2014.07.011.

[38] Gutowsky HS, McCall DW, Slichter CP. Nuclear Magnetic Resonance Multiplets in Liquids. J Chem Phys 1953;21(2):279–92. https://doi.org/10.1063/1.1698874.

[39] Kouvetakis J, Todd M, Wilkens B, Bandari A, Cave N. Novel Synthetic Routes to Carbon-Nitrogen Thin Films. Chem Mater 1994;6(6):811–4. https://doi.org/10.1021/cm00042a018.

[40] Lauterbur PC. C13 Nuclear Magnetic Resonance Spectra. J Chem Phys 1957;26(1): 217–8. https://doi.org/10.1063/1.1743253.

[41] Hu JZ, Solum MS, Taylor CMV, Pugmire RJ, Grant DM. Structural determination in carbonaceous solids using advanced solid state NMR techniques. Energy Fuels 2001;15(1):14–22. https://doi.org/10.1021/ef0001888.

[42] Sethi NK, Pugmire RJ, Facelli JC, Grant DM. Quantitative determination of different carbon types in fusinite and anthracite coals from carbon-13 nuclear magnetic resonance chemical shielding line-shape analysis. Anal Chem 1988;60 (15):1574–9. https://doi.org/10.1021/ac00166a020.

[43] Agrawal V, Sharma S. Molecular characterization of kerogen and its implications for determining hydrocarbon potential, organic matter sources and thermal maturity in Marcellus Shale. Fuel 2018;228:429–37. https://doi.org/10.1016/j.fuel.2018.04.053.

[44] Smernik RJ, Schwark L, Schmidt MWI. Assessing the quantitative reliability of solid-state 13C NMR spectra of kerogens across a gradient of thermal maturity. Solid State Nucl Magn Reson 2006;29(4):312–21. https://doi.org/10.1016/j.ssnmr.2005.10.014.

[45] Dennis LW, Maciel GE, Hatcher PG, Simoneit BRT. 13C Nuclear magnetic resonance studies of kerogen from Cretaceous black shales thermally altered by basaltic intrusions and laboratory simulations. Geochim Cosmochim Acta 1982;46 (6):901–7. https://doi.org/10.1016/0016-7037(82)90046-1.

[46] Witte EG, Schenk HJ, Müller PJ, Schwochau K. Structural modifications of kerogen during natural evolution as derived from 13C CP/MAS NMR, IR spectroscopy and Rock-Eval pyrolysis of Toarcian shales. Org Geochem 1988;13(4-6):1039–44. https://doi.org/10.1016/0146-6380(88)90286-0.

[47] Vučelić D, Juranić N, Vitorović D. Potential of proton-enhanced 13C n.m.r. for the classification of kerogens. Fuel 1979;58(10):759–64. https://doi.org/10.1016/0016-2361(79)90077-2.

[48] Burdelnaya N, Bushnev D, Mokeev M, Dobrodumov A. Experimental study of kerogen maturation by solid-state 13C NMR spectroscopy. Fuel 2014;118:308–15. https://doi.org/10.1016/j.fuel.2013.11.003.

[49] Dawson PE, Kent SBH. Synthesis of native proteins by chemical ligation. Annu Rev Biochem 2000;69(1):923–60. https://doi.org/10.1146/biochem.2000.69.issue-110.1146/annurev.biochem.69.1.923.

[50] Hedges JI, Oades JM. Comparative organic geochemistries of soils and marine sediments. Org Geochem 1997;27(7-8):319–61. https://doi.org/10.1016/S0146-6380(97)00056-9.

[51] Konstantinou IK, Albanis TA. Photocatalytic transformation of pesticides in aqueous titanium dioxide suspensions using artificial and solar light: Intermediates and degradation pathways. Appl Catal B Environ 2003;42:319–35. https://doi.org/10.1016/S0926-3373(02)00266-7.

[52] Lodewyk MW, Siebert MR, Tantillo DJ. Computational prediction of 1H and 13C chemical shifts: A useful tool for natural product, mechanistic, and synthetic organic chemistry. Chem Rev 2012;112(3):1839–62. https://doi.org/10.1021/cr200106v.

[53] Varani G, Aboul-ela F, Allain F-T. NMR investigation of RNA structure. Prog Nucl Magn Reson Spectrosc 1996;29(1-2):51–127. https://doi.org/10.1016/0079-6565(96)01028-X.

[54] Schutt KT, Gastegger M, Tkatchenko A, Muller KR, Maurer RJ. Unifying machine learning and quantum chemistry with a deep neural network for molecular wavefunctions. Nat Commun 2019;10:5024. https://doi.org/10.1038/s41467-019-12875-2.

[55] LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521(7553):436–44. https://doi.org/10.1038/nature14539.

[56] Kvålseth TO. Cautionary note about R2. Am Stat 1985;39(4):279–85. https://doi.org/10.1080/00031305.1985.10479448.

[57] Abadi M, Barham P, Chen JM, Chen ZF, Davis A, Dean J, et al. TensorFlow: A system for large-scale machine learning. 2016.

[58] Kelemen SR, Afeworki M, Gorbaty ML, Sansone M, Kwiatek PJ, Walters CC, et al. Direct characterization of kerogen by X-ray and solid-state 13C nuclear magnetic resonance methods. Energy Fuels 2007;21(3):1548–61. https://doi.org/10.1021/ef060321h.

[59] Longbottom TL, Hockaday WC, Boling KS, Li G, Letourmy Y, Dong H, et al. Organic structural properties of kerogen as predictors of source rock type and hydrocarbon potential. Fuel 2016;184:792–8. https://doi.org/10.1016/j.fuel.2016.07.066.

[60] Burdel'naya NS, Bushev DA. Fragment of the chemical structure of type II and II-S kerogen in the Upper Jurassic and Upper Devonian formations of the East European Platform. Geochemistry Int 2010;48(5):492–504. https://doi.org/10.1134/S0016702910050058.

[61] Leif RN, Simoneit BRT. The role of alkenes produced during hydrous pyrolysis of a shale. Org Geochem 2000;31(11):1189–208. https://doi.org/10.1016/S0146-6380(00)00113-3.

[62] Zhang T, Wang Y, Wang Z, Wei Z, Qian Yu, Sun Z, et al. Characterization of n-alkanes and n-alkylbenzenes from different sediments by Py-GC/MS. Pet Sci Technol 2017;35(17):1784–90. https://doi.org/10.1080/10916466.2017.1391843.

[63] Douglas AG, Sinninghe Damsté JS, Fowler MG, Eglinton TI, de Leeuw JW. Flash pyrolysis of Ordovician kerogens: unique distributions of hydrocarbons and sulphur compounds released from the fossil alga Gloeocapsamorpha prisca. Geochim Cosmochim Acta 1991;55:275–91.

[64] Almendros G, Zancada MC, González-Vila FJ, Lesiak MA, Álvarez-Ramis C. Molecular features of fossil organic matter in remains of the Lower Cretaceous fern Weichselia reticulata from Przenosza basement (Poland). Org Geochem 2005;36 (7):1108–15. https://doi.org/10.1016/j.orggeochem.2005.03.008.

[65] Wang X, Zhu Y, Liu Y, Li W. Molecular Structure of Kerogen in the Longmaxi Shale: Insights from Solid State NMR, FT-IR, XRD and HRTEM. Acta Geol Sin 2019;93: 1015–24. https://doi.org/10.1111/acgs.v93.410.1111/1755-6724.13870.

[66] Kim S, Thiessen PA, Bolton EE, Chen J, Fu G, Gindulyte A, et al. PubChem substance and compound databases. Nucleic Acids Res 2016;44(D1):D1202–13. https://doi.org/10.1093/nar/gkv951.

[67] Willcott MR. MestRe Nova. J Am Chem Soc 2009;131(36):13180. https://doi.org/10.1021/ja906709t.

[68] RDKit: Open-Source Cheminformatics Software. https://www.rdkit.org/; 2021 [accessed 1 March 2021].

[69] Behar F, Vandenbroucke M, Teermann SC, Hatcher PG, Leblond C, Lerat O. Experimental simulation of gas generation from coals and a marine kerogen. Chem Geol 1995;126(3-4):247–60. https://doi.org/10.1016/0009-2541(95)00121-2.

[70] Mao J, Fang X, Lan Y, Schimmelmann A, Mastalerz M, Xu L, et al. Chemical and nanometer-scale structure of kerogen and its change during thermal maturation investigated by advanced solid-state 13C NMR spectroscopy. Geochim Cosmochim Acta 2010;74(7):2110–27. https://doi.org/10.1016/j.gca.2009.12.029.

[71] Duan D, Zhang D, Ma X, Yang Y, Ran Y, Mao J. Chemical and structural characterization of thermally simulated kerogen and its relationship with microporosity. Mar Pet Geol 2018;89:4–13. https://doi.org/10.1016/j.marpetgeo.2016.12.016.

[72] Du J, Geng A, Liao Z, Cheng B. Potential Raman parameters to assess the thermal evolution of kerogens from different pyrolysis experiments. J Anal Appl Pyrolysis 2014;107:242–9. https://doi.org/10.1016/j.jaap.2014.03.007.

[73] Kingma DP, Adam BJL. A method for stochastic optimization. 3rd Int Conf Learn Represent ICLR 2015 -. Conf Track Proc 2015:1–15.

[74] Zou F, Shen L, Jie Z, Zhang W, Liu W. A sufficient condition for convergences of adam and rmsprop. Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit 2019:11119–27. https://doi.org/10.1109/CVPR.2019.01138.