Research Article

Source Rock Evaluation of Hydrocarbons in Deep-Water Offshore Areas Based on a BP Neural Network

Jizhong Wu,1 Ying Shi,2 Qianqian Yang,1 and Yanan Wang3

1School of Bohai Rim Energy, Northeast Petroleum University, Daqing, China
2School of Earth Sciences, Northeast Petroleum University, Daqing, China
3School of Energy, China University of Geosciences, Beijing, China

Correspondence should be addressed to Jizhong Wu; wjzkgr@163.com

Received 17 November 2022; Revised 18 March 2023; Accepted 20 March 2023; Published 9 September 2023

Academic Editor: Chao Zhang

Due to the lack of drilling data and poor quality of seismic data in deep-water offshore areas, conventional methods cannot effectively predict the total organic carbon (TOC) content. In this paper, the BP neural network method is used to predict the TOC of the strata overlying the target layer, which adds to the TOC information in the study area. Then, the highest TOC value of the strata overlying the target layer is used to select the most sensitive seismic attributes. Finally, the sensitive seismic attributes are used to evaluate the source rocks with no or few wells. A set of TOC prediction technology flows is established for TOC combined with seismic attributes under the condition of no wells and few wells in deep-water areas. The application example shows the reliability of TOC prediction by this technical process, and the study has a certain reference significance for the evaluation of hydrocarbon source rocks in offshore deep water.

1. Introduction

Oil and gas exploration and development in offshore deep-water areas are characterized by high investment and high risk. As the material basis of oil and gas systems and accumulation, the accurate prediction of the distribution and quality of hydrocarbon source rocks is of great significance for searching for oil and gas accumulation zones, clarifying the pattern of oil and gas accumulation, and calculating the resource amount [1–3].

The total organic carbon (TOC) content is one of the most important parameters for evaluating the hydrocarbon generation potential of source rocks. At present, the methods to obtain the TOC content mainly include geochemical, logging, and seismic methods. Due to the lack of cores, the TOC acquisition by geochemical analysis is costly, which makes it difficult to carry out systematic studies and meet the exploration needs of low-medium offshore oil-bearing basins. Considering the availability and continuity of conventional logging data, many researchers have performed exploratory work in predicting the TOC content using logging information, forming two types of methods. The first category is the ΔlogR method and its modification, that is, the superposition of resistivity and porosity curves to predict the total organic carbon content [4–10]. The problems of this method are mainly in two-fold: first, the maturity parameters need to be determined, which restricts the use of this method, so its improved algorithm is used more often now; second, the process of determining the logR model is susceptible to human interference. This model requires the artificial determination of fine-grained non-source rock as the baseline, so the appropriate selection of the baseline determines the prediction accuracy of ΔlogR. In practice, it is difficult to find a suitable baseline for some wells, even when the scale is well defined, so calculation accuracy is not guaranteed. The second type of method is to use single characteristic curves or multicharacteristic curves for fitting calculations. For example, the statistical relationship between the natural gamma curve and TOC value is established, the method of compensating for the density curve and TOC single factor fitting is proposed based on the concept of small kerogen density, and
multivariate fitting modelling is carried out using the compensation density curve, compensation neutron curve, acoustic time difference curve, deep resistivity curve, etc. [11–14]. The advantage of this kind of algorithm is that the method is simple, but the prediction accuracy needs to be further improved. These two methods require a certain amount of logging data in the target zone and require a large amount of sample data. Therefore, they are not suitable for areas with few or no wells. TOC prediction by the seismic method is mainly based on measured seismic data combined with sedimentary geological information and establishes the relationship between seismic data and the TOC of source rocks by means of geophysical inversion and multiattribute prediction [14–16]. Loseth et al. [17] used the attribute of the maximum trough amplitude to predict the thickness of the source rock. Cao et al. [18] used the seismic attribute method to predict the thickness of the source rock in the new exploration area of the northern basin of the South Yellow Sea. Lin et al. [15] found the best attribute combination by optimizing the seismic multiattribute in the Weixinan Sag. Leedberg [19] used seismic inversion to predict the distribution of shale oil source rocks in the Northern Alaska basin. Qin et al. [20] used 3D seismic data to predict the abundance of organic matter in the entire formation through seismic multiattribute inversion. Tao et al. [21] used frequency division inversion technology to predict source rocks, and Liu et al. [22] used a multiattribute fusion method to predict source rocks. This method requires the measured TOC of the target layer and then obtains TOC information of the whole target layer by using the continuity of seismic data. However, it cannot be applied to the situation where no well is present in the target layer.

In recent years, artificial neural networks have developed rapidly. As a mathematical model imitating biological neural networks, it has many advantages in solving high-dimensional and nonlinear problems. Many scholars try to use neural network algorithms to predict TOC, that is, to improve the accuracy of TOC calculation by using the strong approximation function of neural networks, among which the BP neural network is the most widely used network. Guo et al. calculated the TOC content by combining plate classification with a BP neural network [23]. Xiong et al. used a BP neural network to predict the TOC of shale [24]. Zhang et al. used a BP neural network to predict and evaluate the TOC content in complex lithologies [25]. Bolandi et al. predicted the TOC content of mudstone by combining a BP neural network with acoustic, resistivity, and density logging [26]. Such methods further improve the calculation accuracy of TOC and are widely used. Compared with deep learning methods represented by convolutional neural network (CNN), BP neural network is a shallow neural network, which has many advantages such as simple and reliable structure, simple sample data production, and fast error convergence. It is very suitable for nonlinear fitting problems such as parameter prediction.

The study area is located in the northern deep-water area of the Yinggehai Basin, covering an area of 6000 km². The area is composed of the Lingao uplift, Haikou nose-like structural belt, and east–west slope belt, which is a Cenozoic sedimentary depression with a thickness of more than 8 km. The structures were deposited mainly during the Eocene (Lingtou Formation), Oligocene (Yacheng Formation and Lingshui Formation), Miocene, Pliocene, and Quaternary. The source rocks are mainly the Oligocene Yacheng Formation, which is composed of littoral and neritic facies and is also the target layer of this study. In the study area, 24 exploration wells have been drilled, among which 7 wells have measured 137 sets of discrete TOC contents. Due to the limitation of drilling depth, none of the wells reached the target layer, and the TOC of only 1 well was taken from the Lingshui Formation, which overlies the target layer. The Lingshui Formation and the target formation are Oligocene in age, while the TOC values of the other wells are from shallower formations of Miocene age and younger. TOC was not measured in the remaining 17 wells, of which only one well was drilled into the Lingshui Formation, the formation overlying the target formation, while the other wells were drilled into shallower formations. The study area is located in the deep-water area, which is not conducive to the prediction of source rocks due to the extremely harsh conditions, such as less TOC measured in drilling, no logging information in the target layer, and poor seismic data quality. The single logging method, seismic method, or neural network method cannot overcome the constraints of the above conditions and obtain satisfactory source rock prediction results. In this paper, based on the nonlinear characteristics between TOC and logging information, the BP neural network method is used to predict the TOC of the strata overlying the target layer and add to the TOC information in the study area. Then, the highest TOC value of the strata overlying the target layer is used to select the most sensitive seismic attributes. Finally, the selected seismic attributes are used to evaluate the source rock without wells in the deep-water area.

2. Logging Characteristics of the Total Organic Carbon Content

Conventional logging series have the characteristic of low cost and are more widely used than unconventional logging series, so this paper discusses only the relationship between conventional logging curves and TOC. The correlation analysis of 137 cores from 7 wells in the study area shows that TOC has a certain correlation with gamma logging, resistivity logging, and acoustic time difference logging (Figure 1). Generally, lacustrine or marine argillaceous source rocks tend to absorb more radioactive uranium due to their small grain size and large specific surface area, resulting in higher natural gamma logging values. However, the TOC of argillaceous source rocks in this area is negatively correlated with natural gamma radiation (Figure 1(a)), which may be because the argillaceous source rocks in this area are coal-measure source rocks. As organic carbon is enriched, the contents of humus and absorbed radioactive elements decrease, leading to a decrease in natural gamma rays. The resistivity of organic matter is large, and the resistivity log value increases with increasing TOC. The acoustic log value increases with increasing TOC. Due to the high acoustic propagation time of organic matter, the acoustic log value increases with increasing TOC.
However, the correlation between the acoustic time difference log value and TOC content is not obvious in this area (Figure 1(c)), which may be related to the fact that this area contains coal-measure source rock with a large burial depth.

As shown in Figure 1, although there is a certain degree of correlation between the response of the above characteristic curves and TOC content, such a relationship can be used only for qualitative analysis of TOC, which has too low of precision for quantitative calculation, especially when the TOC content is low, thereby showing obvious dispersion. This phenomenon occurs because conventional logging curves measure the comprehensive response of formations. When the TOC content is not high, the information expressed by logging curves is suppressed by other information, which is difficult to show. Therefore, it is necessary to use a neural network to mine TOC information.

3. Predicting TOC Using a BP Neural Network

3.1. The Architecture of the BP Neural Network. A back propagation (BP) neural network is a multilayer feed-forward network trained by an error inverse propagation algorithm that can easily deal with complex multidimensional nonlinear mapping problems in engineering research and has a good effect on data classification, clustering, and prediction. Its main principle is to use the method of error back propagation to train on the basis of the known learning sample set and build a network with the training results [23, 27–31]. The learning process of the BP neural network is divided into two kinds: forward propagation learning and back propagation learning. In the forward learning process, the input vector is processed layer by layer from the input layer through the hidden layer and then transmitted to the output layer. In this case, the state of neurons in each layer affects only the state of neurons in the next layer. Once the desired results cannot be obtained in the output layer, reverse propagation is carried out again, and the error signal is returned along the original path. In this way, the error is minimized by constantly modifying the weight of neurons in each layer (Figure 2). The input data of the input layer of the BP neural network are normalized logging data, such as RT data (resistivity logging), which can be expressed as follows:

\[ x_i = \frac{RT - RT_{\text{min}}}{RT_{\text{max}} - RT_{\text{min}}} \]  

where \( x_i \) is the normalized sample data obtained from the input layer, \( RT_{\text{min}} \) is the minimum resistivity logging data, and \( RT_{\text{max}} \) is the maximum resistivity logging data. In order to ensure the normal calculation of equation (1), abnormal cases of \( RT_{\text{min}} \) and \( RT_{\text{max}} \) should be excluded, and the difference between them should not be 0.

The most basic algorithm of the BP neural network is “the fastest descent method,” which is applied to two processes of signal forward propagation and error back propagation. The output formula of the output layer in the forward propagation process is as follows:

\[ Y_k = f \left( \sum_{j=1}^{n} (w_{jk} \times x_j) + b_k \right) \]

where \( f = 1 / (1 + e^{-x}) \) is the activation function, \( w_{jk} \) is the weight from the \( j \) neuron in the hidden layer to the \( k \) neuron in the output layer, \( b_k \) is the offset of the \( k \) neuron in the output layer.
hidden layer, \( x_j \) is the output of the last hidden layer, and \( Y_k \) is the output result of the output layer.

In the process of error back propagation, the output error signals of neurons in each layer are calculated from the output layer, and then the weights and thresholds of each layer are adjusted according to the error gradient descent method to minimize the mean square error (MSE). The expression of the mean square error function is as follows:

\[
MSE = \frac{1}{2n} \sum_{i=1}^{n} (E_i^2),
\]

where \( E_i \) is the error of the \( i \)th input data and \( n \) is the number of input data point. In Formula (2), the weight adjustment formula is expressed as follows:

\[
w_{ij}(t+1) = e \times \beta_j \times Y_k + \alpha \times \Delta w_{ij},
\]

where \( w_{ij}(t+1) \) is the weight from input layer \( i \) to hidden layer \( j+1 \), \( \Delta w_{ij} \) is the weight from input layer \( i \) to hidden layer \( j \), \( e \) is the learning step, \( \beta_j \) is the error term, \( \alpha \) is the momentum factor, and \( Y_k \) is the result of the output layer. The parameter \( \alpha \) is used to prevent local error minimization by using additional momentum to slide over these minima. The correction weight should not only consider the effect of error on the gradient but also consider the influence of the error trend.

3.2. Training and Validation. During BP neural network modelling, the logging data collected in the study area, including gamma, resistivity, and acoustic time difference data, were first preprocessed and standardized uniformly. The measured TOC values of 116 samples from 6 wells were collected, and the corresponding three groups of logging data were input into the BP neural network. The actual measured TOC of the remaining 1 well was used as validation data. The input data are a \( 3 \times 116 \) matrix of 116 three-element vectors, and the output data are a \( 1 \times 116 \) matrix of 116 one-element vectors. Sufficient sample data can guarantee the generalization ability of the network. However, limited by the quantity and quality of actual logging data, we can only obtain 116 sets of sample data to participate in the training and verification of the network. Although the subsequent test and verification results of the network are satisfactory, it is always required to obtain as many samples as possible. The network debugging process is as follows: 100 groups of data are randomly selected as training sets, and the weights of the neural network are constantly updated to minimize the MSE. The training set was used to determine the network structure, and the remaining 16 groups of data were used as test sets to independently test the performance of the neural network. The expected value was set, that is, the overall accuracy was greater than 90%. The final training time was 10 min, and the final epoch number was 297. As shown in Figure 3, when the number of iterations is 297, the training accuracy is the best. Here, the output of each layer needs to be deactivated by activation functions; commonly used activation functions include sigmoid, tanh, and ReLU functions. Among them, sigmoid and tanh functions are prone to the problem of gradient disappearance, and the deep network does not easily converge during training. Therefore, the ReLU function is adopted here, which can adequately solve the above problems and carry out training smoothly. In network training, the trainlm function is used as the training function, and the momentum factor is 0.93. Figure 4 shows the distribution of the predicted TOC and measured TOC values in the training and test phases. The correlation coefficient \( (R) \) in the training phase is 0.83, and that in the test phase is 0.81. Figure 5 shows TOC values predicted from the validation well compared to measured values. As shown in Figure 5, the two values are in good agreement. Although there is some deviation in the values of individual well sections, the trend of increasing or decreasing values on the two curves is basically the same. The results show that the BP neural network has high reliability in predicting the TOC of source rocks and can provide a basis for resource potential evaluation.

4. The Use of High TOC to Select the Most Sensitive Seismic Attribute

There are no drilling data of the target formation in the study area, and only two wells (C and D) are drilled into the Lingshui Formation adjacent to the target formation. Well C has a small number of measured discrete TOC values in the Lingshui Formation, while well D has no measured TOC values. Standardized acoustic time difference, resistivity, and gamma log data from these two wells were used as input data for continuous TOC prediction using trained neural networks (Figure 6). For the whole study area, although the TOC values of the two wells adjacent to the target layer are known, the TOC distribution of the target layer cannot be directly obtained vertically. In addition, due to the small number of wells in the horizontal direction and the large area of the study area, the TOC distribution of the source rock cannot be predicted horizontally. Seismic data are
information bodies that can be continuously distributed both vertically and horizontally. Given the seismic information at one location, we can infer the seismic information at another location at a certain distance according to the continuous distribution of seismic data. The change in geological conditions indicated by seismic information is a gradual process, which is consistent with the characteristics of geological deposition [16]. Based on the continuity and gradual characteristics of seismic data, we propose an evaluation route for deep source rocks in the case that there is no drilling in the target layer in deep water. First, the high TOC values of two wells in the Lingshui Formation of the strata overlying the target layer were correlated with seismic attributes, and the seismic attribute with the highest correlation was found. Then, the high TOC area of the deep target layer is predicted by using the same seismic attribute based on the continuity of seismic data. Here, the key to successful TOC prediction using attributes is to extract and select seismic attributes that are most relevant to source rocks. In this study, GeoEast software was used to extract 39 seismic attribute bodies of the instantaneous class, wavelet class, statistical class, and mathematical class, further calculate the correlation between the attribute values of each attribute body at these two wells and the predicted TOC, and screen out 8 kinds of attribute bodies with a correlation of more than 30% (Table 1). The perigram (smooth reflection intensity) has the greatest correlation with TOC. It should be noted that, when calculating the correlation between seismic attributes and TOC values, it is necessary to extract the attribute values and TOC values of the same depth of the well in advance and then use the common correlation calculation method in mathematics to calculate the correlation. Since TOC values are discrete vertically and seismic attributes are continuous, both values must be guaranteed to exist at the same depth when extracting these two values.

Figure 7 shows the main survey line profile of the perigram passing through well C. Figure 8 shows the local enlargement of Figure 7. The above figure shows the perigram attribute, and the following figure shows the seismic data profile corresponding to the above figure. Figure 8 shows that the high TOC area of well C is in good agreement
with the perigram of the Lingshui Formation. In addition, the seismic data corresponding to high TOC are characterized by strong amplitude, medium and low frequencies, and good continuity.

Figure 9 shows the connection line profile of the perigram across well C. Figure 10 is a partial enlargement of Figure 9, with the perigram attribute shown above and the corresponding seismic data profile shown below. The features observed in Figure 10 are the same as those in Figure 8.

Figure 11 shows the mainline section of the perigram across well D. Figure 12 is a partial enlargement of Figure 11, with the perigram shown above and the seismic

### Table 1: Correlation between seismic attributes and TOC.

<table>
<thead>
<tr>
<th>Seismic attribute</th>
<th>Relevance to TOC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perigram (smooth reflection intensity)</td>
<td>0.78</td>
</tr>
<tr>
<td>RST (the intensity of reflection)</td>
<td>0.37</td>
</tr>
<tr>
<td>Prg*CIP (the product of reflection intensity and phase cosine)</td>
<td>0.36</td>
</tr>
<tr>
<td>Entropy (pseudo entropy)</td>
<td>0.42</td>
</tr>
<tr>
<td>Normal var (the standard variance)</td>
<td>0.41</td>
</tr>
<tr>
<td>Kurt (kurtosis)</td>
<td>0.40</td>
</tr>
<tr>
<td>Variance</td>
<td>0.33</td>
</tr>
<tr>
<td>Absamp (absolute value of amplitude)</td>
<td>0.43</td>
</tr>
</tbody>
</table>
data section corresponding to the figure shown below. Figure 12 shows that the high TOC area of well D is in good agreement with the perigram of the Lingshui Formation. In addition, the seismic data corresponding to high TOC also have the characteristics of strong amplitude, medium and low frequencies, and good continuity.

Figure 13 shows the perigram connection line profile across well D. Figure 14 is a partial enlargement of Figure 13, with the perigram attributes shown above and the corresponding seismic data profile shown below. The features observed in Figure 14 are the same as those in Figure 12. In addition, the high TOC area of well D matches the perimeter of the Lingshui Formation very well. In addition, the seismic data corresponding to high TOC also have the characteristics of strong amplitude, medium and low frequencies, and good continuity.
Laterally, the distance between the two wells, C and D, is 5 km, and the high TOC area of the two wells in different positions coincides with the perigram attribute of the same formation. This indicates two points of view: one viewpoint is that the accuracy of TOC prediction by the BP neural network is very high and reliable; the other is that it is reasonable and feasible to predict the spatial distribution of TOC by using seismic attributes.

5. Technical Process for TOC Prediction

The technical process for TOC prediction based on the BP neural network and seismic attributes is presented below. This technical process can be applied to the evaluation of source rocks in areas with no or few wells in the target layer, as well as other areas with general conditions.
(1) The logging curves are collected and preprocessed, and this step includes acoustic logging, gamma logging, and resistivity logging; collecting measured TOC values; and generating training data sets.

(2) A BP neural network is built, a training data set is used to train the neural network, and then the effectiveness and accuracy of the network are verified.

(3) A neural network is used to predict continuous TOC values of unmeasured TOC wells.

(4) A variety of seismic attributes is obtained, and then the most sensitive attribute associated with high TOC is selected.

(5) This sensitive attribute is used to predict the high TOC distribution area of the target layer.

6. TOC Prediction Results

Based on the aforementioned TOC prediction and evaluation process of source rocks, the regional lateral
Figure 14: Local enlargement of Figure 13.

Figure 15: Distribution characteristics of high TOC in the target layer.
characteristic map of the high TOC content of the target formation is obtained (Figure 15). High TOC areas (yellow) are distributed laterally in strips and sheets. Since well CD was not drilled into the target formation, we projected the wellhead of vertical well CD onto the Yacheng Formation for easy display, as shown in Figure 15. TOC is high in the area where well CD is located, and it can be inferred that the source rocks there are relatively developed.

In general, the distribution characteristics of TOC predicted by the BP method can provide a basis for source rock evaluation, but there are some limitations of this method. These limitations are present because there are no wells in the target layer in the study area, and very few wells are drilled into the strata adjacent to the target layer. Therefore, limited by the small number of samples, it is impossible to establish a one-to-one mapping relationship between the predicted TOC values and the values of sensitive attributes, which has adverse effects on the numerical characterization of TOC in the target layer. Therefore, this method can only describe the TOC of the study area qualitatively rather than quantitatively.

7. Conclusion

The application in the study area shows that the TOC prediction technique combined with the BP neural network and seismic attributes has good prospects in the study of source rocks. Through the research in this paper, the following conclusions are drawn:

(1) The BP neural network has the characteristics of a simple structure, high prediction accuracy, and strong adaptability. The predicted TOC has high reliability, which provides a basis for resource quantity evaluation and a technical idea for the TOC evaluation of source rocks under such geological conditions

(2) The technical process is still unable to achieve the spatial quantitative characterization of TOC, and further targeted research should be carried out

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this article.

Acknowledgments

The authors are grateful to the Key Project of Natural Science Foundation of China (41930431), the Natural Science Foundation of Heilongjiang Province (LH2023D010), and the Northeast Petroleum University’s special funds (2021YDQ-01) for supporting this work.

References


