A dynamic adaptive radial basis function approach for total organic carbon content prediction in organic shale

Maojin Tan¹, Qiong Liu², and Songyang Zhang³

ABSTRACT

Total organic carbon (TOC) is an important parameter for characterizing shale gas and oil reservoirs. Estimation of TOC from well logs has previously been achieved by an empirical model. The radial basis function (RBF) neural network is a new quantitative method that can generate a smooth and continuous function of several input variables to approximate the unknown forward model. We investigated the basic principles of the RBF including network structure, basis function, network training method, and its application in the TOC prediction. The nearest neighbor algorithm was selected for the network training. Then, the Gaussian width was investigated to improve the TOC prediction accuracy through leave-one-out cross-validation. Finally, field cases of organic shale were studied for the TOC prediction, and the prediction results using the RBF method were compared with those of the $\Delta \log R$ method. Furthermore, according to sensitive attribute ranking, the impacts of different input logs on the predicted results were also investigated through various experiments, and the best network model was finally chosen. The error analysis between the prediction results and lab-measured TOC in some examples indicated that the new approach is more accurate than a single empirical regression method and more flexible than the $\Delta \log R$ method.

INTRODUCTION

As important unconventional oil and gas resources, shale oil and shale gas reservoirs are hot-spot fields of petroleum exploration and production at the present. Total organic carbon (TOC) is one of the most important parameters in the evaluation of reservoir quality and hydrocarbon potential (Sachsenhofer et al., 2010; El Sharawy et al., 2012). Furthermore, TOC content is an indispensable parameter of geophysical characterization of shale-gas plays, which can significantly influence the geophysical response of organic-rich rocks (Vermik and Milovac, 2011; Zhu et al., 2011, 2012; Bandyopadhyay et al., 2012). There are many evaluation methods of the TOC based on well logs. Fertl and Rieke (1980) and Fertl and Chillingar (1988) use the gamma-ray spectral log to identify organic-rich rocks and analyze their relationships with total gamma ray, uranium (U), and thorium-potassium (TH/K) ratio. Mendelson and Toksoz (1985) use multivariate methods to construct the quantitative relationships between well logs and the TOC determined with core samples, and the regression equation has a high coefficient of determination. Carpentier et al. (1991) propose CARBOLOG method, which combined physical properties with long sonic transit time and high resistivity to derive an in situ estimation of the organic matter content. But, it requires a calibration based on the lab-measured TOC data available (Yun et al., 2000; Liu et al., 2003). In addition, Passey et al. (1990) propose a technique called $\Delta \log R$, which employs the overlaying of porosity logs (sonic, density, and neutron) and resistivity logs for identifying and calculating the TOC contents. Xu et al. (1998) select the acoustic log baseline and resistivity log baseline of conventional mudstone and calculate the TOC content from well logs. The calculation results are consistent with the core experimental results. Kamali and Mirshady (2004) apply the $\Delta \log R$ method in combination with neuro-fuzzy approaches for determining TOC content. Moreover, it can also be used in reference log response to shale-gas reservoirs. Pan et al. (2009) apply the $\Delta \log R$ method into the TOC calculation of shale gas reservoirs. Passey et al. (2011) revise calibration of the $\Delta \log R$ method to high maturity organic-rich rocks and identify the mature hydrocarbon source rocks of shale gas reservoirs. Bakhtiar et al. (2011) apply $\Delta \log R$ and neural network methods to estimate the TOC content for the source rock evaluation. Therefore, the $\Delta \log R$ technique is the most-used reference and the most popular TOC evaluation method using well
Recently, a radial basis function (RBF) interpolation method was proposed to solve inverse problems encountered in experimental data analysis, which can approximate smooth and continuous multivariate functions of many variables and derive a nonlinear mapping function to solve well-logging and geophysical inverse problems associated with unknown forward models. Freed (2004) first puts forward the new approach and applies it to determine the characteristics and composition of a fluid sample. Heaton and Freedman (2005) use the RBF method to predict the properties of live oils from nuclear magnetic resonance (NMR) relaxation time. Freedman (2006) generally summarizes the RBF theory, and introduces some case studies, whose predicting results were in excellent agreement with the experimental results. Anand and Freedman (2009) adopt the RBF method to determine molecular properties of hydrocarbon mixtures from NMR data. Huang et al. (2011) also use the RBF network for well log inversion and compute the true formation resistivity partly to the input signal; that is, the hidden nodes will produce a larger output when the input signal is close to the central area of basis functions. The node of the output layer is usually a simple linear function.

The TOC content is an integrated function of many wireline logs. If a variety of wireline logs are chosen as the input layer and the unknown TOC content as the output layer, then the TOC prediction based on this network can be realized (Figure 1).

Basis functions

The basis functions are effective function of hidden nodes for the RBF neural network, which are generally named the RBF. Some common RBFs (Powell, 1987; Oyang et al., 2005) are

\[ f(x) = \exp \left( -\frac{x - c_i}{r_i} \right)^2, \]
\[ f(x) = \frac{1}{(r^2 + x^2)\alpha}, \quad \alpha > 0, \]
\[ f(x) = (\alpha^2 + x^2)^\beta, \quad \alpha < \beta < 1. \]  

(1)

These functions are of radial symmetry. The Gaussian function has the advantages of a simple form, radial symmetry, smoothness, and good analyticals for the theoretical analysis (Xu and Torres-Verdin, 2013). So, it is often chosen as the RBF for parameter prediction.

In vector space, the Gaussian basis function is chosen as the RBF:

\[ G_i(x) = \exp \left( -\frac{\|x - c_i\|^2}{2r_i^2} \right), \quad i = 1, 2, \ldots, m, \]  

(2)

where \( x \) is an \( n \)-dimensional input vector; \( c_i \) is the center of the \( i \)th basis function, which has the same dimension as \( x \); \( r_i \) is the \( i \)th apperceive variable, which may be chosen freely, and it determine the width of basis function around the center \( c_i \); \( m \) is the number of the apperceive units, namely the number of hidden nodes. The symbol \( \|x - c_i\| \), the norm of the \( x - c_i \) vector, usually means the distance between \( x \) and \( c_i \); \( G_i(x) \) has only a maximum value at \( c_i \), and with \( \|x - c_i\| \) increasing, \( G_i(x) \) will decrease to zero rapidly. For a given input vector \( x \in \mathbb{R}_n \), only a small part near the center of the \( x \) is activated. This reflects the response characteristics of brain
cortex, which shows the local approximation ability of the network, so the RBF network is also called the local apperceived field network.

From Figure 1, it can be seen that the mapping \( x \rightarrow G_i(x) \) from the input layer to the hidden layer is nonlinear, and the mapping from the hidden layer to the output layer \( G_i(x) \rightarrow y^k \) is a linear one, namely,

![Crossplots](image)

**Figure 2.** Crossplots of lab-measured TOC of cores versus wireline logs such as (a) CNL, (b) AC, (c) DEN, (d) GR, and (e) ILD in well A of the literature (Passey et al., 1990). The correlations of determination between lab measured TOC and wireline logs are list in the figure.
\[ y_k = \sum_{i=1}^{m} w_{i,k} G_i(x), \quad k = 1, 2, \ldots, m, \]  

where \( y_k \) is the \( k \)th output node (output variable), \( w_{i,k} \) is the output-weighting coefficient of the RBF network, and \( m \) is the number of the hidden node. Because \( G_i(x) \) is a Gaussian function and \( G_i(x) > 0 \) for any \( x \), it has the advantage of local adjusting weighting constant. In fact, when \( x \) is far away from \( c_i \), the basis function \( G_i(x) \) becomes negligibly small, so it can be treated as zero. When \( G_i(x) \) is larger than a threshold value (i.e., \( G_i(x) = 0.05 \)), the corresponding weighting coefficient \( w_{i,k} \) is not zero, which is meaningful to the RBF network. Through the treatment above, the RBF network has some advantages of local approaches network and fast learning.

**The nearest neighbor-clustering learning algorithm**

The learning algorithms of the RBF neural network mainly include random algorithm, self-organizing learning algorithm and the nearest neighbor-clustering learning algorithm. These learning algorithms are used to determine the center of the RBF network. The random algorithm and self-organizing learning algorithm are applicable for off-line learning of the static model, which must have all possible sample data beforehand, and it cannot be used in online learning of the dynamic input mode. Before learning, the number of input data centers, that is, the number of RBF network hidden nodes, must be determined, which is often difficult. The nearest neighbor learning algorithm is a dynamic adaptive RBF network model, and this algorithm does not need to determine the number of hidden units beforehand. Furthermore, it is online learning, and the RBF network after clustering may be optimal.

The process of the nearest neighbor-clustering learning algorithm is as follows:

1. An appropriate width \( r \) of the Gaussian function is chosen. A vector \( A(l) \) is defined to store the sum of the output vector belonging to various samples. A counter \( B(l) \) is defined to count the number of various types of samples, where \( l \) is the number of categories.

2. Start from the first data pair \( (x^1, y^1) \), establish a cluster center on \( x^1 \), namely, let \( c_1 = x^1 \). \( A(1) = y^1, B(1) = y^1 \). The established RBF network has only one hidden layer unit, the center of the hidden layer unit is \( c_1 \), the weight vector of the hidden layer units to the output layer is \( w_1 = A(1)/B(1) \).

3. Consider the second sample data pair \( (x^2, y^2) \), and calculate the distance \( ||x^2 - c_1|| \) from \( x^2 \) to the cluster center \( c_1 \). If \( ||x^2 - c_1|| \leq r \), \( c_1 \) is the nearest neighbor clustering of \( x^2 \), and let \( A(1) = y^1 + y^2, B(1) = 2, w_1 = A(1)/B(1) \).

4. If \( ||x^2 - c_1|| > r \), then treat \( x^2 \) as a new cluster center and let \( c_2 = x^2, A(2) = y^2, B(2) = 1 \). Add one more hidden layer unit in the RBF network above, and the weight vector of the hidden layer units to the output layer is \( w_2 = A(2)/B(2) \).

5. If we consider the \( k \)th sample data pair \( (x^k, y^k) \) \( (k = 3, 4, \ldots, N) \), there are \( M \) cluster centers, whose center points are \( c_1, c_2, \ldots, c_M \), respectively. So, there are \( M \) hidden layer unit in the RBF network above. And then, calculate the distance \( ||x^k - c_j|| (j = 1, 2, \ldots, M) \), from \( x^k \) to the \( M \)th cluster center. Let \( ||x^k - c_j|| \) be the minimum among these distances and \( c_j \) is the nearest neighbor clustering of \( x^k \). That is, if \( ||x^k - c_j|| > r \), treat \( x^k \) as a new cluster center. Let \( c_{M+1} = x^k \), \( M = M + 1, A(M) = y^k, B(M) = 1 \), and keep the value of
Figure 4. Comparison of the RBF method with ILD/DEN and ILD/CNL as inputs and comparison of the $\Delta \log R$-derived results with and core TOC measurement. (a) Crossplot of the RBF derived TOC with ILD/DEN as inputs and core TOC measurement, and the correlation coefficient is approximately 0.851. (b) Crossplot of the RBF-derived TOC with ILD/CNL as inputs and core TOC measurement, and the correlation coefficient is approximately 0.803. (c) Comparison of the $\Delta \log R_{DEN}$-derived TOC (overlap of deep induction resistivity and density log) and lab-measured TOC results with the correlation coefficient of approximately 0.721. (d) Comparison of the $\Delta \log R_{CNL}$-derived TOC (overlap of deep induction resistivity and neutron log) and lab-measured TOC results with the correlation coefficient of approximately 0.790.

Table 1. Comparison of the RBF predicted and the $\Delta \log R$-derived results of well A in Passey et al. (1990).

<table>
<thead>
<tr>
<th>No.</th>
<th>RBF method</th>
<th>Prediction accuracy of the RBF method</th>
<th>$\Delta \log R$ method</th>
<th>Calculation accuracy of $\Delta \log R$ method</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Different inputs</td>
<td>Gaussian spread</td>
<td>The number of neurons</td>
<td>MSE (wt.%), MAE (wt.%), MRE (%)</td>
</tr>
<tr>
<td>1</td>
<td>RT/AC</td>
<td>0.18</td>
<td>10</td>
<td>1.851, 1.023, 18.743</td>
</tr>
<tr>
<td>2</td>
<td>RT/DEN</td>
<td>0.05</td>
<td>20</td>
<td>1.979, 1.095, 21.571</td>
</tr>
<tr>
<td>3</td>
<td>RT/CNL</td>
<td>0.10</td>
<td>15</td>
<td>2.512, 1.297, 23.278</td>
</tr>
</tbody>
</table>

The correlation coefficient refers to the relationship between TOC-calculated results from well logs and lab-measured TOC values. The standard deviation refers to the standard deviation of the absolute error (in wt.%) between $\Delta \log R$-derived TOC results and lab-measured TOC.
Figure 5. Comparison of lab-measured TOC and RBF-predicted results with different logs as inputs (a) five-log (AC/GR/DEN/CNL/ILD), (b) four-log (AC/GR/DEN/CNL), (c) three-log RBF (AC/GR/DEN/CNL/ILD), and (d) two-log (AC/GR), and (e) one-log (AC) as inputs in Well A of (Passey et al., 1990).
A(i), B(i), (i = 1, 2, . . . , M) unchanged. Add the Mth hidden layer unit in the RBF network above, the weight vector of the hidden layer units to the output layer is \(w_M = A(M)/B(M)\).

If \(||x^i - c_j|| \leq r\), then \(A(j) = A(j) + y^i\), and \(B(j) = B(j) + 1\). When \(i \neq j\), the value of \(A(i)\) and \(B(i)\) is kept unchanged, the weight vector of the hidden layer units to the output layer is \(w_i = A(i)/B(i)\) (i = 1, 2, . . . , M).

After such a learning algorithm is established, the RBF network output is

\[
f(x^i) = \frac{\sum_{i=1}^{m} w_{ik} \exp\left(-||x - c_i||^2/r^2\right)}{\sum_{i=1}^{m} \exp\left(-||x - c_i||^2/r^2\right)}, \tag{4}
\]

where \(r\) is the Gaussian width, namely, the spread constant. Because \(r\) is a 1D parameter, the appropriate spread constant could be determined by multitime cross-validation experiments and the errors contrast. The RBF method is much more convenient than the backpropagation (BP) neural network, which needs to determine the number of the hidden layers at the same time and whose results are also not steady. In fact, because each pair of input and output data may create a new clustering, this kind of dynamic adaptive RBF network is realized by adaptive adjustments of the network structure including different inputs and parameters at the same time.

**RBF PREDICTION EXPERIMENTS AND PARAMETERS OPTIMIZATION**

**Data source**

The target reservoir of well A from the literature (Passey et al., 1990) is organic shale. The wireline logs of this well include natural gamma ray (GR), deep induction log (ILD), compensated acoustic (AC), compensated neutron log (CNL) and compensated density (DEN). The crossplots of core-measured TOC and well logs are illustrated in Figure 2. The \(\Delta \log R\) method in the literature (Passey et al., 1990) was used for the TOC prediction.

**Leave-one-out cross validation**

In this research, the leave-one-out experiment is chosen to investigate the influence on prediction accuracy of Gaussian spread or width. The leave-one-out experiment is that one from all samples is chosen for verifying the TOC prediction, and the others are regarded as training data set to do a series of tests with different spread, which is used to construct the optimum network to get the coefficient matrix. Finally, the TOC prediction result of the leaved one sample is calculated. Similarly, the TOC results of the other samples are predicted one by one.

First, the input log attributes are considered as RT and AC in the TOC prediction experiment. In the RBF experiment, different

<table>
<thead>
<tr>
<th>Wells</th>
<th>Correlation coefficient</th>
<th>Gaussian spread</th>
<th>The number of neurons</th>
<th>MAE (wt.%)</th>
<th>MRE (%)</th>
<th>Correlation coefficient</th>
<th>Standard deviation of absolute error (wt.%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.865</td>
<td>0.13</td>
<td>14</td>
<td>1.023</td>
<td>1.360</td>
<td>0.846</td>
<td>1.449</td>
</tr>
<tr>
<td>B</td>
<td>0.871</td>
<td>0.18</td>
<td>9</td>
<td>1.216</td>
<td>1.305</td>
<td>0.845</td>
<td>1.425</td>
</tr>
<tr>
<td>C</td>
<td>0.915</td>
<td>0.14</td>
<td>9</td>
<td>0.897</td>
<td>0.878</td>
<td>0.846</td>
<td>1.341</td>
</tr>
<tr>
<td>D</td>
<td>0.855</td>
<td>0.25</td>
<td>5</td>
<td>0.957</td>
<td>1.118</td>
<td>0.836</td>
<td>1.241</td>
</tr>
<tr>
<td>E</td>
<td>0.754</td>
<td>0.11</td>
<td>10</td>
<td>0.990</td>
<td>1.596</td>
<td>0.836</td>
<td>1.468</td>
</tr>
<tr>
<td>F</td>
<td>0.671</td>
<td>0.24</td>
<td>8</td>
<td>0.590</td>
<td>0.755</td>
<td>0.836</td>
<td>1.241</td>
</tr>
</tbody>
</table>
Gaussian widths are chosen to calculate the TOC, respectively. To compare the prediction precision in different Gaussian spreads, some errors such as the mean square error (MSE), mean absolute error (MAE), and mean relative error (MRE) between the predicted results and core experiments are calculated. These errors above as well as the number of neurons are illustrated in Figure 3. It can be seen from Figure 3 that the MAE and MRE first decrease and then increase gradually with the Gaussian width increasing, whereas the number of neurons decreases steadily. Therefore, the corresponding Gaussian width in place of the “inflection point” should be chosen for training and the prediction, and the prediction results of the RBF network with this spread is the most accurate, which is approximately 0.1 in Figure 3a and 3b. Figure 3c shows the contrast between predicted results using the RBF method and core measurement results, and the correlation coefficient ($R$) is approximately 0.865. Figure 3d shows the comparison of the RBF-predicted

Figure 6. Crossplots of lab measured TOC of cores versus various wireline logs such as (a) CNL, (b) GR, (c) PE, (d) DEN, (e) LLD, (f) AC, (g) U, (h) TH, (i) K, (j) TH/U ratio, (k) KTH, and (l) TH/K ratio in case study well. The correlations of determination ($R^2$) between lab-measured TOC and wireline logs are different; the one with TH/U ratio is the highest, and the one with CNL is lowest.
and the $\Delta \log R$-derived results, in which the $\Delta \log R$ method of resistivity log and acoustic log is called $\Delta \log R_{AC}$. From Figure 3c and 3d, it can be seen that the RBF result is slightly better than that of the $\Delta \log R$-derived method, whose correlation coefficient between the $\Delta \log R$-derived and core-measured TOC is approximately 0.854.

Then, ILD/DEN and ILD/CNL are also chosen for the RBF prediction experiment, respectively, and the RBF-predicted results are listed in Figure 4a and 4b. Furthermore, the $\Delta \log R_{DEN}$ method of the resistivity log and density log and the $\Delta \log R_{CNL}$ method of the resistivity log and neutron log are implemented to calculate the TOC contents, their calculated results are listed in Figure 4c and 4d.

Figure 7. Comparison of different $\Delta \log R$-derived TOC curves and lab-measured TOC of core samples. Three $\Delta \log R$ methods with acoustic log, density log, and neutron log are implemented. The results of the target shale gas layer are not well in agreement with the lab-measured TOC.
respectively. From the data point distribution in Figure 4, the results of the RBF method in Figure 4a and 4b are better than those of both \( \Delta \log R \) methods in Figure 4c and 4d. In addition, Table 1 shows the error analysis and the comparison of the experimental results of the RBF and the \( \Delta \log R \) method. By contrast, the error of the RBF prediction is lower and the correction coefficient is higher than that of the \( \Delta \log R \) method, which can be seen in the right column of Table 1. Moreover, in the Passey et al. (1990), the standard deviation of error of approximately 0.570.

Figure 8. Comparison of lab-measured TOC of cores and \( \Delta \log R \)-derived TOC results with (a) resistivity log and acoustic log, (b) resistivity log and neutron log, and (c) resistivity log and density log.

Figure 9. RBF prediction experiments with nine logs as input in the case study well: (a) The MSE and MAE both decrease with the Gaussian spread increase. (b) The MRE decreases and the number of neurons increases with the Gaussian spread increase. (c) Comparison between RBF-derived TOC and lab-measured TOC results with a correlation coefficient (R) of approximately 0.848 and a standard deviation of error of approximately 0.570.
deviation of absolute error between the predicted results based on the \( \Delta \log R \) method and lab measured results was approximately 1.4 wt.%, the standard deviation between the prediction results based on the \( \Delta \log R \) method and lab-measured results are both less than 1.236%, and the correlation coefficient (\( R \)) is greater than that of the corresponding \( \Delta \log R \) method. Therefore, it is indicated that the RBF method is more accurate than the \( \Delta \log R \) method in this field case.

Multi-input prediction experiments based on RBF

To study the influence of different input logs on the predicted results in the RBF prediction experiments, different log inputs in well A are chosen for the TOC prediction using the RBF method. The various experimental results are illustrated in Figure 5, and the prediction accuracy and the error analysis are listed in Table 2. From Figure 5 and Table 2, it can be seen that RBF experiments with GR/AC/DEN and GR/AC as inputs are both more accurate than other inputs. This indicates that the calculation precision is relative to the correlation coefficient of well logs. In fact, the coefficient of determination of GR and AC, namely, refer to linear correlation of determination between core-measured TOC and both wireline logs, are 0.7485 and 0.7125, respectively, which are the highest of all wireline logs (Figure 2).

In addition, Passey et al. (1990) make similar tests for wells B, C, D, E, and F. The RBF-predicted and the \( \Delta \log R \)-derived results are listed in Table 3. Generally, from the standard deviation of absolute errors between the predicted TOC from well logs and lab-measured TOC, the standard deviations of the RBF predicted TOC results in six wells are all lower than those of the \( \Delta \log R \)-derived results. This proves further that the RBF method is more accurate than the \( \Delta \log R \) method in six wells.

**CASE STUDY**

A case study well is a pilot borehole from the SINOPEC Company, which is located in the Pengshui area, Chongqing, China. The target formation is the Wufeng Formation of the lower Ordovician and Longmaxi Formation of up Ordovician, whose drilling goals include evaluating the hydrocarbon potential and gas-bearing content. Geophysical logs of this well include: natural GR, spontaneous potential (SP), dual laterolog (LLD + LLS), acoustic log (AC), litho-density log (DEN + PE), CNL and natural GR spectrometry logs including U, TH, and K. At the same time, 35 core samples were drilled in the target organic shale of this well and the TOC measurement of core samples was made in the laboratory. The crossplots of the experimental results and well logs are illustrated in Figure 6. It shows that the TOC has higher correlation coefficient with U, TH, and the TH/U ratio, natural GR, whereas, it has lower correlation with AC, DEN, PE, CNL, especially with true resistivity (Rt) generally displaced by LLD.

**TOC calculation based on the \( \Delta \log R \) method**

First, we apply the \( \Delta \log R \) method to calculate the TOC contents for this case study well. In this experiment, the resistivity baseline is set as approximately 21.5 ohm·m. When the acoustic is chosen to combine with resistivity log, the baseline is approximately 249 μs/m and the calibration coefficient is approximately ~0.01. When the neutron log is used, the baseline is approximately 19.1% and the calibration coefficient is approximately ~0.086. When the density log and resistivity are combined to calculate, the density baseline is 2.67 g/cm³ and the calibration coefficient is approximately 1.94. The results of three \( \Delta \log R \) methods in the target shale gas layer are illustrated in Figure 7. Figure 8 shows the crossplots of the calculated results and core lab-measured TOC. From the comparison, the results of the three \( \Delta \log R \) methods are all not in good agreement with lab-measured TOC, but the \( \Delta \log R_{\text{DEN}} \) method is better than the \( \Delta \log R_{\text{AC}} \) and the \( \Delta \log R_{\text{CNL}} \) methods.

Therefore, it is not good to evaluate the organic carbon content of the target shale gas layer using the \( \Delta \log R \) method and empirical formula. Hence, the forecast method using the RBF network is really necessary.

**TOC prediction experiments based on RBF**

Data preprocessing is indispensable. The logging data were first normalized before RBF data processing, and the resistivity logging data are taken in logarithmic scale.

<table>
<thead>
<tr>
<th>No.</th>
<th>Different inputs</th>
<th>RBF method</th>
<th>Gaussian spread</th>
<th>The number of neurons</th>
<th>MSE</th>
<th>MAE</th>
<th>MRE (%)</th>
<th>Correlation coefficient</th>
<th>Standard deviation of error (wt.%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>nine-log input (U/TH/GR/Rt/DEN/AC/K/PE/CNL)</td>
<td>0.22</td>
<td>30</td>
<td>0.331</td>
<td>0.415 28.133</td>
<td>0.848</td>
<td>0.570</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>eight-log input (U/TH/GR/Rt/DEN/AC/K/PE)</td>
<td>0.21</td>
<td>29</td>
<td>0.338</td>
<td>0.412 28.002</td>
<td>0.841</td>
<td>0.581</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>seven-log input (U/TH/GR/Rt/DEN/AC/K)</td>
<td>0.19</td>
<td>28</td>
<td>0.323</td>
<td>0.404 27.064</td>
<td>0.850</td>
<td>0.566</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>six-log input (U/TH/GR/Rt/DEN/AC)</td>
<td>0.19</td>
<td>26</td>
<td>0.310</td>
<td>0.394 26.572</td>
<td>0.856</td>
<td>0.556</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>five-log input (U/TH/GR/Rt/DEN)</td>
<td>0.13</td>
<td>29</td>
<td>0.329</td>
<td>0.398 26.186</td>
<td>0.846</td>
<td>0.573</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>four-log input (U/TH/GR/Rt)</td>
<td>0.16</td>
<td>23</td>
<td>0.319</td>
<td>0.400 26.514</td>
<td>0.851</td>
<td>0.565</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>three-log input (U/TH/GR)</td>
<td>0.23</td>
<td>10</td>
<td>0.387</td>
<td>0.461 33.22</td>
<td>0.816</td>
<td>0.622</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>two-log input (U/TH)</td>
<td>0.9</td>
<td>8</td>
<td>0.390</td>
<td>0.452 38.756</td>
<td>0.839</td>
<td>0.611</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>one-log input (U)</td>
<td>0.06</td>
<td>10</td>
<td>0.402</td>
<td>0.465 28.999</td>
<td>0.821</td>
<td>0.633</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Cross-validation based on leave-one-out in RBF training and testing (Freedman, 2006) are adopted. In the experiment, 35 samples are used for RBF neutral training in the different Gaussian widths, and the optimum Gaussian width parameter is determined by comparing the MSE, MAE, and MRE with different spreads, and the TOC prediction results are calculated accordingly.

Attribute ranking of sensitive logs is studied in advance. Based on logging data, the weight of an attribute is evaluated by repeatedly sampling an instance and considering the value of the given attribute for the nearest instance of the same class. The attribute ranking of sensitive logs is: U, TH, GR, LLD, DEN, AC, K, PE, and CNL. To investigate the experimental effects of different input logs, a series of RBF experiments is performed in accordance with the correlation between the logs and the lab measured TOC of cores.

First, the nine wireline logs, including U, TH, GR, LLD, DEN, AC, K, PE, and CNL, are chosen as inputs for the RBF experiment. Figure 9 shows the prediction experiment with different a Gaussian width or spread, Figure 9a illustrates the MSE and MAE with the Gaussian width increasing, and Figure 9b shows the MRE and the number of neurons of hidden layer with the Gaussian spread increasing. It can be seen that the MSE, MAE, and MRE first gradually decrease and then increase as the spread increases. The best value of the Gaussian width is selected in the inflection point, at which the spread is approximately 0.22 and the RBF network with the spread is optimum. The MSE and MAE are approximately 0.3309 and 0.415 wt.%, respectively, the MRE is approximately 28.133%, and the neuron number is approximately 30. Finally, the correlation coefficient \( R \) between the prediction results and core experiment results is approximately 0.848, and the standard deviation of error is approximately 0.570, which is illustrated in Figure 9c.

The authors have used a linear resistivity scale for the RBF experiment, and the mean absolute error and MRE are both larger than those with logarithmic resistivity. It is obvious that the prediction accuracy is significantly higher with resistivity in logarithmic scale than in linear scale. So, it is suggested that the resistivity log should be logarithmic scale for RBF prediction experiment.

According to the sensitivity ranking of the logs on organic carbon content, a series of different inputs is used to do a similar RBF experiment, and the experimental results are shown in Table 4.

From Table 4, it can be seen that when the inputting logs reduce, its Gaussian width becomes smaller to obtain the best prediction results, while the neuron number of the hidden layer will reduce correspondingly. From the mean absolute error and correlation coefficient between prediction results and core test results, when the six-log input (U/TH/GR/LLD/DEN/AC) is selected, the network for the TOC prediction is the best, at which the mean absolute error is minimum with the MAE of around 0.394 wt.% (see Figure 10a and 10b), and the RBF-predicted results are shown in the crossplot with lab-measured TOC in Figure 10c, the correlation coefficient is approximately 0.869, and the standard deviation of the error is approximately 0.547. Moreover, the experiment of different input variables can also get a good network model by selecting the appropriate Gaussian width, and the prediction accuracy is also higher. The correlation coefficients between RBF predicted results and core experiment results in Table 4 are all larger than the largest correlation coefficients of empirical formula of a single log in Figure 6. Therefore, the RBF method for the prediction of the TOC content is flexible.

So, when we use RBF method to predict the TOC content of the whole well, to achieve the correct evaluation of hydrocarbon potential, we will use all limited experimental samples to construct the optimum neutral network. Figure 8 shows some RBF prediction experiments with six logs as input in case study well: (a) The MSE and MAE both decrease with Gaussian spread increasing. (b) The MRE decreases and the number of neurons increases with the Gaussian spread increase. (c) The correlation coefficient between the RBF derived TOC and lab measured TOC results of cores is approximately 0.869, and the standard deviation of the error is approximately 0.547.
experiment results with four, six, seven, and nine logs, which are recorded as TOC4, TOC6, TOC7, and TOC9, respectively. In addition, we use two empirical formulas of U and TH/U ratio with a high correlation coefficient to calculate the TOC curve of the whole well; the prediction results above are all listed in the right two tracks in Figure 11.

The right two tracks in Figure 11 illustrate that the contrast of the RBF prediction results with the result calculated by the empirical formula method, the identical degree of the forecast results (TOC4, TOC6, TOC7, and TOC9) based on the RBF method, is better than that based on the empirical formula of the U and TH/U ratio. According to the calculated TOC value, the target layer is classified: The organic carbon content of interval X057.6 to X137.5 m is low, and the hydrocarbon potential is poor; its TOC value is approximately 1.0 wt.%. The TOC value of the interval from X057.6 to X071.5 m is greater than 1.0 wt.%, which is highest among the interval from X137.5 to X160.4 m, namely, the no. 2 layer, the TOC value is approximately 3–4 wt.%, and the hydrocarbon-generating potential is the best.

DISCUSSION

We used the RBF method for the TOC prediction of one study area. First, we can use lab-measured TOC results of limited cores from the same layers of different wells to construct the optimal network, and then we apply the optimal network to predict TOC

![Figure 11. TOC prediction results with various logs as inputs based on the RBF method in the case study well. The right two tracks show the RBF-predicted results with different inputs, TOC prediction curves from empirical formulas relative to the TH/U ratio and U, and lab-measured TOC of cores, which is displayed as discrete data.](image-url)
contents of other wells in the study area. But in the Pengshui area, the case study well is a pilot borehole, and other wells are not available. So, we use all core experimental results of samples to perform a series of RBF experiments.

Obviously, the RBF method is dependent of the experimental results of the core samples, and the training network is applicable to the formation where the training sample data are. So, the method still has some regional limitations. Therefore, the sample points for training the RBF network should cover different lithologies, and low- and high-TOC sections, and such a RBF network will have good generalization ability.

However, compared to the BP neural network method, the RBF method doesn’t need to determine the number of hidden layer in advance, and it has only one hidden layer, so it is steadier and much more practical.

CONCLUSIONS

Through the research and analysis on the RBF method in the organic shale formation, the following conclusions can be drawn:

• The RBF network is a feed forward network with good performance, and it has global approximation properties and better approximation performance. The learning algorithm of the RBF neural network is an online self-adaptive cluster learning algorithm, and the training method is rapid and feasible.

• It can easily construct the mapping between various logs and TOC contents for hydrocarbon potential assessment through the optimization of Gaussian spread. When the TOC content of organic shale is calculated, the optimal Gaussian width is an important parameter and is determined by multitime cross-validation before the RBF method is used to predict the TOC contents.

• The TOC prediction using the RBF method for organic shale may be achieved by different sensitive logs. The correlation coefficient of the RBF experiments with different input variables is highest at the optimal Gaussian spread. When the sensitive logs to the TOC are chosen as input variable, the TOC prediction results can also achieve high accuracy through optimizing the Gaussian spread, and it is also better than the empirical formulas method. The error analysis indicates that the RBF method is considered as the integrated responses of a variety of logs; therefore, it is more accurate than the Δ log R method and a simple empirical formula in this case.

ACKNOWLEDGMENTS

This research proposal is sponsored by the National Natural Science Foundation of China (41172130), the Fundamental Research Funds for the Central Universities (265201248), and the National Major Projects Development of Major Oil & Gas Fields and Coal Bed Methane (2011ZX05014-001), CNPC Innovation Foundation (2011D-5006-0305). Furthermore, M. Tan would like to acknowledge the support from the China Scholarship Council (no. 201206405002). A special note of gratitude goes to Dr. C. Torres-Verdin and Dr. C. Xu at the University of Texas at Austin for their technical help.

REFERENCES


