ARCHIVES OF ELECTRICAL ENGINEERING VOL. X(Y), pp. 1-20 (2026)

This paper has been accepted for publication in the AEE journal. This is the version, which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.24425/aee.2026.1524

Temporal correlation-aware prediction and early warning method for dissolved gases in transformer oil

JIAN SHAO, XUAN LI*, TONGLEI WANG, YONG MA, YUANDI LIN, PENG WU, YUNCAI LU

Electric Power Research Institute, State Grid Jiangsu Electric Power Co., Ltd.
Nanjing 211103, China
*e-mail: *xuanya183174@163.com*

Abstract: A method is proposed for predicting and warning the dissolved gas content in transformer oil. This method combines the Granger causality test and neural basis expansion analysis with exogenous variables, addressing dynamic coupling and complex temporal correlations in gas component data. First, the Granger causality test is used to analyze temporal correlations in gas component concentrations and their relationship with transformer load, to identify mutual influence between time-series data and achieve dynamic variable selection of the prediction model. Then a dissolved gas prediction model is established using neural basis expansion analysis with exogenous variables. Finally, transformer status warning is achieved under threshold constraints by combining confidence interval distribution analysis of predicted gas component concentrations. The results of a case study show that this model can achieve a mean absolute percentage error below 5.0%. Dynamic variable screening based on Granger causality verification can significantly enhance prediction accuracy, providing a more reliable reference for early warning of transformer status.

Keywords: dissolved gas analysis, Granger causality, NBEATSx, prediction model, time-series correlation

1. Introduction

As an important fundamental equipment in the power system, the status of power transformers directly determines the operation safety and stability of the grid. Dissolved gas analysis (DGA), as a commonly used method to analyze transformer defects and fault diagnosis, is also an important indicator for reflecting equipment operating status. Based on the historical data from online monitoring of dissolved gases in transformer oil, accurately predicting the changing trends of gas component concentrations can help timely grasp the evolution law of equipment status and achieve early warning of equipment defects [1–2].

This paper has been accepted for publication in the AEE journal. This is the version, which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.24425/aee.2026.1524

The methods for predicting dissolved gases in transformer oil can be divided into two categories: one is statistical prediction methods based on traditional statistical analysis, and the other is artificial intelligence prediction methods based on machine learning or deep learning. Statistical prediction methods mainly include the grey model [3] and time-series models [4–5] characterized by mature theory, simple models, and fast operation speed. However, due to their reliance on a single fitting, their poor performance in fitting and predicting nonlinear and non-stationary data (such as dissolved gases in transformer oil) is poor. Early artificial intelligence prediction methods were mainly machine learning models, including the support vector machine (SVM) [6], extreme learning machine (ELM) [7], and random forest (RF) [8]. Compared with statistical prediction methods, machine learning methods have advantages in dealing with nonlinear and non-stationary data. However, most studies have not taken into account the temporal correlations in time-series data, resulting in limited predictive effectiveness.

A series of deep learning methods such as recurrent neural networks (RNNs) [9] can capture temporal features and show broad application prospects in predicting dissolved gas content in transformer oil. The RNN uses a recurrent feedback network structure to effectively capture temporal dependencies in time-series data. However, the cumulative effect caused by derivatives during backpropagation can easily lead to gradient vanishing or explosion. To address these problems, the long short-term memory (LSTM) [10] introduces a memory module and utilizes gating units for adaptive regulation, effectively reducing gradient vanishing and explosion. Combined the empirical mode decomposition (EMD) with the LSTM, Liu *et al*. [11] proposed a prediction method for dissolved gases in transformer oil. Using EMD to decompose the gas content sequence can significantly improve the prediction performance of the LSTM model. Combined variational mode decomposition (VMD) with gated recurrent unit (GRU) neural networks, Xie *et al*. [12] introduced a prediction model to eliminate the impact of non-stationarity in dissolved gas data on the accuracy of VMD prediction. Cui *et al*. [13] adopted an LSTM model integrated with dual attention mechanisms of features and time-series. This approach enables automatic mining of correlations between the target gas and other state information (such as environmental and operational data). Liu *et al*. [14] proposed an adaptive prediction method for dissolved gas content in transformer oil considering multiple factors. The weights of predictors were calculated through an attention mechanism, and the sparrow search algorithm (SSA) was used to improve the prediction accuracy of the bidirectional gated recurrent unit (BiGRU) model. In addition to using attention mechanisms and other "black-box models", there is relatively little research on dissolved gas prediction. These studies explore the dynamic coupling and temporal correlations of component gases, making it difficult to directly determine causal relationship between gas contents and transformer operating conditions. Most studies have focused on trend analysis of gas content, without connecting it with the assessment of transformer operating status, providing limited guidance for the practical operation of field transformers.

To this end, based on Granger causality (GC) tests and the NBEATSx (neural basis expansion analysis with exogenous variables) model, a prediction and early warning method is proposed for dissolved gases in transformer oil. This method can predict dissolved gases after decoupling

This paper has been accepted for publication in the AEE journal. This is the version, which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.24425/aee.2026.1524

component gases and assess equipment status. First, the Granger causality test is used to screen key influencing variables for predicting different gas components and establish an exogenous variable matrix. Then a prediction model for dissolved gas content in transformer oil is built using the NBEATSx driver. At the same time, the KDE (kernel density estimation) method is used to calculate and form confidence interval predictions for various gas components. Finally, a method based on the dissolved gas threshold standard is proposed to determine the operating trend of transformers, which can provide early warnings for equipment operating status.

2. Time-series data correlation analysis

2.1. Granger causality test

The Granger causality test is a statistical test method based on time-series data [15], is used to explore predictive causal relationship between variables. Under stationary time-series, GC verifies causality by constructing a vector autoregressive model (VAR): assuming two-time series variables *X* and *Y*, if the historical lagged observations of *X* can significantly improve the predictive accuracy of future values of *Y*, *X* is determined as the Granger cause of *Y*; *X* is defined as the cause variable and *Y* as the effect variable [16]. The *q*-th order VAR model for time-series variables *X* and *Y* is expressed as follows:

$$X_t = c_1 + \sum_{i=1}^{q} \alpha_{1i} Y_{t-i} + \sum_{j=1}^{q} \beta_{1j} X_{t-j} + u_{1t}, \tag{1}$$

$$Y_t = c_2 + \sum_{i=1}^{q} \alpha_{2i} X_{t-i} + \sum_{j=1}^{q} \beta_{2j} Y_{t-j} + u_{2t}, \tag{2}$$

where the lag order q reflects the delayed effect of influence between variables [17]; c_1 and c_2 are constant terms; X_t and Y_t represent the time-series variables *X* and *Y* at time *t*; α_{1i}, α_{2i}, β_{1i} and β_{2i} are the lagged autoregressive coefficients of the variables; u_{1t} and u_{2t} are the error terms.

Taking *Y* as an example, when testing whether *X* is a Granger cause of *Y*, if the lagged autoregressive coefficients in Eq. (2) are all zero, it indicates that the current value of *Y* is only predicted based on its own historical values. This model is called as the unrestricted model. Conversely, if not all of these coefficients are zero, it signifies that historical information about *X* is introduced to predict the current value of *Y*. This model is termed as the restricted model. Using the null hypothesis that "*X* is not a Granger cause of *Y*", the statistic $F_{x\to y}$ is composed of the sum of squared residuals calculated using the unrestricted and restricted models:

$$E_{\mathrm{RSS}} = \sum_{i=1}^{n} (y_i - \tilde{y}_i)^2, \tag{3}$$

where E_{RSS} represents the residual sum of squares; y_i and $\tilde{y}_i$ are the true value and predicted value of the sample, respectively.

This paper has been accepted for publication in the AEE journal. This is the version, which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.24425/aee.2026.1524

$$F_{x\to y} = \frac{(E_{\mathrm{RSS0}} - E_{\mathrm{RSS1}})/q}{E_{\mathrm{RSS1}}/(n-k)}, \tag{4}$$

where n is the sample size of the time-series data; $k = 2q + 1$ is the total number of parameters to be estimated in the unrestricted model; E_{RSS0} and E_{RSS1} are the residual sums of squares obtained from the unrestricted model and its restricted counterpart, respectively. $F_{x\to y}$ is a statistic that follows the *F*-distribution. The validity of the null hypothesis is determined by comparing its corresponding *P*-value (i.e., the probability of achieving the current result when the null hypothesis is assumed to be true) with the selected significance level (adopted as 0.05). Should the *P*-value be smaller than the significance level, the null hypothesis is rejected; on the contrary, if it exceeds the significance level, the null hypothesis is retained.

2.2. Analysis of temporal correlation of gas components

When there are latent defects in a transformer, the insulating oil undergoes chemical decomposition under electrical and thermal stress, generating hydrocarbon gases [18]. As temperature increases, hydrocarbon gases are generated sequentially in the following order: methane, ethane, ethylene, and acetylene. However, due to the types of defects, gas dissolution, and diffusion conditions, there are some differences in the composition and dynamic coupling characteristics of dissolved gas components in transformer oil [19]. Transformer defect diagnosis can be achieved based on the ratio relationships between the concentrations of these gases, although most analysis rely on data from specific points in time. In addition, in studies that focus on predicting dissolved gas concentrations, the analysis of coupling relationship is relatively limited. There is little attention paid to the temporal correlations between gas components. Therefore, the Granger causality test is introduced to analyze the temporal correlations between the dissolved gas components and their relationship with the transformer load rate. On one hand, it can help screen effective features to enhance prediction accuracy; on the other hand, it can utilize the correlations to identify potential defect types within the transformer.

Taking the in-service transformer of a 500 kV substation in China as an example, using the ODFS-334000/500 model, the oil-dissolved gas online monitoring data and the load factor variation curve are shown in Fig. 1. The time interval between adjacent data points of all curves is 12 h, indicating that the graph shows the data changes of this main transformer over 117 d. As shown in Fig. 1, the hydrocarbon gases in this main transformer have increased, indicating that there may be latent defects in the transformer.

This paper has been accepted for publication in the AEE journal. This is the version, which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.24425/aee.2026.1524

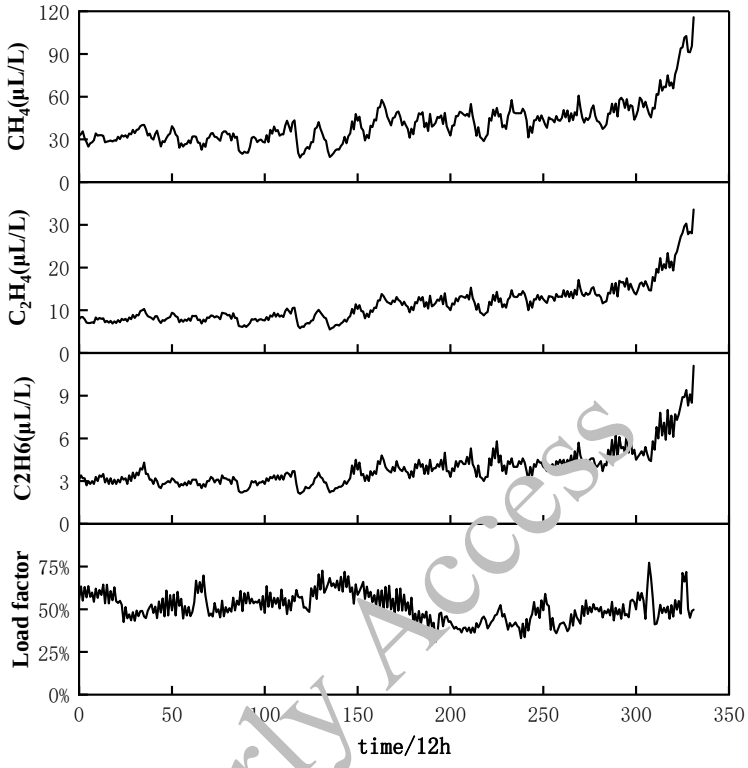


Fig. 1. Variation curve of dissolved gases in transformer oil and load rate

Using three hydrocarbon gases as the outcome variables, a lag order of 3 was selected after normalization. The normalization uses Min-Max normalization to map the data at the interval [0, 1]:

$$x_{\text{norm}} = \frac{x - x_{\min}}{x_{\max} - x_{\min}}, \tag{5}$$

where, x is the original concentration value of a certain gas; $x_{\min}$ is the minimum concentration value of this gas, and $x_{\max}$ is the maximum concentration value of this gas. Figure 2 shows the normalization results of methane gas.

This paper has been accepted for publication in the AEE journal. This is the version, which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.24425/aee.2026.1524

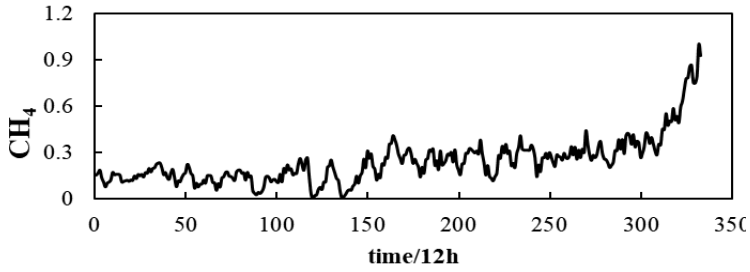
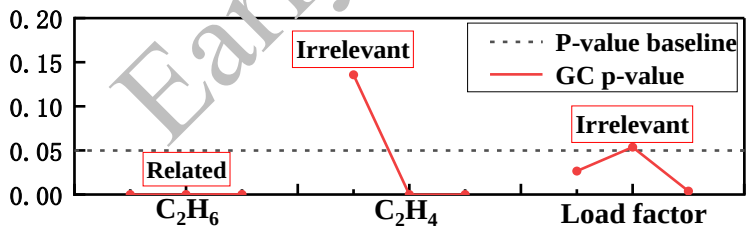
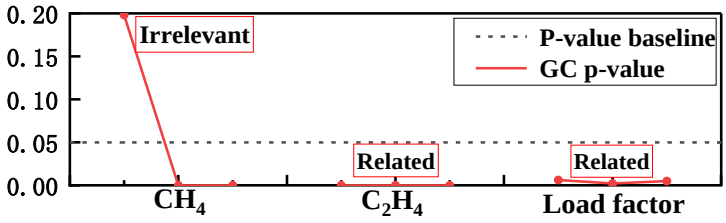


Fig. 2. Normalization results of the methane gas

The Granger causality test is used to sequentially analyze the temporal correlations between gas components and their relationship with load rate, as shown in Fig. 3. A lag order of 3 is selected to determine whether the current value of the gas component content is influenced by the historical values of other time-series data over the past three time points. Therefore, the delayed response characteristics of hydrocarbon gases to load rate variations caused by dissolution and diffusion are taken into account. If the lag order is too small, it may overlook critical correlation information in the time-series data; if it is too large, it may introduce interference noise. In engineering applications, the 3rd-order lag is often adopted as a general benchmark for the Granger causality test. The selection of the lag order should be based on the sampling interval and oil flow conditions. The Akaike information criterion (AIC), Bayesian information criterion (BIC) and the regularization theory should be used to determine it.

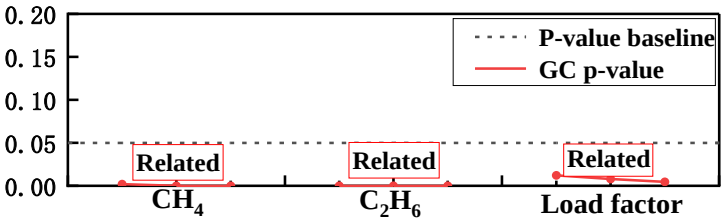


(a) Methane as the outcome variable



(b) Ethane as the outcome variable

This paper has been accepted for publication in the AEE journal. This is the version, which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.24425/aee.2026.1524



(c) Ethylene as the outcome variable

Fig. 3. Correlation analysis results of time-series data

Figure 3 shows that when used as the outcome variable, methane is only correlated with ethane. Ethane, in turn, is correlated with ethylene and the load rate, and ethylene has correlations with methane, ethane, and the load rate. It should be noted that the Granger causality test results are temporal prediction correlations at the statistical level, rather than a strict physical causal mechanism. Only the auxiliary predictive effect of the previous variables on the target variable are characterized. Figure 3 only shows that combining historical values of ethane may provide some auxiliary predictive power for predicting methane content. In this case, the temporal influence of ethylene and the load rate can be ignored. In addition, there is no temporal correlation between methane and the load rate. This means that the change in the main component gas, methane, is not a lagged effect in the load rate. This indicates that the total hydrocarbon gas content is unrelated to the load rate of the transformer, effectively eliminating overheating defects caused by the change in transformer load.

3. Transformer oil dissolved gas content prediction method

3.1. NBEATSx

NBEATS (neural basis expansion analysis for interpretable time-series forecasting) is a model specifically designed to address the problem of point prediction in time-series. It is a fully connected deep neural network (FC-DNN) architecture with forward and backward residual connections [20]. The NBEATSx model enhances NBEATS through exogenous variables, achieving the integration of multiple valuable information sources [21–22]. NBEATSx achieves the decomposition of the expected sequence by nonlinearly mapping fragments of the target time-series to basic functions in different modules. Basis vectors and their corresponding coefficient parameters can be obtained by training. The prediction results of each decomposed series are calculated, and then the final prediction is reconstructed by superimposing these individual results. Compared with neural network models such as the LSTM, NBEATSx has a simpler structure and better interpretability, making it a predictive model with greater applicability and stronger generalization. NBEATSx is composed of stacks and blocks, as shown in Fig. 4. The stack is formed by combining multiple blocks based on the principle of dual

This paper has been accepted for publication in the AEE journal. This is the version, which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.24425/aee.2026.1524

residual stack. The basic building block of this model is a multi-layer FC network containing rectified linear units (ReLU).

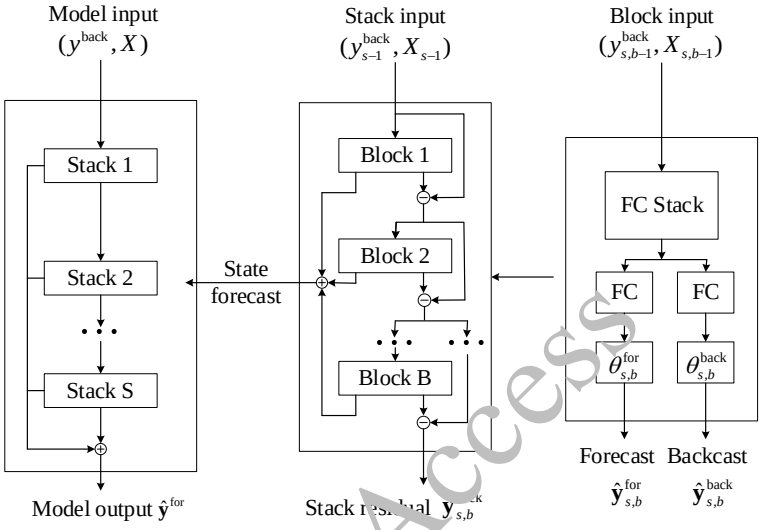


Fig. 4. Architecture diagram of the NBEATSx model

The *b*-th block within the s-th stack of the NBEATSx model consists of two components. The first is a fully connected (FC) network designed to calculate the forward and backward expansion coefficients factors θ for *s*, *b* and θ back *s*, *b*. The calculation process is as follows:

$$\begin{cases} \boldsymbol{h}_{s,b,i} = ReLU_{s,b,i}\left(\boldsymbol{W}_{s,b,i}\boldsymbol{x}_{s,b,i} + \boldsymbol{b}_{s,b,i}\right) \quad (i = 1,2,3,4) \\ \theta_{s,b}^{\text{for}} = \text{LINEAR}_{s,b}^{\text{for}}\left(\boldsymbol{h}_{s,b,4}\right) = \boldsymbol{W}_{s,b}^{\text{for}}\boldsymbol{h}_{s,b,4} \\ \theta_{s,b}^{\text{back}} = \text{LINEAR}_{s,b}^{\text{back}}\left(\boldsymbol{h}_{s,b,4}\right) = \boldsymbol{W}_{s,b}^{\text{back}}\boldsymbol{h}_{s,b,4} \end{cases}, \tag{6}$$

where $\boldsymbol{h}_{s,b}$ represents the hidden state; the LINEAR layer is a simple linear projection layer. The second consists of the forward basis function *g* for *s*, *b* and the backward basis function *g* back *s*, *b*. They are used to calculate the backward prediction output $\hat{\boldsymbol{x}}_{s,b}$ and the forward prediction output $\hat{\boldsymbol{y}}_{s,b}$. The calculation process is as follows:

$$\begin{cases} \hat{\boldsymbol{x}}_{s,b} = g_{s,b}^{\text{back}}\left(\theta_{s,b}^{\text{back}}\right) = \sum_{i=1}^{\dim\left(\theta_{s,b}^{\text{back}}\right)} \theta_{s,b,i}^{\text{back}} \boldsymbol{V}_i^{\text{back}} \\ \hat{\boldsymbol{y}}_{s,b} = g_{s,b}^{\text{for}}\left(\theta_{s,b}^{\text{for}}\right) = \sum_{i=1}^{\dim\left(\theta_{s,b}^{\text{for}}\right)} \theta_{s,b,i}^{\text{for}} \boldsymbol{V}_i^{\textbf{for}} \end{cases}, \tag{7}$$

where $\boldsymbol{V}_i^{\text{back}}$ and $\boldsymbol{V}_i^{\text{for}}$ are the backward basis vector and the prediction basis vector, respectively.

As shown in Fig. 4, NBEATSx uses a hierarchical dual residual topology. This topology utilizes two residual branches to calculate the backward and forward prediction vectors for each layer, respectively. The calculation process is as follows:

This paper has been accepted for publication in the AEE journal. This is the version, which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.24425/aee.2026.1524

$$\begin{cases} \boldsymbol{x}_{s,b+1} = \boldsymbol{x}_{s,b} - \hat{\boldsymbol{x}}_{s,b} \\ \hat{\boldsymbol{y}} = \sum_{s=1}^{S} \hat{\boldsymbol{y}}_s = \sum_{s=1}^{S} \sum_{b=1}^{B} \hat{\boldsymbol{y}}_{s,b} \end{cases}, \tag{8}$$

where, $x_{s,b+1}$ and $x_{s,b}$ are the residual outputs of two adjacent blocks; $\hat{\boldsymbol{y}}_s$ is the forward prediction vector of each stack, and $\hat{\boldsymbol{y}}$ is the final prediction result vector.

To improve the prediction accuracy and interpretability of oil-dissolved gas time-series data, the NBEATSx model is divided into a trend stack and a seasonal stack to breakdown the target data into trend and seasonal components for independent processing. Specifically, the trend stack is used to characterize the historical variation trend of component gases, where the basis vectors *g* for *s*, *b* and *g* back *s*, *b* are set as slowly changing *P*-th degree polynomial functions:

$$\hat{\boldsymbol{y}}_{s,b} = \sum_{i=0}^{P} t^i \theta_{s,b,i}^{\text{for}} = \boldsymbol{T}\theta_{s,b}^{\text{for}}, \tag{9}$$

where the time vector $t = [0, 1, 2, \ldots, H\text{-}2, H\text{-}1]^T/H$, and H is the prediction horizon; $\boldsymbol{T} = [1, t, t^2, \ldots, t^p]$ is the power matrix of t.

The seasonal stack mainly characterizes the dynamic equilibrium characteristics of adsorption and desorption of dissolved gases in oil under the influence of periodic temperature changes. Its basis vectors *g* for *s*, *b* and *g* back *s*, *b* are represented using Fourier series functions:

$$\hat{\boldsymbol{y}}_{s,b} = \sum_{i=0}^{[H/2-1]} \cos(2\pi \boldsymbol{it})\theta_{s.b.i}^{\text{for}} + \sin(2\pi \boldsymbol{it})\theta_{s.b.i+[H/2]}^{\text{for}} = \boldsymbol{S}\theta_{s.b.i}^{\text{for}}, \tag{10}$$

where, $\boldsymbol{S}$ is the sine wave matrix.

In addition, except the reciprocal influences between different gas components, the concentration of dissolved gases in the oil can also be affected by external factors such as ambient temperature and transformer load. Therefore, an exogenous stack is introduced into the NBEATSx model. For the exogenous variable matrix *X*, the basic vectors *g* for *s*, *b* and *g* back *s*, *b* of the exogenous stack are set as functions related to the exogenous variable matrix:

$$\hat{\boldsymbol{y}}_{s,b} = \sum_{i=0}^{Nx} \boldsymbol{X}^i \theta_{s,b,i}^{\text{for}} = \boldsymbol{X}\theta_{s,b}^{\text{for}}, \tag{11}$$

where $\boldsymbol{X} = [\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_N]$, and *N* represents the quantity of exogenous variables.

3.2. KDE

KDE is a non-parametric estimation method that is independent of prior information about data distribution. Its outputs are fully determined by the distribution of the sample data [23]. Suppose independent and identically distributed sample points x_i ($i = 1, 2, \ldots, n$) are sampled from a continuous distribution *X*, the KDE of the probability density function $f(x)$ at an arbitrary point *x* is expressed as follows:

$$\hat{f}_d(x) = \frac{1}{Nd} \sum_{i=1}^{N} K\left(\frac{x - x_i}{d}\right), \tag{12}$$

where *N* is the total number of samples; *d* is the bandwidth; $K(\cdot)$ is the kernel function, and we adopt the Gaussian kernel function:

This paper has been accepted for publication in the AEE journal. This is the version, which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.24425/aee.2026.1524

$$K(u) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-u^2}{2}\right). \tag{13}$$

The kernel density estimation (KDE) algorithm is used to calculate the probability density of prediction errors for each gas component and evaluate interval prediction at different confidence levels to determine a suitable confidence level and interval prediction results. According to the industry standard DL/T 722-2014 [24], which specifies the limit of dissolved gas content in transformer oil during operation, the confidence interval prediction results of each component gas are superimposed and reconstructed to achieve early warning of transformer operation status. Specifically, for 500 kV power transformers, the attention threshold for the total hydrocarbon content is set to 150 μL/L as the corresponding limit.

3.3. Evaluation indicators

To evaluate the predictive performance of dissolved gas content in transformer oil, three key indicators: mean absolute percentage error (MAPE), root mean square error (RMSE), and coefficient of determination (R^2), are used. These metrics are expressed as follows:

$$E_{\mathrm{MAPE}} = \frac{1}{n}\sum_{i=1}^{n} \frac{|y'_i - y_i|}{y_i} \times 100\%, \tag{14}$$

$$E_{\mathrm{RMSE}} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y'_i - y_i\right)^2}, \tag{15}$$

$$R^2 = \frac{\sum_{i=1}^{n}\left(y'_i - \bar{y}_i\right)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2}, \tag{16}$$

where n represents the count of test data samples; y_i denotes the true value; y'_i is the forecast value, and $\bar{y}_i$ is the mean of the actual values across all test data. To further assess the interval prediction of gas content values, two evaluation indicators: prediction interval coverage probability (PICP) and prediction interval normalized average width (PINAW), are used. The indicators are calculated as follows:

$$E_{\mathrm{PICP}} = \frac{\xi}{N} \times 100\%, \tag{17}$$

$$E_{\mathrm{PINAW}} = \frac{1}{N(y_{\max} - y_{\min})}\sum_{i=1}^{N}\theta_i \times 100\%, \tag{18}$$

where N represents the count of test samples; ξ denotes the number of actual values falling within the prediction interval; $y_{\max}$ and $y_{\min}$ correspond to the maximum and minimum of the actual values, respectively, and θ_i stands for the prediction interval width of the i-th test sample.

3.4. Multimodal fusion

Based on the above methods, the Granger causality test is introduced to screen for significant influencing factors in the target gas prediction. It adopts the NBEATSx algorithm to consider the comprehensive impact of subsequences at different time scales and external factors on the

This paper has been accepted for publication in the AEE journal. This is the version, which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.24425/aee.2026.1524

prediction results. In addition, it combines kernel density estimation to achieve interval prediction for component gases and equipment status early warning. The application flowchart is shown in Fig. 5.

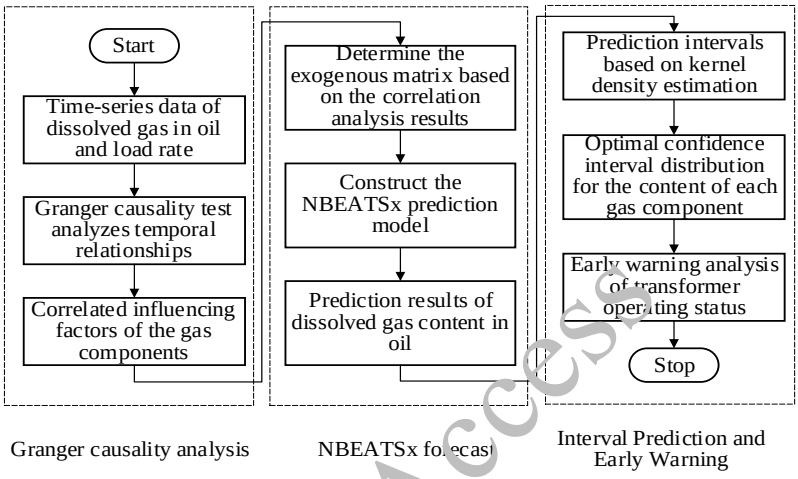


Fig. 5. Flow chart for prediction and early warning of dissolved gases in transformer oil

1) Collect online monitoring datasets of the load rate and dissolved gas content in transformer oil to form time-series data on a unified time scale;
2) Using the Granger causality test to investigate temporal correlations between component gas content and its correlation with load rate data, and screen relevant influencing factors when predicting each component gas;
3) Based on the results of the Granger causality test, determine the model inputs corresponding to each target gas, that is, the dimension of the exogenous matrix and its constituent data;
4) Combine the time-series data of the target gas with its associated exogenous matrix data, divide the dataset into training and test sets, and construct an NBEATSx prediction model;
5) Based on the test set data, the predictive results of dissolved gas content in oil are obtained;
6) Using kernel density estimation to calculate the prediction interval distribution of gas content for each component, and obtain the optimal confidence interval distribution for each component gas;
7) Combining the Granger causality test with standard regulatory limits and the interval prediction results to provide early warnings for transformer operating status and output the warning results.

This paper has been accepted for publication in the AEE journal. This is the version, which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.24425/aee.2026.1524

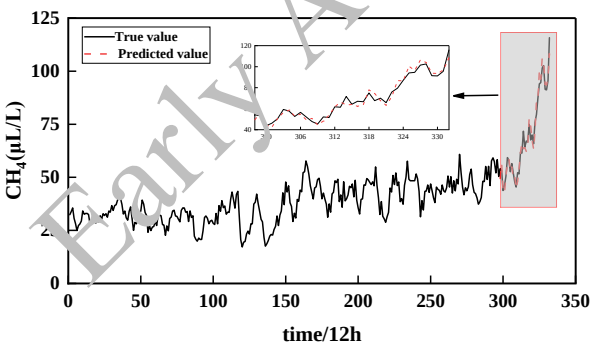
4. Experiments and analysis of results

4.1. Case 1: Prediction and early warning based on three hydrocarbon gases

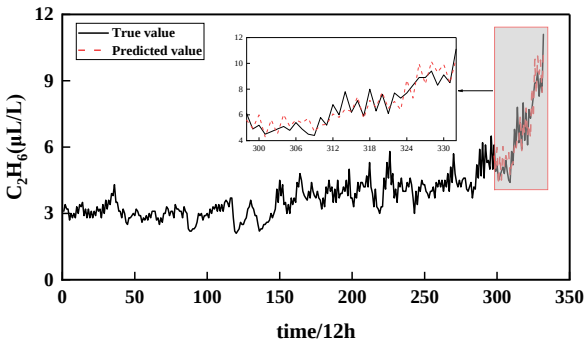
Based on dissolved gas time-series data and load rates in transformer oil (in Fig. 1), normalization results and Granger causality tests, prediction and early-warning models are used to compare the algorithms.

According to the Granger causality test results shown in Fig. 3, when predicting methane gas content, time-series data of ethane content can be used to construct the exogenous matrix; when predicting ethane gas content, time-series data of ethylene content and load rate can be used to construct the exogenous matrix; when predicting ethylene gas content, time-series data of methane content, ethane content, and load rate can be used to construct the exogenous matrix.

The NBEATSx forecasting model is established with a block count of 3, a neuron counts of 256, and a batch size of 32. Ultimately, the forecasting results of the three gas component concentrations are shown in Fig. 6, and the corresponding indicator results are shown in Table 1. The findings demonstrate that the predicted values of each gas component are consistent with the actual values. This indicates that the model can effectively reveal the variations of dissolved gases in oil, providing a technical reference for predicting the development trends of component gas.

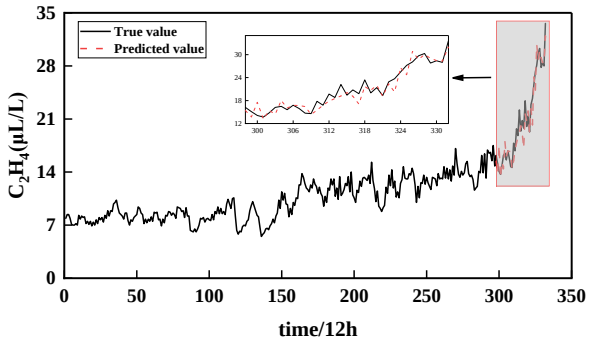


(a) Predicted results of methane



(b) Predicted results of ethane

This paper has been accepted for publication in the AEE journal. This is the version, which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.24425/aee.2026.1524



(c) Predicted results of ethylene

Fig. 6. Prediction results of GC-NBEATSx

Table 1. Prediction performance of GC-NBEATSx models

Gas	E_{MAPE}	E_{RMSE}	R^2
CH_4	4.37%	3.75	0.965
C_2H_6	8.55%	0.68	0.848
C_2H_4	6.01%	1.59	0.919

Taking methane gas as an example, the GC-NBEATSx model is used to train the prediction error of the dataset, and the kernel density estimation (KDE) method is used to calculate interval prediction results at different confidence levels. The bandwidth parameter d is sequentially set to 0.1, 0.25, 0.4517, 1 and 2 to plot the frequency histogram of methane gas prediction errors and the KDE probability density curves under different bandwidths, as shown in Fig. 7. The bandwidth parameter $d = 0.4517$ is derived using cross-validation by optimizing the mean integrated squared error (MISE).

As shown in Fig. 7, the bandwidth parameter d in KDE significantly influences the shape characteristics of the probability density curve. If the bandwidth parameter is too small, the probability density curve has local non-physical oscillations; if it is too large, it cannot capture the detailed features of error distribution. As the bandwidth d increases, the sensitivity of the probability density curve to the dynamic characteristics of the prediction error frequency distribution gradually decreases.

The bandwidth parameter $d = 0.4517$ optimized by the MISE criterion achieves a balanced representation of global distribution trend and local feature characterization, significantly improving the accuracy of the KDE method. Taking $d = 0.4517$ as the bandwidth parameter, the KDE method is applied to generate interval prediction results at different confidence levels, as shown in Fig. 8. The evaluation metrics for the interval predictions are shown in Table 2.

This paper has been accepted for publication in the AEE journal. This is the version, which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.24425/aee.2026.1524

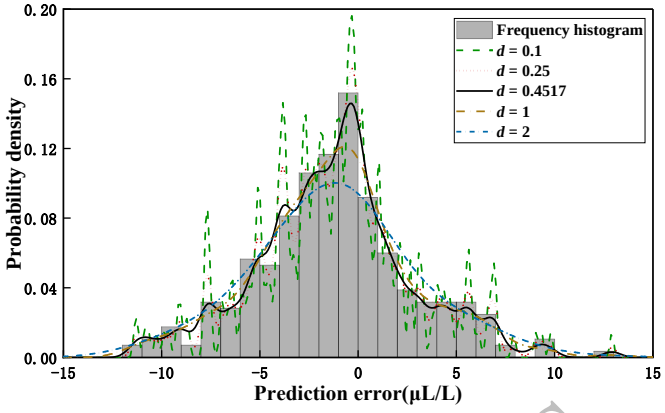


Fig. 7. KDE probability density curves of methane gas prediction errors

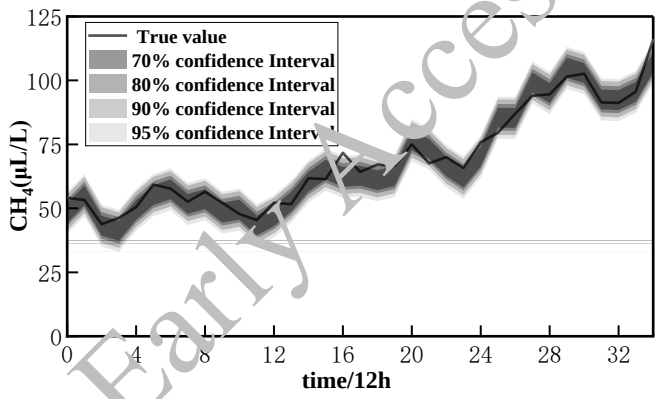


Fig. 8. Interval prediction results of methane gas under different confidence levels

Table 2. Evaluation results of methane gas interval prediction under different confidence levels

Confidence levels	E_{PICP}	E_{PINAW}
95%	0.943	0.267
90%	0.943	0.226
80%	0.771	0.169
70%	0.657	0.126

Figure 8 and Table 2 show that a low confidence level leads to a narrow prediction interval, and increases the probability of the prediction based on the true methane gas content. Conversely, an excessively wide interval will significantly reduce the predictive accuracy. Therefore, when applying confidence interval predictions, it is essential to balance these two metrics: ensuring a

This paper has been accepted for publication in the AEE journal. This is the version, which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.24425/aee.2026.1524

high interval coverage probability and minimizing the average interval width as much as possible. According to Table 2, at a 90% confidence level, the interval coverage probability remains the same as that at the 95% confidence level, but with a smaller average interval width. Therefore, the prediction results at a 90% confidence level are selected as the final confidence interval prediction.

Using the above method, the confidence interval predictions for ethane and ethylene gas contents under the GC-NBEATSx model can be obtained. They are then reconstructed and superimposed with the methane prediction results to derive the total hydrocarbon confidence interval predictions, as shown in Fig. 9.

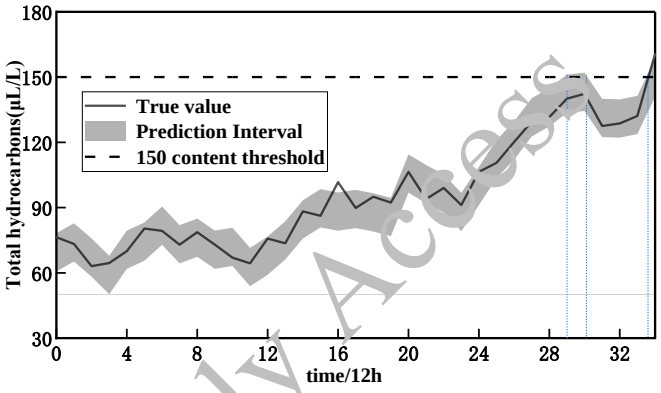


Fig. 9. Confidence interval prediction for total hydrocarbons in Case 1

As shown in Fig. 9, the prediction interval of the total hydrocarbons basically covers the true values. This indicates that the confidence interval prediction results under the GC-NBEATSx model can effectively reflect the change of trend in total hydrocarbon gas content. The model sets a threshold of 150 μL/L for the total hydrocarbon content of the 500 kV transformer. An alert signal is triggered at the 29th data point, lasting for more than one day. It is about 2 days earlier than the warning based on the true value. Based on the analysis results of the Granger causality test, it is determined that the excessive total hydrocarbons in this main transformer are not related to overheating defects. The disassembly inspection revealed inter-turn short circuits in the medium-voltage winding, indicating a clear discharge. This method can effectively uncover the potential correlations within transformer time-series data. Taking into account the changing patterns of time-series data, it is possible to predict the development trend of equipment status, thereby achieving early warning of abnormal equipment conditions.

4.2. Case 2: Comprehensive prediction using seven dissolved gases

The experimental data for Case 2 were from the dissolved gases in oil of a 500 kV transformer from June 14 to September 10, 2023. The dataset consists of 179 time-series sample points, with a sampling interval of 12 h. The monitored indicators include seven dissolved gases

This paper has been accepted for publication in the AEE journal. This is the version, which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.24425/aee.2026.1524

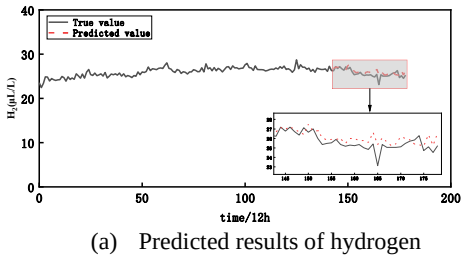
(H_2, CH_4, C_2H_2, C_2H_4, C_2H_6, CO, CO_2) and the transformer load rate. A lag order of 3 was selected, and the temporal correlations between each of the seven gases and the load rate were analyzed sequentially. The results are shown in Table 3.

Table 3. Significant Granger causality relationships among transformer DGA gases and load factor

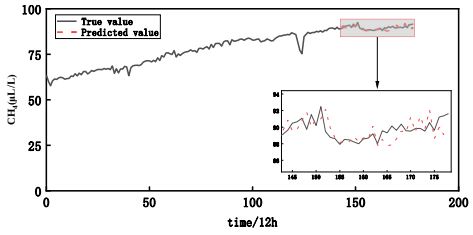
Detection Indicators	H_2	CH_4	C_2H_2	C_2H_4	C_2H_6	CO	CO_2	Load factor
H_2	—	—	—	—	—	—	—	—
CH_4	—	—	Y	Y	Y	—	—	—
C_2H_2	—	—	—	—	—	—	—	—
C_2H_4	—	Y	Y	—	Y	—	—	Y
C_2H_6	—	Y	Y	Y	—	—	—	Y
CO	—	—	Y	—	—	—	—	—
CO_2	—	Y	—	—	—	—	—	Y

In Table 3, the first column shows each indicator as the outcome variable, and the first row presents indicators as the cause variable. “Y” denotes a statistically significant Granger causal correlation between the two detection indicators. As shown in Table 3, when methane is taken as the outcome variable, it shows significant Granger causal correlations with acetylene, ethylene, and ethane, different from those obtained in Case 1. Such discrepancies in Granger causality test outcomes increase due to variations in transformer mode is, operating durations, and service environments. Therefore, when the gas prediction method is used, it is necessary to re-perform the Granger causality test on the dataset and incorporate the correlated gases into the neural network as an exogenous variable matrix to enhance the prediction accuracy.

Based on the Granger causal correlations between gases in Table 3, the correlated variables were incorporated into the NBEATSx model. The structural parameters and configurations of the neural network were consistent with those in Case 1, and the prediction results are shown in Fig. 10.

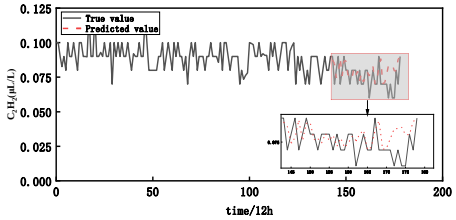


(a) Predicted results of hydrogen

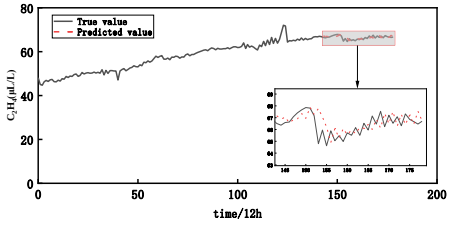


(b) Predicted results of methane

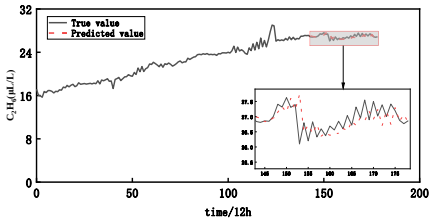
This paper has been accepted for publication in the AEE journal. This is the version, which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.24425/aee.2026.1524



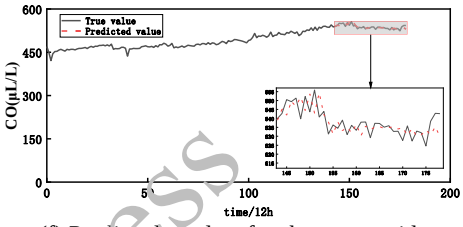
(c) Predicted results of acetylene



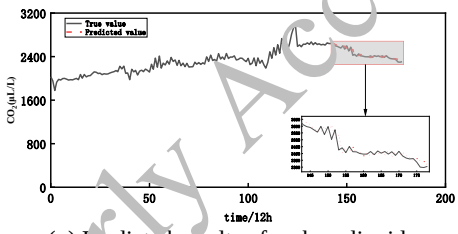
(d) Predicted results of ethylene



(e) Predicted results of ethane



(f) Predicted results of carbon monoxide



(g) Predicted results of carbon dioxide

Fig. 10. GC-NBEATSx forecasting of seven transformer DGA gases

Based on the Granger causal correlations between gases derived from Table 3, the correlated variables were incorporated into the NBEATSx model. The structural parameters and configurations of the neural network were kept consistent with those in Case 1, and the prediction results are presented in Fig. 10.

Table 4. Performance comparison of NBEATSx and GC-NBEATSx for transformer DGA gas prediction

	Gas	E_{MAPE}	E_{RMSE}	R^2
NBEATSx	H_2	1.83%	0.6599	0.762
	CH_4	1.08%	1.2065	0.402
	C_2H_2	10.38%	0.0091	0.571
	C_2H_4	0.77%	0.7115	0.618
	C_2H_6	1.14%	0.4335	0.494
	CO	1.23%	8.1126	0.643

This paper has been accepted for publication in the AEE journal. This is the version, which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.24425/aee.2026.1524

	CO_2	2.12%	58.9146	0.857
GC-NBEATSx	H_2	1.83%	0.6599	0.762
	CH_4	0.86%	0.9569	0.825
	C_2H_2	10.38%	0.0091	0.571
	C_2H_4	1.08%	0.9043	0.615
	C_2H_6	1.38%	0.4451	0.752
	CO	0.99%	6.5307	0.827
	CO_2	1.26%	39.1511	0.949

As shown in Fig. 10, the GC-NBEATSx model still has high prediction performance on gas data in Case 2. Meanwhile, as shown in Table 4, the prediction accuracy of most gas components is improved, except for hydrogen (H_2) and acetylene (C_2H_2). Therefore, incorporating the results of the Granger causality test as exogenous variables into the neural network can effectively enhance the overall performance and predictive accuracy of the gas prediction model. The prediction accuracy can be further improved by accumulating more transformer cases and adopting optimization algorithms to optimize the network parameters.

Using the same procedure as in Case 1, interval-based early warning was performed for the total hydrocarbon gases, and the results are shown in Fig. 11.

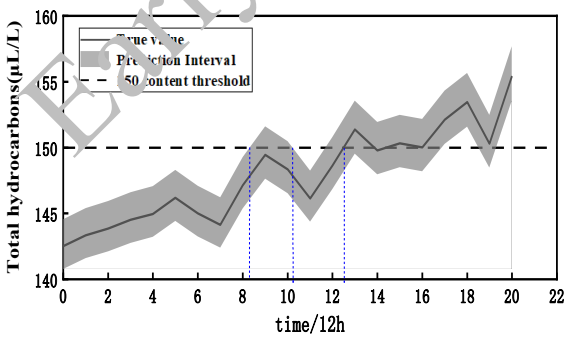


Fig. 11. Confidence interval prediction for total hydrocarbons in Case 2

As shown in Fig. 11, the prediction interval for the total hydrocarbon concentration effectively covers the actual values. For the transformer dataset in Case 2, the GC-NBEATSx model issued an early warning at the 9th sampling point, which lasted for about half a day. This is about two days earlier than an alarm only based on actual measurements.

This paper has been accepted for publication in the AEE journal. This is the version, which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.24425/aee.2026.1524

5. Conclusion

Based on the NBEATSx model, the Granger causality test method is used to analyze and dynamically select variables. A predictive and early-warning model, GC-NBEATSx, is put forward for dissolved gas content in transformer oil. The main findings are drawn as follows:

1) The Granger causality test method is used to analyze the temporal correlations between gas component contents and their relationship with transformer load rates, enabling dynamic variable selection of the predictive model. This approach can effectively uncover the correlations in transformer time-series data;
2) Case studies demonstrate that this model can achieve high predictive accuracy and generalization. This helps to develop on-site application-oriented early-warning strategies for transformer operational status, providing a valuable reference for transformer fault prevention and diagnosis.

References

[1] Ali M.S., Omar A., Jaafar A.S.A., Mohamed S.H., *Conventional methods of dissolved gas analysis using oil-immersed power transformer for fault diagnosis: A review*, Electric Power Systems Research, vol. 216, 109064 (2023), DOI: 10.1016/j.epsr.2022.109064.

[2] Wajid A., Rehman A.U., Iqbal S., Pushkarna M., Hussain S.M., Kotb H., Alharbi M., Zaitsev I., *Comparative performance study of dissolved gas analysis (DGA) methods for identification of faults in power transformer*, International Journal of Energy Research, vol. 2023, no. 1, 9960743 (2023), DOI: 10.1155/2023/9960743.

[3] Xiao Y., Zhu H., Chen X., *Concentration prediction of dissolved gas-in-oil of a power transformer with the multivariable grey model*, Automation of Electric Power Systems, vol. 30, no. 13, pp. 64–67 (2006).

[4] Deng Y., Zhang W., Jiang J., Wu R., *Prediction of the trends in the data of dissolved gas production in transformer oil based on time series analysis*, Smart Grid, vol. 4, no. 8, pp. 744–748 (2016), DOI: 10.14171/j.2095-5944.sg.2016.08.002.

[5] Liu H., Zhang J., Lian H., Zheng D., Lai Y., *Prediction of the gasses dissolved in transformer oil by sequential learning*, High Voltage Apparatus, vol. 55, no. 12, pp. 193–199 (2019), DOI: 10.13296/j.1001-1609.hva.2019.12.028.

[6] Wang Y., *Application of improved support vector machine model in fault diagnosis and prediction of power transformers*, Advanced Control for Applications, vol. 6, no. 4, e170 (2024), DOI: 10.1002/adc2.170.

[7] Kari T., He Z., Rouzi A., Zhang Z., Ma X., Du L., *Power Transformer Fault Diagnosis Using Random Forest and Optimized Kernel Extreme Learning Machine*, Intelligent Automation & Soft Computing, vol. 37, no. 1, 691 (2023), DOI: 10.32604/iasc.2023.037617.

[8] Zhou X., Tian T., He N., Ma Y., Liu W., Yan Z., Luo Y., Li X., Ni H., *Prediction Method of Dissolved Gas in Transformer Oil Based on Firefly Algorithm-Random Forest*, 2022 Asia Power and Electrical Technology Conference (APET), Shanghai, China, IEEE, pp. 51–55 (2022), DOI: 10.1109/APET56294.2022.10073321.

[9] Aseeri A.O., *Effective RNN-based forecasting methodology design for improving short-term power load forecasts: Application to large-scale power-grid time series*, Journal of Computational Science, vol. 68, 101984 (2023), DOI: 10.1016/j.jocs.2023.101984.

This paper has been accepted for publication in the AEE journal. This is the version, which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.24425/aee.2026.1524

[10] Hu C., Zhong Y., Lu Y., Luo X., Wang S., *A prediction model for time series of dissolved gas content in transformer oil based on LSTM*, Journal of Physics: Conference Series, vol. 1659, no. 1, 012030 (2020), DOI: 10.1088/1742-6596/1659/1/012030.

[11] Liu Y., Xu Z., Dong W., Li Z., Gao S., *Concentration prediction of dissolved gases in transformer oil based on empirical mode decomposition and long short-term memory neural networks*, Proceedings of the CSEE, vol. 39, no. 13, pp. 3998–4008 (2019), DOI: 10.13334/j.0258-8013.pcsee.182431.

[12] Xie L., Qiu W., Li Z., Liu Y., Jiang Q., Liu D., *Prediction model of dissolved gas in transformer oil based on variational modal decomposition and recurrent neural network with gated recurrent unit*, High Voltage Engineering, vol. 48, no. 2, pp. 653–660 (2022), DOI: 10.13336/j.1003-6520.hve.20201808.

[13] Cui Y., Hou H.J., Xu M.K., Li S.W., Sheng G.H., Jiang X.C., *A prediction method for dissolved gas in power transformer oil based on dualstage attention mechanism*, Proceedings of the CSEE, vol. 40, no. 1, pp. 338–347+400 (2020), DOI: 10.13334/j.0258-8013.pcsee.191060.

[14] Liu Z., Wang S., Tang B., *Prediction of dissolved gas content in transformer oil based on SSA-BiGRU-attention model*, High Voltage Engineering, vol. 48, no. 8, pp. 2972–2981 (2022), DOI: 10.13336/j.1003-6520.hve.20220099.

[15] Shojaie A., Fox E.B., *Granger causality: A review and recent advances*, Annual Review of Statistics and Its Application, vol. 9, no. 1, pp. 289–319 (2022), DOI: 10.1146/annurev-statistics-040120-010930.

[16] Zhang Z., Wu L., *Graph neural network-based bearing fault diagnosis using Granger causality test*, Expert Systems with Applications, vol. 242, 122827 (2024), DOI: 10.1016/j.eswa.2023.122827.

[17] Kostaridou E., Siatis N., Zafeiriou E., *Resource Price Interconnections and the Impact of Geopolitical Shocks Using Granger Causality: A Case Study of Ukraine–Russia Unrest*, Journal of Risk and Financial Management, vol. 17, no. 6, 240 (2024), DOI: 10.3390/jrfm17060240.

[18] Amalanathan A.J., Sarathi R., Zdanowski M., Vinu R., Nadolny Z., *Review on gassing tendency of different insulating fluids towards transformer applications*, Energies, vol. 16, no. 1, 488 (2023), DOI: 10.3390/en16010488.

[19] International Electrotechnical Commission, *IEC 60599: 2022: Mineral oil-filled electrical equipment in service–Guidance on the interpretation of dissolved and free gases analysis*, International Electrotechnical Commission, 60 (2022).

[20] Shaikh A.K., Nazir A., Khan I., Shah A.S., *Short term energy consumption forecasting using neural basis expansion analysis for interpretable time series*, Scientific Reports, vol. 12, no. 1, 22562 (2022), DOI: 10.1038/s41598-022-26499-y.

[21] Olivares K.G., Challu C., Marcjasz G., Weron R., Dubrawski A., *Neural basis expansion analysis with exogenous variables: Forecasting electricity prices with NBEATSx*, International Journal of Forecasting, vol. 39, no. 2, pp. 884–900 (2023), DOI: 10.1016/j.ijforecast.2022.03.001.

[22] Souto H.G., Moradi A., *Introducing NBEATSx to realized volatility forecasting*, Expert Systems with Applications, vol. 242, 122802 (2024), DOI: 10.1016/j.eswa.2023.122802.

[23] Zandieh A., Han I., Daliri M., Karbasi A., *Kdeformer: Accelerating transformers via kernel density estimation*, Proceedings of the 40th International Conference on Machine Learning, PMLR, vol. 202, pp. 40605–40623 (2023).

[24] Zhou X., Tian T., Liu N., Bai J., Luo Y., Li X., He N., Zhang P., Jun S., *Research on gas production law of free gas in oil-immersed power transformer under discharge fault of different severity*, Frontiers in Energy Research, vol. 10, 1056604 (2023), DOI: 10.3389/fenrg.2022.1056604.