electronics

Article

A Method for Predicting Transformer Oil-Dissolved Gas Concentration Based on Multi-Window Stepwise Decomposition with HP-SSA-VMD-LSTM

Tie Chen [1,2], Shinan Guo [1,2,*], Zhifan Zhang [1,2], Yimin Yuan [1,2] and Jiaqi Gao [1,2]

1 Hubei Provincial Key Laboratory for Operation and Control of Cascade Hydropower Station, China Three Gorges University, Yichang 443002, China; chent@ctgu.edu.cn (T.C.)
2 School of Electrical Engineering and New Energy, China Three Gorges University, Yichang 443002, China
* Correspondence: 202208580121063@ctgu.edu.cn

Abstract: Predicting the concentration of dissolved gases in transformer oil is a critical activity for the early detection of potential faults. To address the prevalent issue of data leakage in current prediction methods, this paper proposes a prediction method that completely avoids data leakage. First, the Hodrick Prescott (HP) filter is used for stepwise decomposition to obtain the long-term trend and high-frequency periodic component. The high-frequency periodic component is further decomposed using singular spectrum analysis (SSA) to extract periodic features. Dispersion entropy (DE) and fuzzy entropy (FE) are utilized alongside the HP and SSA methods to determine the optimal decomposition windows during the process, enhancing the ability of the model to acquire time series features. Then, variational mode decomposition (VMD) is applied to remove noise from the high-frequency component. Finally, the long short-term memory network (LSTM) is employed to predict each decomposed component, and the network parameters undergo optimization through the sparrow search optimization algorithm (SSOA). The two case studies in this work verify that the proposed model excels over other prediction means, providing strong support for subsequent fault prediction.

Keywords: dissolved gas in transformer oil; data leakage; multi-window stepwise decomposition; variational mode decomposition; long short-term memory

Citation: Chen, T.; Guo, S.; Zhang, Z.; Yuan, Y.; Gao, J. A Method for Predicting Transformer Oil-Dissolved Gas Concentration Based on Multi-Window Stepwise Decomposition with HP-SSA-VMD-LSTM. *Electronics* **2024**, *13*, 2881. https://doi.org/10.3390/electronics13142881

Academic Editor: Yide Wang

Received: 3 June 2024
Revised: 19 July 2024
Accepted: 20 July 2024
Published: 22 July 2024

1. Introduction

Oil-immersed power transformers play a critical role in the power system, and their uninterrupted operation holds significant economic and safety implications for both electric utilities and industrial power uses. Dissolved gas analysis (DGA) is a widely used technique to evaluate the health of oil-filled electrical equipment [1,2]. When a transformer experiences discharge faults or operates under overheating conditions, the oil and solid insulation materials will decompose, producing various characteristic gases that dissolve in the insulating oil [3]. As the equipment operates for longer periods of time, the content of the dissolved gases gradually rises, leading to a decline in the insulation performance and posing significant risks of serious failures. Through DGA, it is possible to assess the state of the transformer and diagnose faults [4,5].

Online monitoring devices for dissolved gases in transformer oil can automatically measure these gases in real time. These devices are now extensively used in substations, providing support for the condition maintenance of transformers [6,7]. By analyzing historical data collected by these devices, the precise estimation of dissolved gas concentrations in transformer oil can help service personnel anticipate changes in transformer conditions and detect potential faults early [8–10]. This, in turn, contributes to an enhancement in the reliability of power operations.

Owing to the synergistic effects of operational conditions, environmental influences, and intricate electromagnetic environments, the time series data of dissolved gases in transformer oil display pronounced nonlinearity and instability [11,12]. Traditional statistical methods [13,14] have limited capability in fitting nonlinear time series and face challenges in uncovering the underlying features of dissolved gas time series data in transformer oil. The combination of sequence decomposition techniques and deep learning models is a current research focus. Researchers employ techniques such as variational mode decomposition (VMD), empirical mode decomposition (EMD), and ensemble empirical mode decomposition (EEMD) [15–18] to disaggregate the original series into a set of stable, highly periodic modal components before prediction, thereby enhancing predictive accuracy. However, these studies generally use the overall decomposition sampling (ODS) technique, which first decomposes the time series as a whole and then splits it into training and prediction sets. This method introduces future 'unknown' data during decomposition, resulting in data leakage [19], which does not align with actual engineering scenarios. In contrast, the stepwise decomposition sampling (SDS) technique avoids this issue [20], making it more suitable for practical applications.

EMD suffers from severe mode mixing during decomposition, and the number of modes is uncontrollable, making it unsuitable for SDS [21,22]. In EEMD, the modal components change significantly as the time series boundary values are updated, resulting in difficulty in controlling time series patterns during SDS [23]. Under current circumstances, there is a lack of efficient decomposition methods specifically designed for SDS.

Transformers operating under high loads for extended periods experience continuous aging of oil–paper insulation materials, which leads to a stable long-term trend in the concentration of dissolved gases. Influenced by the ambient temperature, humidity, and load levels of the transformer environment, the concentration of dissolved gases in the oil exhibits periodic features over certain time scales. Additionally, sudden grid failures, intrusion of external pollutants, or abrupt changes in transformer operation can cause instantaneous strong fluctuations in the concentration of dissolved gases. Drawing from traditional decomposition prediction approaches, in SDS, the data on dissolved gas content in transformer oil is decomposed into long-term trends, periodic components, and fluctuating components. This reduces the complexity of the time series of gas concentrations in the oil.

To extract long-term trends in time series data, this paper introduces the Hodrick Prescott (HP) filter [24], which is a data analysis method commonly used in macroeconomics. Compared to Kalman filters and wavelet smooth, the HP filter does not require prior information and is simple to use. Reducing the volatility of time series data enhances the long-term trend, allowing for better analysis of its changes and decreasing the impact of data noise [25–27]. Therefore, it has received increasing attention in the field of data prediction [28]. Singular spectrum analysis (SSA) can effectively extract periodic oscillations in time series [29,30] and does not require the selection of prior basis functions, making it suitable for SDS sampling.

Based on this, this paper proposes the HSV (HP-SSA-VMD) sequence decomposition framework under SDS. First, the HP is used to obtain the long-term trend from the gas concentration time series. Then, SSA is employed to extract the periodic features and high-frequency components from the detrended data. Finally, VMD is applied to denoise the high-frequency component [31]. Considering that the decomposition window affects subsequent prediction accuracy, the dispersion entropy (DE) [32] is utilized to optimize the HP decomposition window for the extraction of the long-term trend feature sequence most similar to the original series trend [33]. The fuzzy entropy (FE) [34] is used to improve the SSA decomposition window to obtain the most significant periodic component.

Finally, each component is predicted separately using a long short-term memory (LSTM) network [35]. These predictions are then aggregated to achieve the prediction of dissolved gases in transformer oil.

Experimental results demonstrate that while avoiding the risk of data leakage, the proposed method realizes higher accuracy compared to traditional methods and possesses a greater practical engineering value.

2. Fundamental Principle

2.1. Stepwise Decomposition Sampling (SDS)

For the specified time series $TS = \{y_t | t = 1, \ldots, L + K\}$, select y_1 to y_L as the training set, with the remaining part as the test set. The specific SDS steps are as follows:

Step 1: Decompose the time series $y_{1:T}$ into subsequences $y^{i1}_{1:T}$. Here, T represents the decomposition length, and $i1$ means the i-th subseries of the first decomposition execution.

Step 2: Decompose the time series $y_{2:T+1}$ into subsequences $y^{i2}_{2:T+1}$. The input and output of the first sample of the i-th subsequence come from $y^{i1}_{1:T}$ and $y^{i2}_{2:T+1}$, as shown in Figure 1.

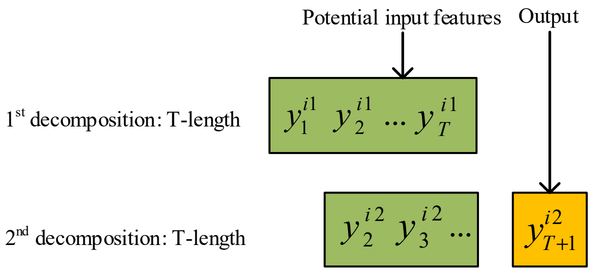

Figure 1. Input and output samples.

Step 3: Decompose the next T-length time series step-by-step until the training dataset is obtained.

Step 4: Build a prediction model for each subsequence.

Based on the above decomposition principle, this method does not introduce 'unknown' data and is closer to practical application scenarios.

2.2. HSV Sequence Decomposition Framework

First, the raw series is stepwise decomposed using the HP filter. The variance of the DE values of the trend components obtained under different decomposition windows is calculated. The window with the smallest DE is selected as the decomposition window T_1. Then, the residual component after HP decomposition is stepwise decomposed using SSA. The FE values of the periodic component under different decomposition windows are computed. The window with the smallest FE is selected as the decomposition window T_2. Finally, VMD is applied to decompose the high-frequency component containing noise, and the multiscale permutation entropy (MPE) [36] of each mode component is calculated. Effective mode components are determined based on the MPE threshold values. A genetic algorithm (GA) is adopted to optimize the entropy values to identify the optimal decomposition window.

2.2.1. Hodrick Prescott Filter (HP)

Hodrick Prescott filter decomposition is designed to smooth the data by minimizing the variance of fluctuations, using a symmetric data moving average method as an approximation of a high-pass filter [25]. DE is a dynamically changing metric that focuses on local features in signals with significant fluctuations and nonlinear characteristics, measuring the similarity between the long-term trend component and the original sequence.

The HP filter can decompose the time series $\boldsymbol{X} = \{x_1, x_2, \ldots, x_T\}$ into a long-term trend series $\boldsymbol{G} = \{g_1, g_2, \ldots, g_T\}$ and a high-frequency periodic series $\boldsymbol{S} = \{s_1, s_2, \ldots, s_T\}$, in which, $x_i = g_i + s_i,\ i = 1, 2, \ldots, T$. The long-term trend series $\{g_i\}$ is the solution that minimizes the loss function:

$$\min_{g_i}\left\{\sum_{i=1}^{T}(x_i - g_i)^2 + \theta\sum_{i=2}^{T-1}[(g_{i+1} - g_i) - (g_i - g_{i-1})]^2\right\} \tag{1}$$

In the equation, θ is a positive number, adjusted as a penalty factor to control the smoothness by altering the weighting, known as the smoothing parameter. A larger value results in a smoother long-term trend series $\{g_i\}$. T represents the number of samples in the time series $\{x_i\}$.

Calculate the first derivatives of $g_1,\ g_2,\ \ldots,\ g_T$ with respect to Equation (1) to further obtain the long-term trend $\boldsymbol{G}$:

$$\boldsymbol{G} = [\boldsymbol{I} + \theta\boldsymbol{F}]^{-1}\boldsymbol{X} \tag{2}$$

$$\boldsymbol{F} = \begin{bmatrix} 1 & -2 & 1 & \cdots & 0 & 0 \\ -2 & 4+1 & -2-2 & \cdots & 0 & 0 \\ 1 & -2-2 & 1+4+1 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 1+4 & -2 \\ 0 & 0 & 0 & \cdots & -2 & 1 \end{bmatrix} \tag{3}$$

In the equation, the sum of each column of the $\boldsymbol{F}$ matrix is 0, and $\boldsymbol{I}$ is the identity matrix.

Normalize the values of the long-term trend component under window T. Map the sequence $\{g_i\}$ to $\{y_i\}$ using the cumulative normal distribution function, where $y = \{y_1, y_2, \ldots, y_T\}$, so that the values of $\{y_i\}$ lie between [0, 1].

Divide the sequence $\{y_i\}$ into c classes:

$$z_i^c = round(c.y_i + 0.5) \tag{4}$$

In the equation, c is the number of classes in the sequence z^c; z_i^c is the i-th data point in the sequence z^c.

Construct a sequence $z_j^{m,c}$ with delay d and dimension m:

$$z_j^{m,c} = \left\{z_j^c, z_{j+d}^c, \ldots, z_{j+(m-1)d}^c\right\}, j = 1, \ldots, N - (m-1)d \tag{5}$$

Map the sequence $z_j^{m,c}$ to the discrete pattern $\pi_{v_0 \sim v_{m-1}}$:

$$\begin{cases} z_i^c = v_0 \\ z_{i+d}^c = v_1 \\ \quad\vdots \\ z_{i+(m-1)d}^c = v_{m-1} \end{cases} \tag{6}$$

In the equation, c^m is the number of discrete patterns assigned to each sequence $z_j^{m,c}$.

Calculate the relative frequency of each discrete mode of c^m:

$$p\left(\pi_{v_0 \sim v_{m-1}}\right) = \frac{Number\left\{t\,\middle|\,t \leq T - (m-1)d, z_j^{m,c}\left(\pi_{v_0 \sim v_{m-1}}\right)\right\}}{T - (m-1)d} \tag{7}$$

The DE of the sequence $\{g_i\}$ is:

$$DE(x, m, c, d) = -\sum p\left(\pi_{v_0 v_1 \ldots v_{m-1}}\right)\ln p\left(\pi_{v_0 v_1 \ldots v_{m-1}}\right) \tag{8}$$

2.2.2. Singular Spectrum Analysis (SSA)

As a method for handling nonlinear time series, singular spectrum analysis decomposes and reconstructs the trajectory matrix of the target time series, breaking the original series down into multiple independent and interpretable parts, such as slow-changing trends, oscillatory components, and structureless noise [37]. FE is an improvement over sample entropy, reflecting the probability of generating new patterns in a time series, which can be used to measure the complexity of each component.

Transform the high-frequency periodic component into a trajectory matrix $\boldsymbol{W}$.

$$\boldsymbol{W}_{L\times K} = (w_{ij})_{i,j=1}^{L,K} = \begin{pmatrix} s_1 & s_2 & \cdots & s_{n-L+1} \\ s_2 & s_3 & \cdots & s_{n-L+2} \\ \vdots & \vdots & \ddots & \vdots \\ s_L & s_{L+1} & \cdots & s_n \end{pmatrix} \tag{9}$$

In the equation, $K = n - L + 1$, and L is the embedding window length.

Suppose d represents the number of non-zero eigenvalues, then the singular value decomposition of the trajectory matrix $\boldsymbol{W}$ can be expressed as:

$$\boldsymbol{W} = \boldsymbol{W}_1 + \boldsymbol{W}_2 + \cdots \boldsymbol{W}_d \tag{10}$$

In the equation, $\boldsymbol{W}_i = \sqrt{\lambda_i}\boldsymbol{U}_i\boldsymbol{V}_i^T$, $\boldsymbol{U}_i$, and $\boldsymbol{V}_i$ correspond to the left and right singular vectors of the i-th singular value of $\boldsymbol{W}$, respectively. The contribution rate η of each matrix $\boldsymbol{W_i}$ to $\boldsymbol{W}$ can be expressed as:

$$\eta = \lambda_i / \sum \lambda \tag{11}$$

Suppose $\boldsymbol{W}$ is an $L \times K$ matrix, let K* = max (L, K) and L* = min (L, K). If L < K, then $w_{ij}{}^* = w_{ij}$; otherwise, $w_{ij}{}^* = w_{ji}$. The formula for computing the reconstructed sequence $Z_c = (z_{c1}, z_{c2}, \ldots, z_{cn})$ is:

$$z_{cl} = \begin{cases} \dfrac{1}{l+1}\sum\limits_{m=1}^{l+1} w^*_{m,\, l-m+2}, & 1 \le l \le L^* \\ \dfrac{1}{L^*}\sum\limits_{m=1}^{L^*} w^*_{m,\, l-m+2}, & L^* \le l \le K^* \\ \dfrac{1}{n-l}\sum\limits_{m=l-K^*+2}^{l-K^*+1} w^*_{m,\, l-m+2}, & K^* \le l \le n \end{cases} \tag{12}$$

Set a contribution rate threshold η_e, and construct sub-sequences according to the size of the eigenvalues. When $\sum\lambda_i/\sum\lambda \ge \eta_e$, the sequence formed is the periodic component of the original sequence after detrending, while the remaining sub-sequences are reconstructed as high-frequency components containing noise.

Perform phase space reconstruction on the periodic sequence $\{s_i\}$ under window:

$$X_i^m = \{s(i), s(i+1), \cdots, s(i+m-1)\} - s_0(i) \tag{13}$$

$$s_0(i) = \frac{1}{m}\sum_{j=0}^{m-1} s(i+j) \tag{14}$$

In the equation, m is the embedding dimension, $s_0(i)$ is the mean value, and $i = 1, 2, \ldots,$ $N - m + 1$.

Calculate the distance $d\left[X_i^m, X_j^m\right]$:

$$d_{ij}^m = d\left[X_i^m, X_j^m\right] = \max_{k\in(0,m-1)} \{|(s(i+k) - s_0(i)) - \ (s(j+k) - s_0(j))|\} \tag{15}$$

Using the fuzzy function $\mu(d_{ij}^m, n, r)$ to define the similarity D_{ij}^m between X_i^m and X_j^m:

$$D_{ij}^m = \mu\left(d_{ij}^m, n, r\right) = \exp\left(-\frac{\left(d_{ij}^m\right)^n}{r}\right) \tag{16}$$

In the equation, r is the similarity tolerance, and n is the gradient of the fuzzy function.

Define the function:

$$\psi^m(r) = \frac{1}{N-m}\sum_{i=1}^{N-m}\left(\frac{1}{N-m-1}\sum_{j=1,i\neq j}^{N-m} D_{ij}^m\right) \tag{17}$$

The FE of the sequence $\{s_i\}$ is:

$$FuzzyEn(m,r,N) = ln\psi^m(r) - ln\psi^{m+1}(r) \tag{18}$$

2.2.3. Variational Mode Decomposition (VMD)

Variational mode decomposition can decompose highly complex and non-stationary signals into multiple stationary IMF components with different central frequencies [38].

Construct a constrained variational optimization problem:

$$\begin{cases} \min\limits_{\substack{\{u_k\}\\ \{w_k\}}}\left\{\sum\limits_k \left\| \partial_t\left[\left(\delta(t) + \frac{j}{\pi t}\right)\times u_k(t)\right]e^{-\mathrm{j}w_k t}\right\|_2^2\right\} \\ \text{s.t} \quad \sum\limits_k u_k(t) = f(t) \end{cases} \tag{19}$$

In the equation, $\{u_k\}$ is the set of K IMF components; $\{w_k\}$ is the set of K frequency centers; $\delta(t)$ is the Dirac delta function; $f(t)$ is the initial signal.

Utilizing the Lagrange multiplier λ and the quadratic penalty factor α, transform Equation (19) into an unconstrained variational problem:

$$L(\{u_k\},\{w_k\},\lambda) = \alpha\sum_k\left\|\partial_t\left[\left(\delta(t)+\frac{j}{\pi t}\right)\times u_k(t)\right]e^{-jw_k t}\right\|_2^2 + \left\|f(t) - \sum_k u_k(t)\right\|_2^2 + \left\langle \lambda(t), f(t) - \sum_k u_k(t)\right\rangle \tag{20}$$

For the multiple IMF components obtained after VMD, MPE is applied as a criterion for distinguishing noise from signal. MPE represents an optimization technique based on permutation entropy (PE), whose core idea involves computing PE over multiple time scales, offering better stability and resistance to interference over PE alone. The calculation process can be found in reference [39].

2.3. Long Short-Term Memory Network (LSTM)

The long short-term memory network structure, as shown in Figure 2, addresses issues such as long-term dependencies and vanishing gradients in recurrent neural networks.

The input to the neuron at time t is the output of the neuron at the previous time step h_{t-1}, the state of the memory cell at the previous time step c_{t-1}, and the neuron input x_t. After processing by the forget gate, input gate, and output gate of the LSTM neuron, the current time step memory cell state c_t and neuron output h_t are outputted. The formulas for calculating the three gates are as follows:

Forget gate:

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right) \tag{21}$$

Input gate:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{22}$$

$$\widetilde{c}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \tag{23}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \widetilde{c}_t \tag{24}$$

Output gate:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{25}$$

$$h_t = o_t \odot \tanh(c_t) \tag{26}$$

The learning rate, number of iterations, and number of neurons in the hidden layer of LSTM will affect the prediction accuracy of the model. In this study, the sparrow search optimization algorithm (SSOA) [40] is employed to optimize these parameters, where the standard deviation between the predicted values of LSTM and the actual values is taken as the fitness value of the sparrow individuals.

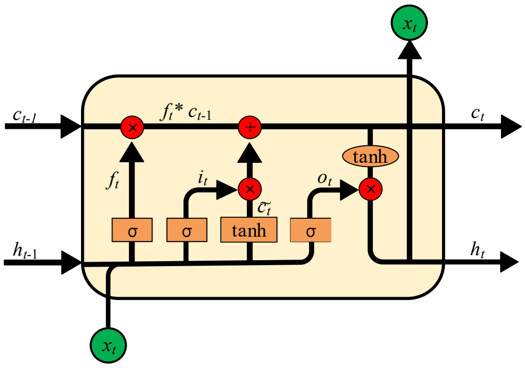

Figure 2. LSTM neurons.

3. Dissolved Gas Concentration Prediction Model in Oil

The proposed method for predicting dissolved gas concentrations in transformer oil based on multi-window stepwise HSV decomposition and LSTM networks is exhibited in Figure 3, where the primary steps are demonstrated as follows:

Step 1: Divide the dissolved gas raw time series in transformer oil into a training set and a testing set.

Step 2: Using the HSV decomposition framework, sequentially perform HP filter decomposition, SSA decomposition, and VMD decomposition to obtain the complete long-term trend component, periodic features component, and low-frequency mode component as training samples.

Step 3: Build LSTM network prediction models separately for the long-term trend component, periodic features component, and low-frequency mode component. Utilize the SSOA algorithm to optimize the hyperparameters of the LSTM network and select the optimal parameter combinations for predicting each component.

Step 4: Sum the predicted values of each component to acquire the forecasted dissolved gas concentration in the original transformer oil.

In this study, the root mean square error (RMSE) and mean absolute error (MAE) are used as evaluation metrics for the model, as displayed in Table 1. Here, y_i represents the actual value, $\hat{y}_{\mathrm{i}}$ represents the predicted value, and n is the number of samples.

Table 1. Error index.

Metrics	Definition	Equations
RMSE	Root mean square error	$RMSE = \sqrt{\frac{1}{n}\sum\limits_{i=1}^{n}(y_{\mathrm{i}} - \hat{y}_{\mathrm{i}})^2}$
MAE	Mean absolute error	$MAE = \frac{1}{n}\sum\limits_{i=1}^{n}\lvert y_{\mathrm{i}} - \hat{y}_{\mathrm{i}}\rvert$

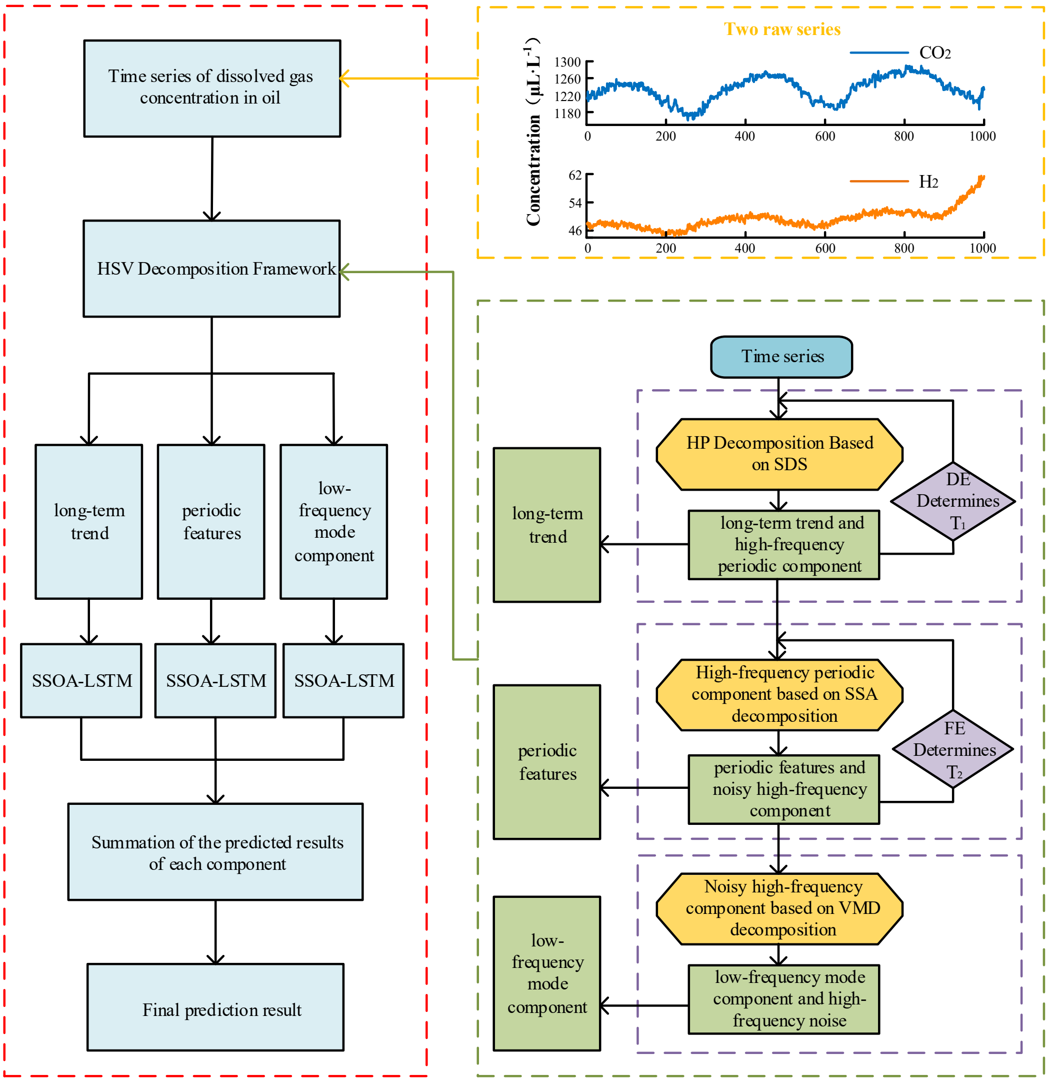

Figure 3. The flow chart of the proposed prediction approach.

4. Experimental Results and Analysis

4.1. Experiment I

The dissolved gas concentration data were collected from the online monitoring device of the 500 kV transformer of Unit 1 at a power station in China from 11 April 2011, to 5 January 2014, for testing purposes. It includes data for five types of gases: CO_2, CH_4, C_2H_4, CO, and H_2. The sampling interval is 1 day, with a total of 1000 samples collected. The data for the first 800 days are used as the training set, while the data for the remaining 200 days are the test set. Due to space limitations, this section will analyze the CO_2 sequence of the transformer as an example, as shown in Figure 4.

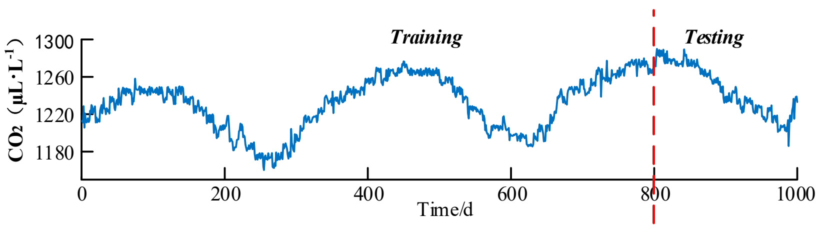

Figure 4. DGA gas sequence of the transformer.

4.1.1. HP Decomposition

First, the CO_2 gas sequence was subjected to HP stepwise decomposition to obtain the trend feature curve and high-frequency periodic component. DE was used to optimize the HP decomposition window, extracting the long-term trend. The initial population size for the GA optimization algorithm was set to 80, with 150 iterations. Through multiple experiments, the minimum fitness value obtained was 0.6948, corresponding to the optimal decomposition window of 102. The GA iteration curve is pictured in Figure 5. After the CO_2 sequence undergoes HP stepwise decomposition based on the optimal value (smoothing parameter θ = 14,600), the long-term trend and high-frequency periodic component are acquired, as shown in Figure 6.

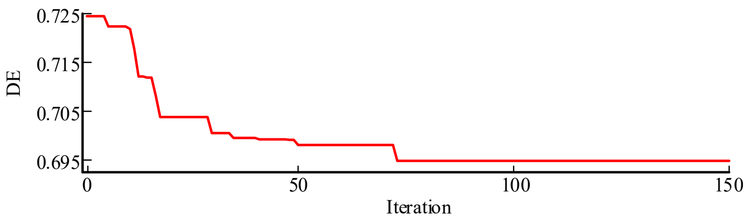

Figure 5. GA iteration curve.

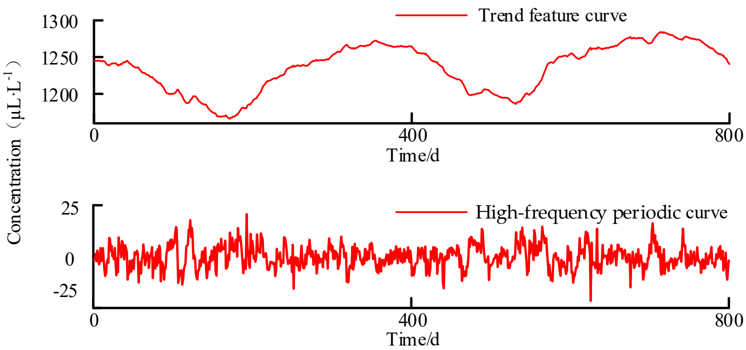

Figure 6. HP decomposition results.

4.1.2. SSA Decomposition

The high-frequency periodic components obtained from the HP decomposition are further decomposed using the SSA stepwise method to derive the periodic features curve and the high-frequency component that contains noise. FE is utilized to optimize the SSA decomposition window, extracting the periodic features. Similarly, the initial population size for the GA optimization algorithm was set to 80, with 150 iterations. Through multiple experiments, the minimum fitness value gained for FE was 0.4763, with an optimal decomposition window of 88. The GA iteration curve is plotted in Figure 7. After SSA stepwise decomposition of the high-frequency periodic component based on the optimal values (embedding window length L = 44, contribution rate threshold η_e = 0.65), the periodic features and high-frequency component containing noise were captured, as displayed in Figure 8.

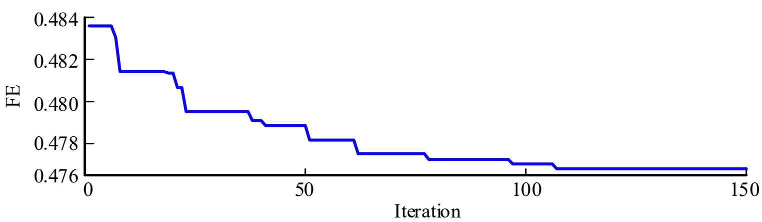

Figure 7. GA iteration curve.

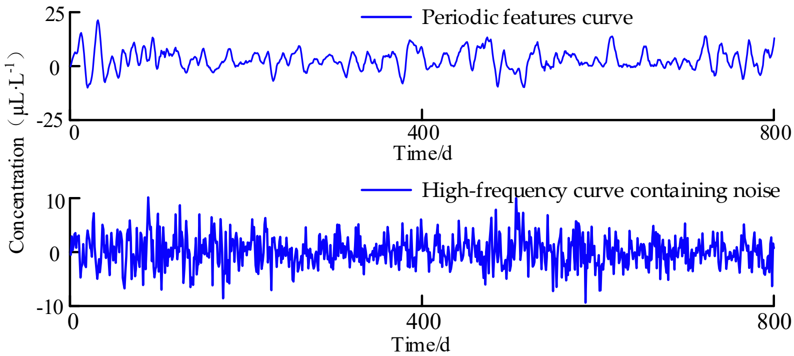

Figure 8. SSA decomposition results.

4.1.3. VMD Decomposition

As viewed in Figure 8, the high-frequency component containing noise exhibits significant nonlinearity and instability. Using the optimal decomposition parameters (K,α) = (7,1793), the high-frequency component, including noise, was decomposed using VMD. The MPE values of each IMF component were calculated, with MPE values closer to 1, indicating greater random fluctuation and higher irregularity in the time series. An MPE threshold of 0.6 is typically used, where IMF components with MPE values greater than 0.6 are considered noise components, and those with values less than 0.6 are regarded as low-frequency signal components [36]. The VMD decomposition results and the MPE values for each IMF component are shown in Figure 9 and Table 2, respectively. The figure and table indicate that the low-frequency components are primarily concentrated in the first two modes. The MPE values of IMF1 and IMF2 are less than 0.6, so the low-frequency signals IMF1 and IMF2 are reconstructed as low-frequency mode components for prediction. IMF3 to IMF7 are treated as noise and filtered out.

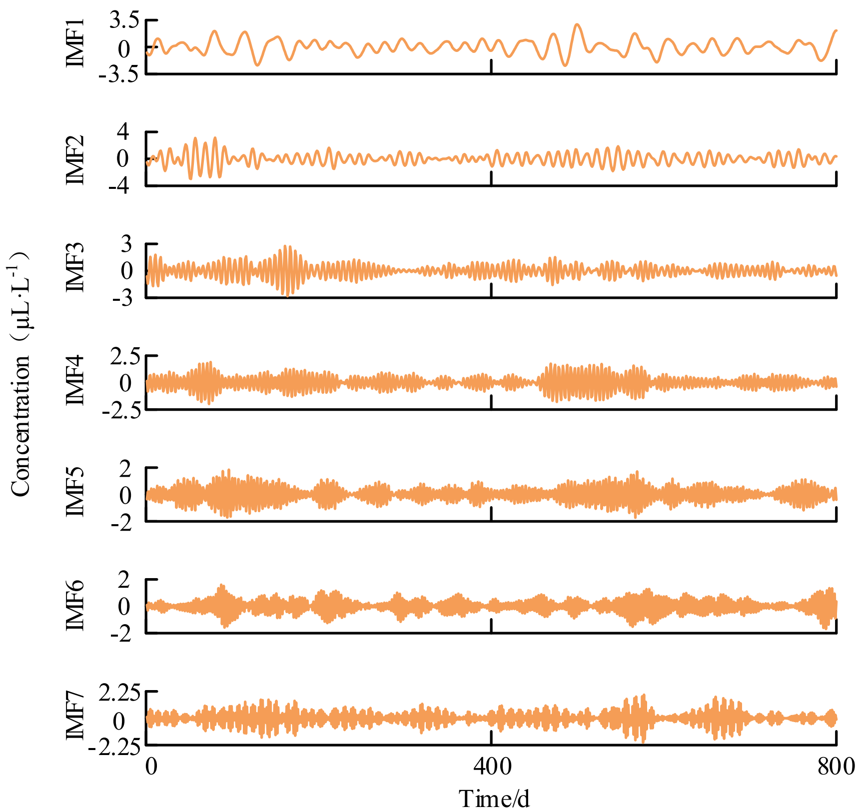

Figure 9. VMD decomposition results.

Table 2. The MPE mean value of each IMF component.

Modal components	IMF1	IMF2	IMF3	IMF4	IMF5	IMF6	IMF7
MPE	0.3758	0.4805	0.6755	0.6702	0.6817	0.6672	0.6901

4.1.4. LSTM Prediction

Then, the LSTM prediction model is established, and the SSOA is applied to optimize the network hyperparameters. After multiple experiments, the SSOA population size was set to 20, with the proportion of discoverers at 10% and the remaining individuals as joiners, with a warning value set to 0.8. The activation function of the LSTM model was tanh, the loss function was a mean squared error, and the optimizer was Adam. The batch size was set to 32. The network model was built using the TensorFlow framework, with the parameter settings displayed in Table 3.

Table 3. LSTM model parameters setting.

Model Parameters	**Long-Term Trend**	**Periodic Feature**	**Low-Frequency Mode Component**
learning rate	0.01	0.01	0.02
iterations	143	35	124
neurons	50	60	52

The long-term trend, periodic features, and low-frequency mode component obtained from the decomposition of the CO_2 sequence were input into the LSTM for prediction. The prediction results were then summed up to acquire the final prediction value. By integrating the single-step prediction results with the historical monitoring sequence, a new single-step prediction model input sequence can be constructed, enabling multi-step rolling prediction.

4.1.5. Analysis of Results

To validate the effectiveness of the proposed model in this study, a comprehensive comparative analysis is conducted among LSTM, SDS-EMD-LSTM, SDS-EEMD-LSTM, SDS-HSV-LSTM, and the SDS-DE-FE-HSV-LSTM model proposed in this work.

Taking the one-step prediction of CO_2 as an example, all prediction performance and error indicators of each model are exhibited in Figure 10; the prediction results for the remaining four gases are drawn in Appendix A. In Figure 10, compared with other models, the prediction curve of the proposed one is closest to the actual data, and the prediction error indicators are smaller, indicating that the model proposed in this study has better predictive performance.

Table 4 shows error indicators for the test sets of the five gases. It can be observed that in the SDS-EMD-LSTM and SDS-EEMD-LSTM models, compared to LSTM, the RMSE indicators increased by 7.89% and 11.68%, respectively, and the MAE indicators by 9.82% and 15.14%, respectively, in one-step prediction. This suggests that under SDS technology, models based on EMD and EEMD can improve prediction accuracy, with EEMD outperforming EMD, albeit with limited overall improvement. Compared to EEMD, the HSV-based decomposition of the RMSE and MAE increased by 14.96% and 13.94%, respectively, in one-step prediction, suggesting that the HSV decomposition framework proposed in this paper is more suitable for SDS than traditional EMD and EEMD methods. The prediction model based on entropy-optimized decomposition windows performed better than HSV, with an increase of 4.34% in RMSE and 7.03% in MAE. This demonstrates that the time window can affect the extraction of long-term trends and periodic components, and using DE and FE to determine the optimal time window can improve the decomposition of long-term trends and periodic components, thereby enhancing the accuracy of feature extraction from the data time series. Compared to the LSTM model, the proposed model in this paper showed an average reduction of 30.98% in RMSE and 36.11% in MAE in one-step prediction. In multi-step prediction, the RMSE and MAE indicators decrease by 33.82% and 39.59%, respectively, implying a more significant improvement in accuracy in multi-step prediction, with a more pronounced advantage. In conclusion, the proposed model manifests potential for application in the short-term prediction of transformer oil gas sequences.

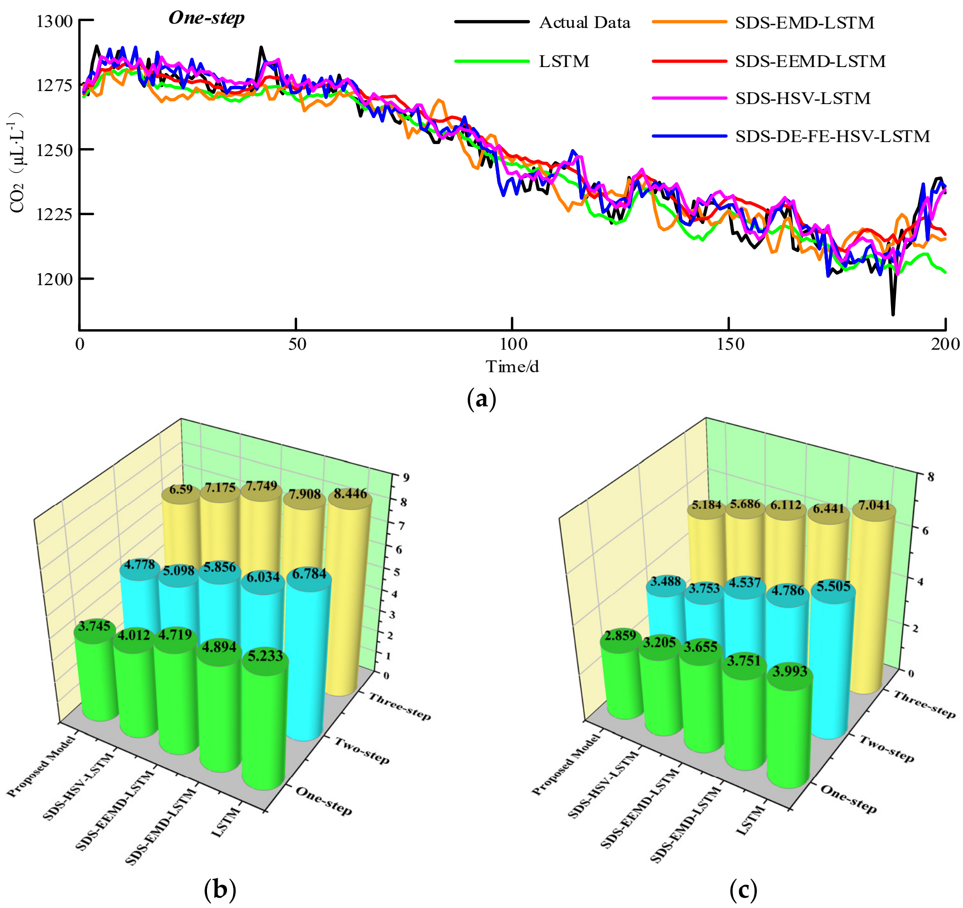

Figure 10. The prediction performance of CO_2 sequence. (**a**) Contrast of one-step forecasting consequences. (**b**) Contrast of RMSE error indicators. (**c**) Contrast of MAE error indicators.

Table 4. Error indicators of different prediction models.

Gases	Models	One-Step		Two-Step		THREE-STEP	
		RMSE	MAE	RMSE	MAE	RMSE	MAE
CO_2	LSTM	5.233	3.993	6.784	5.505	8.446	7.041
	SDS-EMD-LSTM	4.894	3.751	6.034	4.786	7.908	6.441
	SDS-EEMD-LSTM	4.719	3.655	5.856	4.537	7.749	6.112
	SDS-HSV-LSTM	4.012	3.205	5.098	3.753	7.175	5.686
	SDS-DE-FE-HSV-LSTM	3.745	2.859	4.778	3.488	6.590	5.184
CH_4	LSTM	0.418	0.365	0.422	0.368	0.429	0.375
	SDS-EMD-LSTM	0.371	0.310	0.382	0.313	0.391	0.335
	SDS-EEMD-LSTM	0.360	0.304	0.370	0.307	0.379	0.318
	SDS-HSV-LSTM	0.321	0.266	0.330	0.268	0.348	0.278
	SDS-DE-FE-HSV-LSTM	0.279	0.228	0.288	0.227	0.306	0.231

Table 4. *Cont.*

Gases	Models	One-Step		Two-Step		THREE-STEP	
		RMSE	MAE	RMSE	MAE	RMSE	MAE
C_2H_4	LSTM	0.247	0.206	0.319	0.268	0.387	0.328
	SDS-EMD-LSTM	0.232	0.189	0.295	0.243	0.336	0.270
	SDS-EEMD-LSTM	0.222	0.174	0.283	0.236	0.319	0.259
	SDS-HSV-LSTM	0.195	0.153	0.243	0.183	0.266	0.214
	SDS-DE-FE-HSV-LSTM	0.173	0.129	0.211	0.159	0.243	0.188
CO	LSTM	3.970	3.036	6.715	5.819	9.357	8.270
	SDS-EMD-LSTM	3.697	2.776	5.142	4.274	7.201	6.110
	SDS-EEMD-LSTM	3.489	2.503	4.813	3.943	6.801	5.801
	SDS-HSV-LSTM	2.949	2.147	4.489	3.553	6.127	5.148
	SDS-DE-FE-HSV-LSTM	2.681	1.830	4.185	3.212	5.991	4.892
H_2	LSTM	0.546	0.445	0.609	0.497	0.656	0.533
	SDS-EMD-LSTM	0.498	0.394	0.553	0.443	0.647	0.520
	SDS-EEMD-LSTM	0.478	0.373	0.537	0.426	0.613	0.497
	SDS-HSV-LSTM	0.410	0.318	0.488	0.387	0.582	0.463
	SDS-DE-FE-HSV-LSTM	0.378	0.278	0.441	0.349	0.509	0.385

4.2. Experiment II

Due to the specific working environment and intricate external conditions, the data from faulty transformers display pronounced nonlinearity and instability alongside frequent abrupt changes and abnormal trend variations. In this section, the dissolved gas concentration data were collected from the online monitoring device of the 500 kV transformer of Unit 2 at a power station in China from 27 June 2015 to 23 March 2018. The data include three types of gases: H_2, CH_4, and CO. Starting from 30 December 2017, the transformer experienced an incipient high-temperature fault, resulting in the rapid internal generation of H_2, CH_4, and CO. The analysis is conducted using the H_2 sequence of this transformer as an example, as shown in Figure 11.

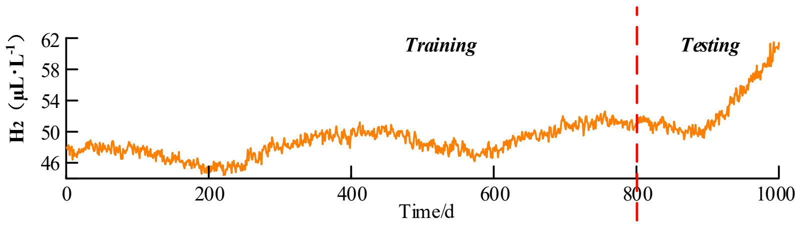

Figure 11. DGA gas sequence of the faulty transformer.

Analysis of Results

To confirm the effectiveness of the proposed model in this study, a comparative analysis was performed between the results obtained from the LSTM model and the proposed model. Taking H_2 prediction as an example, the prediction properties of the models are illustrated in Figure 12.

From Figure 12, it can be viewed that around day 100, the actual values curve shows an upward trend, which corresponds to the onset of an incipient high-temperature fault in the transformer. Although the LSTM model is capable of capturing these abrupt trends, it still faces problems with temporal offsets and has certain prediction accuracy deviations. Compared to the LSTM model, the approach in this study more effectively extracts sudden fluctuations in the sequence when addressing temporal offset issues, thereby understanding the patterns of gas changes in oil more precisely and making more accurate predictions of sequence anomalies during faults.

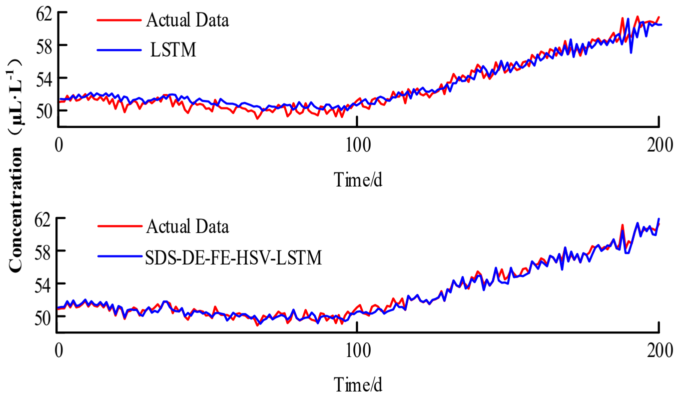

Figure 12. The prediction performance of H_2 sequence.

Table 5 displays the error indicators for the three gases. As seen in Table 5, compared to the LSTM model, the method proposed in this study achieves significant improvement in overall prediction performance. Through experiment II, it has been demonstrated that the proposed model can effectively identify potential trends in gas sequences before faults occur. This aids maintenance personnel in diagnosing transformer faults and assessing their condition.

Table 5. Error indicators of different prediction models.

Gases	Models	RMSE	MAE
H_2	LSTM	0.703	0.573
	SDS-DE-FE-HSV-LSTM	0.408	0.303
CH_4	LSTM	0.867	0.732
	SDS-DE-FE-HSV-LSTM	0.367	0.301
CO	LSTM	6.213	4.951
	SDS-DE-FE-HSV-LSTM	3.466	1.905

5. Conclusions

This research integrates multiple decomposition techniques and deep learning methods to propose a transformer oil-dissolved gas prediction model based on multi-window stepwise decomposition. By combining it with subsequent fault detection and state evaluation models, it can predict potential faults and abnormal conditions in transformers in advance. The following conclusions are drawn:

(1) In the prediction of dissolved gases in transformer oil, the SDS strategy effectively mitigates the data leakage issue inherent to the ODS technique. The proposed HSV sequence decomposition framework in this study proves to be more compatible with SDS than traditional modal decomposition methods.

(2) The selection of the decomposition window affects the extraction of long-term trends and periodic components. DE can improve the decomposition of long-term trends, while FE enhances the decomposition of periodic components. By utilizing DE and FE, optimal windows for extracting long-term trends and periodic features can be determined, thereby improving the accuracy of temporal feature extraction and reducing overall prediction difficulty.

(3) Compared to traditional models, the predictive advantage of the proposed model in multi-step forecasting is more significant than in one-step forecasting.

Author Contributions: Methodology, T.C. and S.G.; software, Y.Y. and J.G.; validation, S.G. and J.G.; writing—original draft, S.G.; writing—review & editing, Z.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the National Natural Science Foundation of China (51741907), The Opening Fund of Hubei Province Key Laboratory of Operation, and Control of Cascade Hydropower Station (2019KJX08).

Data Availability Statement: All the data supporting the reported results have been included in this paper.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

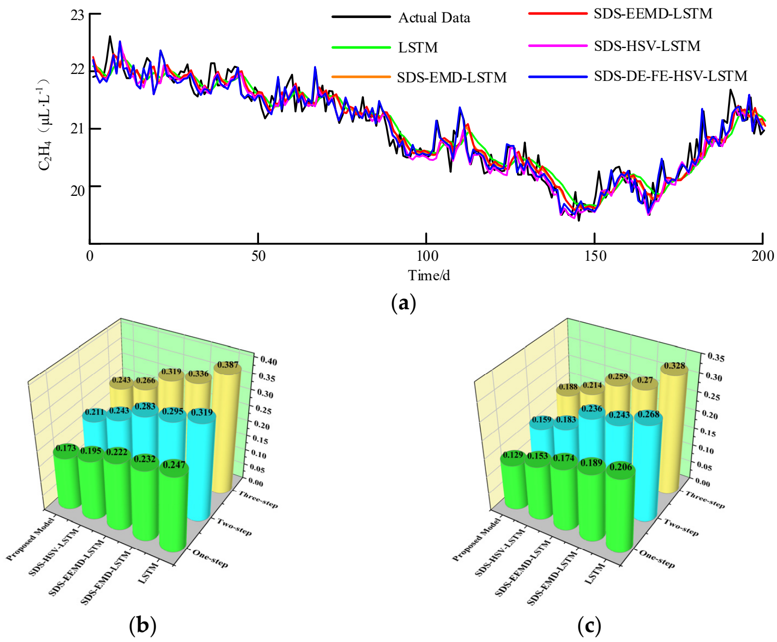

Figure A1. The prediction performance of C_2H_4 sequence. (**a**) Contrast of one-step forecasting consequences. (**b**) Contrast of RMSE error indicators. (**c**) Contrast of MAE error indicators.

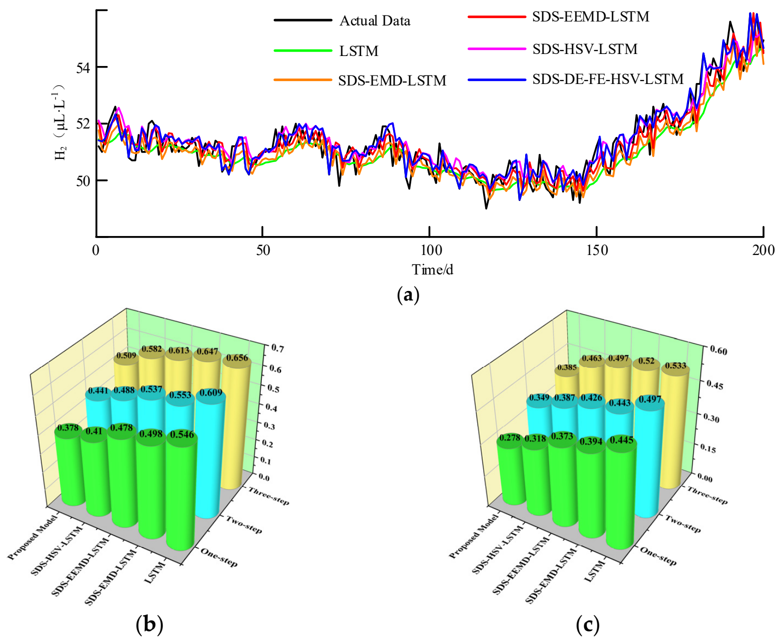

Figure A2. The prediction performance of H_2 sequence. (**a**) Contrast of one-step forecasting consequences. (**b**) Contrast of RMSE error indicators. (**c**) Contrast of MAE error indicators.

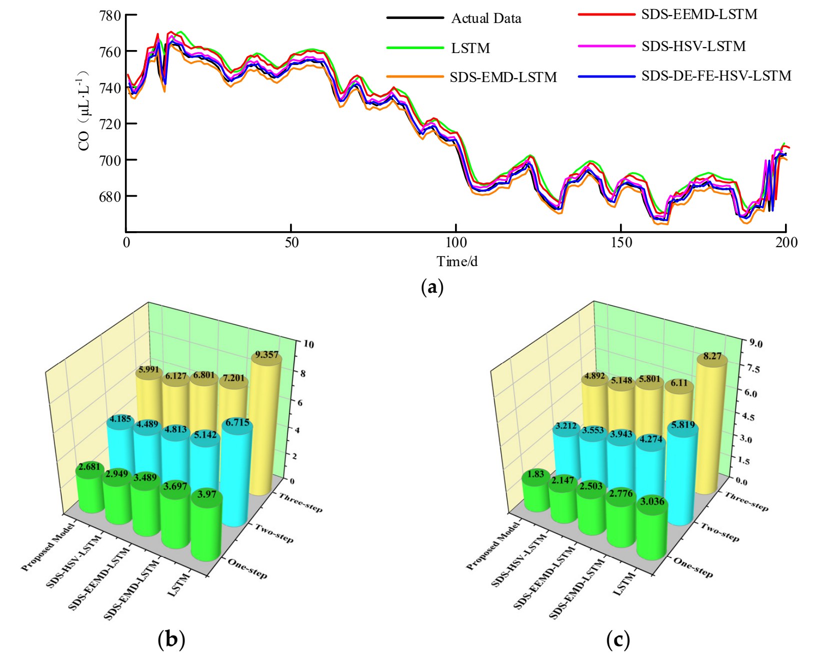

Figure A3. The prediction performance of CO sequence. (**a**) Contrast of one-step forecasting consequences. (**b**) Contrast of RMSE error indicators. (**c**) Contrast of MAE error indicators.

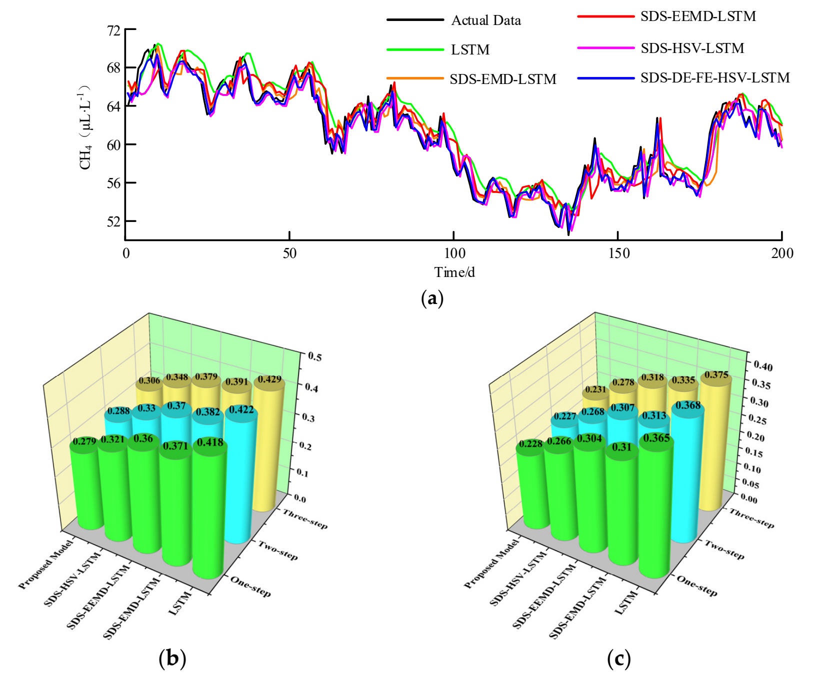

Figure A4. The prediction performance of CH_4 sequence. (**a**) Contrast of one-step forecasting consequences. (**b**) Contrast of RMSE error indicators. (**c**) Contrast of MAE error indicators.

References

1. A Guide to Transformer Oil Analysis. Available online: https://www.satcs.co.za/Transformer_Oil_Analysis.pdf (accessed on 18 December 2023).
2. Wang, Y.; Li, W.; Zhao, H.S.; Zhang, J.L.; Shen, Z.W. Transformer DGA fault diagnosis method based on DBN-SSAELM. *Power Syst. Prot. Control* **2023**, *51*, 32–42.
3. 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 Volt. Eng.* **2022**, *48*, 653–660.
4. Mirowski, P.; Lecun, Y. Statistical machine learning and dissolved gas analysis: A review. *IEEE Trans. Power Deliv.* **2012**, *27*, 1791–1799. [CrossRef]

5. Chen, T.; Leng, H.W.; Li, X.S.; Chen, Y.F. Transformer hierarchical fault diagnosis model based on dissolved gas analysis of insulating oil and class overlap features. *Electr. Power* **2022**, *55*, 22–32, 41.
6. Gao, Q.; Dai, J.C.; Yuan, F.; Pan, F.H.; Li, Z.L. An On-Line monitoring device for dissolved gas in transformer oil based on spectrum technology. In Proceedings of the Earth and Environmental Science, Seoul, Republic of Korea,, 26–28 January 2019.
7. Chen, Y.; Wang, Z.T.; Li, Z.; Zheng, H.Q.; Dai, J.M. Development of an online detection setup for dissolved gas in transformer insulating oil. *Appl. Sci.* **2021**, *11*, 12149. [CrossRef]
8. Zai, H.T.; Chen, W.G.; He, H.Y.; Lee, W.J.; Zhang, Z.Y.; Zhang, K.; Fang, J.; Luo, D.S. Prediction for dissolved gas in power transformer oil based on temporal convolutional and graph convolutional network. In Proceedings of the 2021 IEEE/IAS Industrial and Commercial Power System Asia, Chengdu, China, 18–21 July 2021.
9. Hu, C.; Zhong, Y.; Lu, Y.Q.; Luo, X.T.; Wang, S.R. A prediction model for time series of dissolved gas content in transformer oil based on LSTM. *J. Phys. Conf. Ser.* **2020**, *1659*, 012030. [CrossRef]
10. Lin, J.; Su, L.; Yan, Y.J.; Sheng, G.H.; Xie, D.; Jiang, X.C. Prediction method for power transformer running state based on LSTM-DBN network. *Energies* **2018**, *11*, 1880. [CrossRef]
11. Ding, C.; Ding, Q.C.; Feng, L.; Wang, Z.L. Prediction model of dissolved gas in transformer oil based on VMD-SMA-LSSVM. *IEEJ Trans. Electr. Electron. Eng.* **2022**, *17*, 1432–1440. [CrossRef]
12. Zhang, P.; Qi, B.; Zhang, R.Y.; Shao, M.Y.; Li, C.R. Dissolved gas prediction in transformer oil based on empirical wavelet transform and gradient boosting radial basis. *Power Syst. Technol.* **2021**, *45*, 3745–3754.
13. Liu, J.X.; Zhao, Z.J.; Zhong, Y.C.; Zhao, C.C.; Zhang, G.G. Prediction of the dissolved gas concentration in power transformer oil based on SARIMA model. In Proceedings of the 2021 the 2nd International Conference on Power Engineering, Nanning, China, 9–11 December 2021.
14. Chen, Y.; He, Z.; Shang, Z.; Li, C.; Li, L.; Xu, M. A novel combined model based on echo state network for multi-step ahead wind speed forecasting: A case study of NREL. *Energy Convers. Manag.* **2019**, *179*, 13–29. [CrossRef]
15. Shen, Y.F.; Zheng, W.; Yin, W.J.; Xu, A.G.; Zhu, H.Z. Feature extraction algorithm using a correlation coefficient combined with the VMD and its application to the GPS and GRACE. *IEEE Access* **2021**, *9*, 17507–17519. [CrossRef]
16. Teng, C.Y.; Ding, Y.C.; Zhang, Y.B.; Li, S.; Mo, Y.J. Ultra-short-term photovoltaic power prediction based on VMD-informer-BiLSTM mode. *High Volt. Eng.* **2023**, *49*, 2961–2971.
17. Zeng, B.; Guo, J.; Zhang, F.Q.; Zhu, W.Q.; Xiao, Z.H.; Huang, S.X.; Fan, P. Prediction model for dissolved gas concentration in transformer oil based on modified grey wolf optimizer and LSSVM with grey relational analysis and empirical mode decomposition. *Energies* **2020**, *13*, 422. [CrossRef]
18. Santhosh, M.; Venkaiah, C.; Vinod, K.D.M. Ensemble empirical mode decomposition based adaptive wavelet neural network method for wind speed prediction. *Energy Convers. Manag.* **2018**, *168*, 482–493. [CrossRef]
19. Wang, Y.M.; Wu, L. On practical challenges of decomposition-based hybrid forecasting algorithms for wind speed and solar irradiation. *Energy* **2016**, *112*, 208–220. [CrossRef]
20. Zheng, Q.; Yan, P.; Hamidreza, Z.; Chen, N.Y. A review and discussion of decomposition-based hybrid models for wind energy forecasting applications. *Appl. Energy* **2019**, *235*, 939–953.
21. Fang, W.; Huang, S.Z.; Ren, K.; Huang, Q.; Huang, G.H.; Cheng, G.H.; Li, K.L. Examining the applicability of different sampling techniques in the development of decomposition-based streamflow forecasting models. *J. Hydrol.* **2019**, *568*, 534–550. [CrossRef]
22. Jiang, Y.; Liu, S.; Zhao, N.; Xin, J.; Wu, B. Short-term wind speed prediction using time varying filter-based empirical mode decomposition and group method of data handling-based hybrid model. *Energy Convers. Manag.* **2020**, *220*, 113076. [CrossRef]
23. Wang, T.; Zhang, M.C.; Yu, Q.H.; Zhang, H.Y. Comparing the applications of EMD and EEMD on time-frequency analysis of seismic signal. *J. Appl. Geophys.* **2012**, *83*, 29–34. [CrossRef]
24. Hodrick, R.J.; Prescott, E.C. Postwar U.S. business cycles: An empirical investigation. *J. Money Credit Bank.* **1997**, *29*, 1–16. [CrossRef]
25. Weron, R.; Zator, M. A note on using the Hodrick Prescott filter in electricity markets. *Energy Econ.* **2015**, *48*, 1–6. [CrossRef]
26. Tong, C.F.; Shi, H.B.; Bao, X.Q.; Li, H.P. Application of a combined model based on wavelet analysis for predicting crop water requirement. *Trans. Chin. Soc. Agric. Eng.* **2011**, *27*, 93–98.
27. Zeng, Q.H.; Huang, L.; Liu, J.Y.; Chen, L.J.; Gu, S.S. Real-time filtering methods of FOG random noise based on ARMA model. *J. Chin. Inert. Technol.* **2015**, *23*, 120–124.
28. Li, Z.Q.; Zhang, X.Y.; Tian, H.D. Prediction method of satellite telemetry data using HP filter. *Spacecr. Eng.* **2021**, *30*, 23–30.
29. Wang, C.; Zhang, H.L.; Ma, P. Wind power forecasting based on singular spectrum analysis and a new hybrid laguerre neural network. *Appl. Energy* **2020**, *259*, 114139. [CrossRef]
30. Moreno, S.R.; Coelho, L.D.S. Wind speed forecasting approach based on singular spectrum analysis and adaptive neuro fuzzy inference system. *Renew. Energy* **2018**, *126*, 736–754. [CrossRef]
31. Dragomiretskiy, K.; Zosso, D. Variational mode decomposition. *IEEE Trans. Signal Process.* **2014**, *62*, 531–544. [CrossRef]
32. Azami, H.; Arnold, S.E.; Sanei, S.; Chang, Z.Q.; Sapiro, G.; Escudero, J.; Gupta, A.S. Multiscale fluctuation-based dispersion entropy and its applications to neurological diseases. *IEEE Access* **2019**, *7*, 68718–68733. [CrossRef]
33. Liao, W.Q.; Fu, W.L.; Yang, K.; Tan, C.; Huang, Y.G. Multi-scale residual neural network with enhanced gated recurrent unit for fault diagnosis of rolling bearing. *Meas. Sci. Technol.* **2024**, *35*, 056114. [CrossRef]

34. Chen, W.T.; Wang, Z.Z.; Xie, H.B.; Yu, W.X. Characterization of surface EMG signal based on fuzzy entropy. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2007**, *15*, 266–272. [CrossRef]
35. Agga, A.; Abbou, A.; Labbadi, M.; Houm, Y.E.; Ali, I.H.O. CNN-LSTM: An efficient hybrid deep learning architecture for predicting short-term photovoltaic power production. *Electr. Power Syst. Res.* **2022**, *208*, 107908. [CrossRef]
36. Aziz, W.; Arif, M. Multiscale permutation entropy of physiological time series. In Proceedings of the 2005 Pakistan Section Multitopic Conference, Karachi, Pakistan, 24–25 December 2005.
37. Feng, H.W.; Ji, H.X.; Zhi, F.H.; Xu, G. Stock price prediction based on SSA and SVM. *Procedia Comput. Sci.* **2014**, *31*, 625–631. [CrossRef]
38. Lu, T.D.; Xie, J.X. Deformation monitoring data de-noising method based on variational mode decomposition combined with sample entropy. *J. Geod. Geodyn.* **2021**, *41*, 1–6.
39. Morabito, F.C.; Labate, D.; Foresta, F.L.; Bramanti, A.; Morabito, G.; Palamara, I. Multivariate multi-scale permutation entropy for complexity analysis of Alzheimer's disease EEG. *Entropy* **2012**, *14*, 1186–1202. [CrossRef]
40. Xue, J.K.; Shen, B. A novel swarm intelligence optimization approach: Sparrow search algorithm. *Syst. Sci. Control Eng.* **2020**, *8*, 22–34. [CrossRef]