

Article

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

Bing Zeng ^{1,2} , Jiang Guo ^{1,2,*}, Fangqing Zhang ^{1,2}, Wenqiang Zhu ^{1,2}, Zhihuai Xiao ², Sixu Huang ^{1,2} and Peng Fan ^{3,4}

¹ Intelligent Power Equipment Technology Research Center, Wuhan University, Wuhan 430072, China; zengbing_whu@whu.edu.cn (B.Z.); zfq@whu.edu.cn (F.Z.); 200632650021@whu.edu.cn (W.Z.); huangsixu@whu.edu.cn (S.H.)

² College of Power Mechanical Engineering, Wuhan University, Wuhan 430072, China; xiaozhihuai@whu.edu.cn

³ NARI Group Corporation (State Grid Electric Power Research Institute), Nanjing 211106, China; fanpeng2@sgepri.sgcc.com.cn

⁴ Wuhan NARI Limited Company of State Grid Electric Power Research Institute, Wuhan 430074, China

* Correspondence: guo.river@whu.edu.cn; Tel.: +86-135-5444-5122

Received: 29 November 2019; Accepted: 8 January 2020; Published: 15 January 2020



Abstract: Oil-immersed transformer is one of the most important components in the power system. The dissolved gas concentration prediction in oil is vital for early incipient fault detection of transformer. In this paper, a model for predicting the dissolved gas concentration in power transformer based on the modified grey wolf optimizer and least squares support vector machine (MGWO-LSSVM) with grey relational analysis (GRA) and empirical mode decomposition (EMD) is proposed, in which the influence of transformer load, oil temperature and ambient temperature on gas concentration is taken into consideration. Firstly, GRA is used to analyze the correlation between dissolved gas concentration and transformer load, oil temperature and ambient temperature, and the optimal feature set affecting gas concentration is extracted and selected as the input of the prediction model. Then, EMD is used to decompose the non-stationary series data of dissolved gas concentration into stationary subsequences with different scales. Finally, the MGWO-LSSVM is used to predict each subsequence, and the prediction values of all subsequences are combined to get the final result. DGA samples from two transformers are used to verify the proposed method, which shows high prediction accuracy, stronger generalization ability and robustness by comparing with LSSVM, particle swarm optimization (PSO)-LSSVM, GWO-LSSVM, MGWO-LSSVM, EMD-PSO-LSSVM, EMD-GWO-LSSVM, EMD-MGWO-LSSVM, GRA-EMD-PSO-LSSVM and GRA-EMD-GWO-LSSVM.

Keywords: dissolved gas analysis; empirical mode decomposition; grey relation analysis; grey wolf optimizer; least squares support vector machine

1. Introduction

The transformer is the core equipment of power system and its running state is closely related to the reliability and stability of power grid. The catastrophic failure of the transformer will lead to a power failure accident, and the power system will be damaged, which will bring huge economic loss and social harm. Therefore, it is very important to detect potential faults in transformer. Dissolved gas analysis (DGA) is widely used in transformer internal latent fault diagnosis. A failure of a power

transformer usually results in degradation of the insulation, leading to the release of gases dissolved in oil. The composition of dissolved gas is closely related to the abnormal state inside the transformer. The fault-related characteristic gases mainly include hydrogen (H_2), methane (CH_4), acetylene (C_2H_2), ethylene (C_2H_4), ethane (C_2H_6), carbon monoxide (CO) and carbon dioxide (CO_2). In the case of electrical or thermal failure, the concentration of gases varies gradually and regularly over time. By analyzing the change trend of the dissolved gas concentration, the operating state of the transformer can be obtained to determine the fault type, and the potential risks of the transformer can be detected in time to predict the development trend of the latent faults of the transformer, so as to avoid the occurrence of serious events and minimize the loss. Therefore, it is of great significance to study the prediction method of dissolved gas concentration and the prediction results can provide a basis for transformer state evaluation and fault prediction.

Recently, various artificial intelligence techniques have been used to develop time series prediction models and achieved good results, such as artificial neural network (ANN) [1], grey model (GM) [2], support vector machine (SVM) [3–5] and least squares support vector machine (LSSVM) [6,7]. However, these methods have some drawbacks. For example, training speed of ANN is slow, and it is easy to fall into local minimum, and a large number of training samples are needed in the training. GM (1,1) can reveal the development law of things with a small amount of incomplete information, but only consider the development and change of a certain gas, and lack of comprehensive consideration of a variety of gases.

In addition, many scholars put forward many improved prediction models, which can be applied to the prediction of dissolved gas in power transformer oil. Lu et al. used the Gaussian process regression (GPR) to predict the dissolved gas concentration, in which the grey relational coefficients of gas concentration were analyzed by using grey relational analysis, and then the performance of the model was improved [8]. By using improved fruit fly optimization algorithm (FFOA) to select the smooth factor, Lin et al. proposed a combined prediction model based on kernel principal component analysis (KPCA) and generalized regression neural network (GRNN) to predict gas concentration, obtaining better data fitting and more accurate prediction [9]. Zheng et al. proposed an improved particle swarm optimization algorithm combined with LSSVM based on wavelet technology to predict the dissolved gas concentration. The comparison results showed that the mean absolute percentage error (MAPE) of this method was significantly better than the other four methods [10]. Pereira, F.H. et al. proposed a nonlinear autoregressive neural network model combined with discrete wavelet transform to predict the concentration of dissolved gas, which shows better prediction results compared with the current prediction models and the commonly used time series techniques [11]. Lin et al. proposed a transformer operation state prediction method based on long short-term memory and deep belief network (LSTM_DBN), which predicted the dissolved gas concentration by developing a long short term memory (LSTM) model [12]. On the basis of radial basis function neural network (RBFNN), back propagation neural network (BPNN), LSSVM of two different kernel functions and grey model, Liu et al. proposed a combined prediction model based on cross entropy, in which the weight coefficient of each algorithm is determined by cross entropy theory, and analyzed its application [13]. Peimankar, et al. proposed an integrated time series prediction algorithm based on evolutionary multi-objective optimization algorithm for predicting dissolved gas concentration in power transformers, which has higher accuracy and reliability [14].

Although the intelligent prediction method mentioned above improves the accuracy of the prediction model, there is still room for improvement. Moreover, the possible influence of transformer load, operating oil temperature, ambient temperature and other variables on gas concentration is not considered in the prediction methods. Therefore, the practical application of these algorithms is limited. At present, the commonly used intelligent optimization algorithms, including genetic algorithm (GA), particle swarm optimization (PSO), differential evolution (DE), etc., have achieved good results in parameter optimization of LSSVM. The grey wolf optimizer (GWO) is a novel swarm intelligence optimization algorithm proposed by Mirjalili et al. in 2014 [15], which shows good

performance in strong convergence, few parameters and easy implementation compared with GA, PSO and DE, and has attracted the attention of many scholars [16–18]. Due to the slow convergence rate of GWO in the late stage, it is easy to fall into the local optimization. The modified grey wolf optimizer (MGWO) proposed in this paper adjusts the exploration and exploitation of the algorithm and assigns more weight to the most suitable grey wolf, so as to find the optimal new position of grey wolf in the iterative process. The MGWO is applied to the parameter optimization of LSSVM. In the prediction model, the influence of the transformer oil temperature, load and ambient temperature on the concentration of dissolved gas is taken into consideration. Firstly, grey relational analysis (GRA) is used to evaluate the correlation between gas concentration and transformer load, oil temperature and ambient temperature to extract the main factors influencing the gas concentration as the input of the model. Then the non-stationary series DGA data is decomposed into subsequences with different scales by using empirical mode decomposition (EMD). Finally, the MGWO-LSSVM model is used to predict subsequences of each gas to get the final concentration of dissolved gas and the validity and superiority of the model are verified.

This paper is organized as follows: In Section 2, the basic theory of the GRA-EMD-MGWO-LSSVM model is introduced. The GRA-EMD-MGWO-LSSVM model is proposed in Section 3. In Section 4, the performance of GRA-EMD-MGWO-LSSVM model is verified by comparing with other models and Section 5 provides the conclusion of the work and discussion of potential future work.

2. Related Theory

2.1. Grey Relational Analysis

Grey relational analysis (GRA) is an analysis method based on grey system theory [19]. Its basic idea is to evaluate the correlation degree between various factors according to the similarity degree of the geometric shape of the change curve of each factor. By quantitative analysis of the development trend of the dynamic process, the method compares the geometric relations of the relevant statistical data of time series and calculates the grey relational degree of each factor. The concrete steps of GRA are as follows:

Step 1 Let the reference time series $X_0 = \{X_0(k) | k = 1, 2, 3, \dots, n\}$, the comparison sequence $X_i = \{X_i(k) | k = 1, 2, \dots, n; i = 1, 2, \dots, m\}$. The original data is dimensionless processed according to Formula (5):

$$x_i(k) = \frac{X_i(k)}{X_i(1)}, k = 1, 2, \dots, n; i = 1, 2, \dots, m \quad (1)$$

Step 2 Calculate the grey relational coefficient. Grey relational coefficient is calculated as follows:

$$\xi_i(k) = \frac{\min_i \min_k |x_0(k) - x_i(k)| + \rho \max_i \max_k |x_0(k) - x_i(k)|}{|x_0(k) - x_i(k)| + \rho \max_i \max_k |x_0(k) - x_i(k)|} \quad (2)$$

where $\xi_i(k)$ is the grey relational coefficient $x_0(k)$ and $x_i(k)$, which reflects the close degree of two sequences at some point. Constant ρ is the resolution coefficient and its value range is $(0, 1)$, while a smaller ρ indicating greater discrimination. To increase the difference between correlation coefficients, $\rho = 0.5$ is usually taken.

Step 3 Calculate the grey relational degree. By integrating the grey relational coefficients of all points, the grey correlation degree of X_i and X_0 can be calculated as follows:

$$r_i = \frac{1}{n} \sum_{k=1}^n \xi_i(k), k = 1, 2, \dots, n; i = 1, 2, \dots, m \quad (3)$$

r_i reflects the degree of correlation between X_i and X_0 . A larger r_i means higher correlation degree, closer relationship and closer development trend and rate.

Step 4 Grey relational degree ranking. The grey relational degree of influencing factor sequence to the system behavior characteristic sequence is ranked from large to small.

Since there is no definite qualitative and quantitative description between dissolved gas concentration, oil temperature, transformer load and ambient temperature, and there is uncertainty in the mutual restriction relationship among all gases, the grey relational degree is used to measure the affinity among all factors, and obtains the main factors that influence each dissolved gas concentration.

2.2. Empirical Mode Decomposition

Empirical mode decomposition (EMD) [20–22] is a signal decomposition method based on local characteristics of signals, which absorbs the advantage of multi-resolution of wavelet transform and overcomes the difficulty of choosing wavelet basis and determining decomposition scale in wavelet transform, so it is more suitable for nonlinear non-stationary signal analysis and is an adaptive signal decomposition method. The EMD assumes that any complex signal is composed of simple intrinsic mode functions (IMF), and each IMF is mutually independent. This EMD can decompose different scales or trends in time series data into its component step by step, and a series of data sequences with the same characteristics of the scale are produced, by which the non-stationary nonlinear data is transformed into a smooth linear data. Compared with the original data sequence, the sequence after decomposition is with greater regularity, which is of great help on identifying hidden relationship and can improve the prediction accuracy [23–25]. The steps of EMD for a given time series are as follows:

Step 1: Determine the upper envelope $e_{up}(t)$ and the lower envelope $e_{low}(t)$ from the local maximum and local minimum of time series data $x(t)$, and calculate the mean envelope $m_1(t)$:

$$m_1(t) = \frac{e_{up}(t) + e_{low}(t)}{2} \quad (4)$$

Step 2: Subtract $m_1(t)$ from $x(t)$ to get $h_1(t)$, and consider $h_1(t)$ as a new signal $x(t)$, repeat Step 1, by k times of screening, until $h_1(t) = x(t) - m_1(t)$ meets IMF conditions, then $c_1(t)$ is the first IMF component of the time series, as shown in (2), it contains the shortest periodic component in the original sequence.

$$c_1(t) = x(t) - m_1(t) \quad (5)$$

Step 3: After separating the first IMF component from the time series $x(t)$, the remaining component $r_1(t)$ of $x(t)$ can be obtained as follows:

$$r_1(t) = x(t) - c_1(t) \quad (6)$$

Step 4: Take $r_1(t)$ as a new time series and repeat Step 1–3 to obtain a series of qualified IMF components $c_i(t)$ and residual $r_n(t)$. Then, the original time series $x(t)$ can be described by IMF components and residual component as follows:

$$x(t) = \sum_{n=1}^n c_i(t) + r_n(t) \quad (7)$$

From Step 1 to Step 4, the original time series can be decomposed into sub-sequences of different frequencies, namely IMF and residual r . Then, the trend prediction of each subsequence is carried out, and the prediction results of the subsequence are superimposed to obtain the prediction results of the original sequence.

2.3. Grey Wolf Optimizer

2.3.1. Standard Grey Wolf Optimization

The grey wolf optimizer (GWO) [14] is a novel swarm intelligence algorithm inspired by predation behavior of grey wolves by Mirjalili et al., which mimics hunting behavior and social leadership of grey wolves in nature and uses four types of grey wolves to simulate social hierarchy.

The three wolves with the best fitness are alpha (α), beta (β) and delta (δ), while the remaining wolves are omega (ω). GWO algorithm realizes the global optimization by imitating the predation behavior of grey wolves such as encircling, hunting and attacking. The optimization process is mainly guided by the three best solutions (α , β , and δ) in each generation of the population.

The mathematical model of the behavior of grey wolves encircling their prey is as follows:

$$D = C \cdot X_p(t) - X(t), \quad (8)$$

$$X(t+1) = X_p(t) - A \cdot D, \quad (9)$$

$$A = 2a \cdot r_1 - a, \quad (10)$$

$$C = 2r_2, \quad (11)$$

where: t is the number of current iterations; A and C are coefficient vectors; X_p represents the position vector of prey; $X(t)$ represents the position vector of the current grey wolf. In the whole iteration, a is the convergence factor, which decreases linearly from 2 to 0; r_1 and r_2 are random vectors in (0,1).

Assuming that α , β , and δ are capable of identifying potential prey positions, the three best wolves in the current population are retained during each iteration and the positions of other search agents are updated based on their positions. The mathematical model of this behavior can be expressed as follows:

$$D_\alpha = C_1 \cdot X_\alpha - X, D_\beta = C_2 \cdot X_\beta - X, D_\delta = C_3 \cdot X_\delta - X, \quad (12)$$

where: X_α , X_β , and X_δ represents the current position of α , β , and δ wolf respectively; X is the current location of the ω wolf and D_α , D_β and D_δ represents the distance between the current candidate and the three optimal wolves respectively. C_1 , C_2 and C_3 are random vectors.

The vector positions of the prey can be determined based on the α , β and δ positions using the following equations:

$$X_1 = X_\alpha - A_1 \cdot D_\alpha, X_2 = X_\beta - A_2 \cdot D_\beta, X_3 = X_\delta - A_3 \cdot D_\delta, \quad (13)$$

$$(t+1) = \frac{X_1 + X_2 + X_3}{3} \quad (14)$$

where: X_1 , X_2 , and X_3 represents the current position of the prey determined α , β , and δ wolf respectively; $X(t+1)$ is the final location of the prey determined based on the X_1 , X_2 , and X_3 .

2.3.2. Modified Grey Wolf Optimization

Although GWO algorithm shows its advantages in many fields, it is easy to fall into the local optimum and the calculation speed and the accuracy is a little low, which limit the application of the algorithm. Therefore, this paper improved the original grey wolf optimization algorithm as follows.

In the process of grey wolf population approaching the target, the position updating Equation (14) shows the equal importance of α , β , δ , ignoring the different characteristics of the three wolves, and the proportion of the leading position and the optimal solution of grey wolf is not well reflected in Equation (14). Considering that the contribution of each grey wolf is different, different weights

are given to individual grey wolf of different social hierarchy [26]. The position update Equation (15) was adopted:

$$X(t+1) = \frac{3X_1 + 2X_2 + X_3}{6} \quad (15)$$

In the optimization algorithm, excess exploration of search space may result in lower probability of getting trapped in local optima, while too much exploitation is related to less randomness and the algorithm may not reach the global optimum. Hence, a balance between exploration and exploitation should be maintained during the iterations. In the original GWO, the value of a decreases linearly from 2 to 0 during the iteration, and the updated equation is as follows:

$$a = 2\left(1 - \frac{t}{MaxIter}\right) \quad (16)$$

where $MaxIter$ represents the maximum number of iterations, and t is the current number of iterations. The update equation makes half the iteration for exploration and the other half for exploitation. In this paper, Formula (17) was used for the attenuation of a in the iterative process, and the value of a was reduced from 2 to 0 [27]:

$$a = 2\left(1 - \frac{t^2}{MaxIter^2}\right) \quad (17)$$

The decay function in Formula (17) made the number of iterations used for exploration and exploitation are 70% and 30% respectively. In this way, more iterations were used for exploration, while fewer iterations were used for exploitation, so as to achieve better performance than the original GWO algorithm [27]. The pseudo-code of the MGWO (Algorithm 1) is presented in the following form.

Algorithm 1. MGWO pseudo-code

- (1) Initialize the positions of grey wolf population X_i ($i = 1, 2, 3 \dots, n$) randomly.
 - (2) Initialize a , A , C .
 - (3) Find α , β , and δ as the first three best solutions based on their fitness values.
- $t = 0$.
- while** $t \leq MaxIter$ **do**
- foreach** $Wolf_i \in pack$ **do**
- Update current wolf's position according to Equation (15).
- end**
- Update a as in Equation (17).
 - Update A , and C as in Equations (10) and (11).
 - Evaluate the positions of individual wolves.
 - Update α , β , and δ positions as the first best three solutions in the current population.
 - $t = t + 1$.
- end**
- (4) Select the optimal grey wolf position.
-

2.4. Least Square Support Vector Machine

SVM [28] is a machine learning technology based on the structural risk minimization principle, which has shown excellent learning performance and generalization ability in the fields of regression analysis, pattern recognition, fault diagnosis and time series trend prediction. LSSVM is an extension of standard SVM. LSSVM algorithm replaces the inequality constraint in traditional SVM with

equality constraint and replaces the relaxation variable in SVM with the square of training error, thus transforming the quadratic programming problem into the equality constraint problem of linear equations, which obtains faster solution speed and stronger real-time performance. The mathematical model of least squares support vector machine is as follows [7]:

Given training data set $D = \{(x_i, y_i) | i = 1, 2, 3, \dots, N\}$, $x_i \in R^n$, $y_i \in R$. The optimal decision function is constructed in the feature space as follows:

$$y(x) = \omega^T \varphi(x_i) + b, \quad (18)$$

where ω is the weight vector and b is the deviation. Use the structural risk minimization (SRM) principle to calculate the parameters ω and b by the minimization Formula (19):

$$\min_{\omega, b, \xi} J(\omega, \xi) = \frac{1}{2} \|\omega^2\| + \gamma \sum_{i=1}^N \xi_i^2, \quad (19)$$

Equation (19) satisfies the equation constraint:

$$y_i = \omega^T \varphi(x_i) + b + \xi_i, i = 1, 2, 3, \dots, N, \quad (20)$$

In Equation (19), $\|\omega^2\|$ controls the complexity of the model, $\gamma \sum_{i=1}^N \xi_i^2$ is error control function, γ is the regularization parameter and ξ_i is fitting error of the sample i . By introducing Lagrange multiplier $\alpha > 0$, the unconstrained optimization problem can be obtained as follows:

$$L(\omega, b, \xi, \alpha) = J(\omega, \xi) - \sum_{i=1}^N \alpha_i \{\omega^T \varphi(x_i) + b + \xi_i - y_i\}, \quad (21)$$

According to KKT (Karush-Kuhn-Tucker) optimal conditions, take the partial derivative of ω , b , ξ_i and α_i through L respectively and make them all zero, finally the LSSVM model is obtained:

$$y(x) = \sum_{i=1}^N \alpha_i K(x, x_i) + b, \quad (22)$$

where $K(x, x_i)$ is the kernel function. Due to the non-linear relationship among the dissolved gas concentrations in transformer oil, the radial basis kernel function (RBF) suitable for solving nonlinear problems and with fewer nuclear parameters is selected as the kernel function.

$$K(x, x_i) = \exp\left(-\frac{\|x - x_2\|^2}{2\sigma^2}\right), \quad (23)$$

where σ^2 is the kernel parameter.

In this paper, the MGWO was adopted to optimize LSSVM model parameters γ and σ to improve the prediction accuracy of the model. The steps of MGWO-LSSVM model for time series prediction were as follows:

- Step 1 Normalize the sample data and change the original data linearly to the interval (-1.1) , and use it as the training data set.
- Step 2 Select different parameters for the training data to conduct experiments and generate training tables with different parameters. In this step, the parameters of each set of decomposed IMF sequence prediction models should be optimized.
- Step 3 Select the appropriate optimal parameters according to the training error and the comprehensive performance of different parameters from the training result.

- Step 4 The training model is generated after learning the training data by using the selected parameters, and then the prediction data is used as input to test the prediction results. If the prediction results do not meet with the prediction accuracy requirements, then return to Step 3 and re-select the parameters for learning, while the prediction results conform to the prediction accuracy, proceeding to Step 5.
- Step 5 After the parameters are determined, using the data sequence for prediction, and finally conduct error analysis.

3. Prediction Model Based on GRA-EMD-MGWO-LSSVM

The time series data of dissolved gas concentration in transformer oil have strong nonlinearity and non-stationary. In view of the prominent advantages of EMD technology in non-stationary data processing, this paper proposed a prediction model for dissolved gas concentration that integrated grey relational analysis, empirical mode decomposition and modified grey wolf optimized least squares support vector machine. The prediction steps were as follows:

- Step 1 Sample collection: The DGA data were collected, including the sampling time, oil temperature, load, ambient temperature and gas concentration. The gas concentration, oil temperature, load and ambient temperature were listed as the comparison sequence, and the gas concentration to be predicted as the reference sequence.
- Step 2 Grey relational analysis: The initial value transformation was performed on the original data according to Formula (1), the grey relational coefficient was calculated according to Formula (2), and grey relational degree is calculated according to Formula (3). According to the grey incidence matrix, the main influencing factors of each gas concentration were screened out, and the factors with weak relational degree were eliminated.
- Step 3 EMD processing: The EMD method was used to decompose the time series data of dissolved gas concentration and obtain $c_i(t)$ and residual $r_n(t)$ of each IMF component.
- Step 4 MGWO-LSSVM model prediction: LSSVM regression model was established for each IMF component $c_i(t)$ and $r_n(t)$ to obtain the predicted values of each decomposition sequence, and MGWO was used to select optimal parameters.
- Step 5 The predicted values of each component were superimposed to obtain the predicted values of gas concentration.
- Step 6 Verification of prediction results: compared with actual data, calculated error index and conducted error analysis.

The prediction model proposed in this paper based on MGWO-LSSVM combined with EMD and GRA was shown in Figure 1. It consists of four main parts. Firstly, GRA was used to conduct correlation analysis of the original DGA data. Secondly, EMD was used to decompose the DGA data to get the IMF components of each gas concentration. Thirdly, MGWO was used to optimize LSSVM model to predict the trend of each IMF components and finally, the prediction result was being evaluated.

In order to verify the prediction accuracy of the model, the average absolute percentage error (MAPE) η_{MAPE} , root mean square error (RMSE) η_{RMSE} and maximum average absolute percentage error η_{max} to evaluate the predictive results of the proposed algorithm [29,30], it is defined as:

$$\eta_{MAPE} = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i} \times 100\%, i = 1, 2, \dots, n, \quad (24)$$

$$\eta_{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}, i = 1, 2, \dots, n, \quad (25)$$

$$\eta_{max} = \max\left(\frac{|y_i - \hat{y}_i|}{y_i}\right), i = 1, 2, \dots, n, \quad (26)$$

where y_i is the actual value; \hat{y}_i is the predicted value and n is the number of samples. max means to calculate the maximum relative error of $\frac{|y_i - \hat{y}_i|}{y_i}$. A large average relative error means a low prediction accuracy.

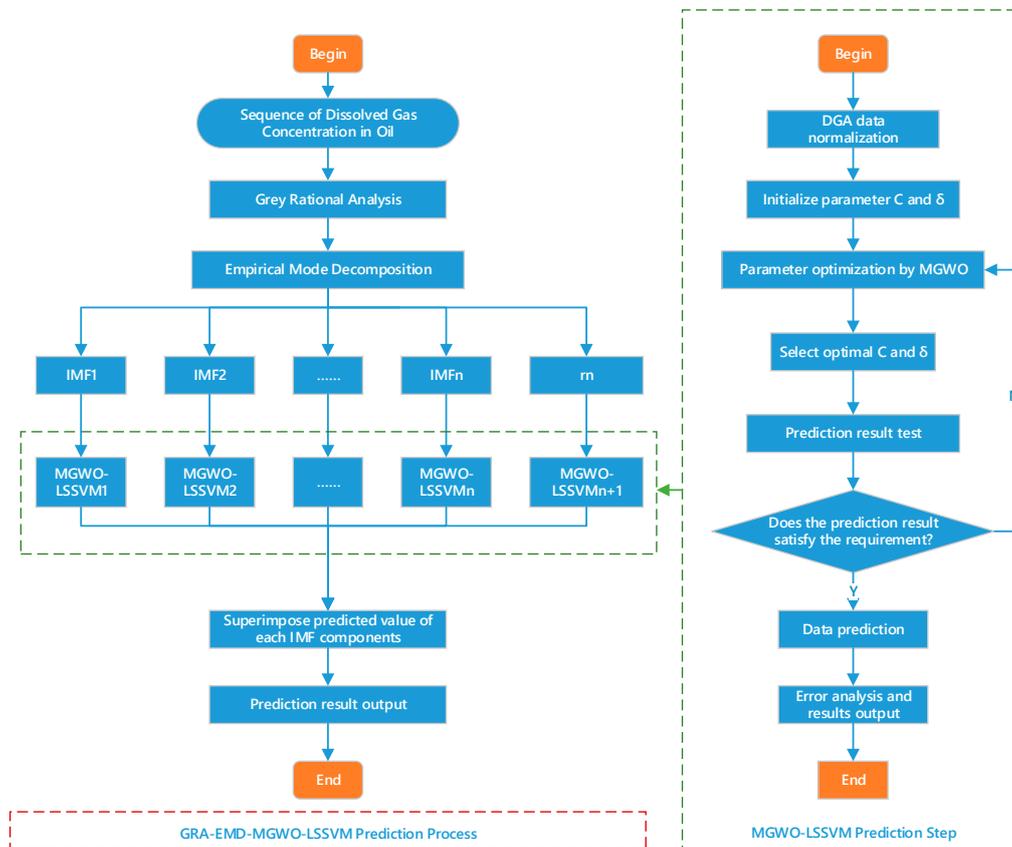


Figure 1. Flowchart of the prediction model based on grey relational analysis-empirical mode decomposition-modified gray wolf optimizer-least squares support vector machine (GRA-EMD-MGWO-LSSVM).

4. Case Study and Analysis

In this paper, the prediction model based on GRA-EMD-MGWO-LSSVM was implemented by the MATLAB simulation platform (R2018b, MathWorks, Natick, Massachusetts, USA) on an 8-core Lenovo laptop (T470P, Lenovo, Beijing, China) with 8 GB memory and 2.8 GHz clock, running Windows 10 enterprise operating system (64-bit). In addition, two examples were given to test and verify the model.

4.1. Prediction Example 1

In this part, the DGA data were from a transformer with a voltage of 750 kV in State Grid Corporation of China. The equipment model was BKD-120000/800. Since 9 January 2012, the workers took DGA samples every three days with the oil temperature, load and ambient temperature recorded.

4.1.1. Sample Collection

The DGA data of the transformer during 9 January 2012 and 3 July 2012 are shown in Table 1, where A1-A11 represents H_2 ($\mu\text{L/L}$), CH_4 ($\mu\text{L/L}$), C_2H_6 ($\mu\text{L/L}$), C_2H_2 ($\mu\text{L/L}$), C_2H_4 ($\mu\text{L/L}$), CO ($\mu\text{L/L}$), CO_2 ($\mu\text{L/L}$), total hydrocarbon ($\mu\text{L/L}$), oil temperature ($^\circ\text{C}$), load (MW) and ambient temperature ($^\circ\text{C}$), respectively.

Table 1. Distribution of transformer sample data.

Date	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11
2012/1/9	97	30.23	4.02	5.39	17.22	160	610	56.86	23.7	332.4	-9
2012/1/13	96.7	30.28	3.81	5.36	16.59	152.38	615.93	56.04	22.9	332.03	-14
2012/1/17	102	34.62	5.7	5.43	20.6	182	647	66.35	22.8	332.05	-12
2012/1/21	110	43.07	6.72	5.5	28.06	153	553	83.35	23.8	330.91	-14
2012/1/25	101	33	5.77	5.53	23	157	524	67.3	29.7	330.17	-12
2012/1/29	112	42.71	6.63	5.59	25.5	162	490	80.43	28	331.29	-11
2012/2/2	96	44.34	7.38	5.63	31.2	154	466	88.55	25.2	330.54	-11
2012/2/6	102	42.71	6.63	5.7	30.25	198	456	85.29	26.1	332.4	-9
2012/2/10	104	48.91	6.7	5.76	28.7	169	437	90.07	24.7	328.3	-9
2012/2/14	89	42.88	4.88	5.78	25.3	165	455	78.84	33	332.03	-6
2012/2/18	102	42.71	6.63	5.8	25.5	200	409	80.64	31.8	327.19	-6
2012/2/22	104	54.09	7.23	5.82	31.8	161	453	98.94	36	326.07	-5
2012/2/26	102	56.23	8.53	5.89	35.8	167	405	106.45	33.8	330.54	-2
2012/3/1	112	43.6	6.63	6.03	39.5	200	414	95.76	36	158	1
2012/3/5	125	53.67	5.64	6.07	36	200	456	101.38	40.1	329.79	1
2012/3/9	144	53.88	5.66	6.1	36.55	201	449	102.19	24.1	329.79	3
2012/3/13	122	52.71	6.63	6.13	35.01	200	439	100.48	40.8	329.42	3
2012/3/17	122	49.63	6.63	6.16	29.5	200	459	91.92	39.3	334.27	5
2012/3/21	122	42.71	6.63	6.23	25.5	200	409	81.07	39.9	332.78	5
2012/3/25	140	52.68	8.63	6.26	35.8	207	582	103.37	39.7	330.54	7
2012/3/29	147	55.76	9.1	6.31	34.9	197	630	106.07	38.5	332.03	8
2012/4/2	125	45.17	7.4	6.33	28.9	166	596	87.8	39.3	329.42	8
2012/4/6	131	47.12	8.21	6.34	29.65	171	609	91.32	39.5	329.33	12
2012/4/10	132	47.21	8.2	6.34	29.73	171	607	91.48	39.3	326.07	12
2012/4/14	145	42.61	5.74	6.3	24.35	171	591	79	42.1	330.54	12
2012/4/18	137	44.6	7.11	6.37	30.44	160	510	88.52	42.1	334.64	14
2012/4/22	122	42.71	6.63	6.36	25.5	196	496	81.2	49.3	332.78	14
2012/4/26	128	50.73	8.05	6.4	32.07	167	487	97.25	56	329.79	14
2012/4/30	135	63.94	10.78	6.46	39.8	121	392	110.98	48.8	335.01	14
2012/5/4	123	47.8	6.4	6.42	24.68	124	399	85.3	51.6	324.95	14
2012/5/8	122	42.7	6.63	6.43	25.5	160	406	81.26	43.1	324.5	15
2012/5/12	135	54.22	8.3	6.48	30.38	122	456	99.38	44	327	15
2012/5/16	145	65.41	10.03	6.54	35.7	138	496	117.68	48.1	256.2	15
2012/5/20	147	65.36	10.16	6.75	42.2	152	616	124.47	48	336	20
2012/5/24	148	70.87	11.67	6.81	45.2	167	606	134.55	49	335	20
2012/5/28	146	68.97	11.07	6.88	48.8	172	617	135.72	51.9	333.89	23
2012/6/1	147	72.31	11.86	6.94	48	192	650	139.11	52	336	23
2012/6/5	145	69.41	12.06	7	45.7	183	686	134.17	53	334	26
2012/6/9	148	70.6	12.3	7.06	49.8	200	698	139.76	58.3	329.05	27
2012/6/13	148	74.94	12.37	7.15	49.4	208	651	143.86	50.7	332.27	24
2012/6/17	146	70	12.35	7.21	45.7	198	666	135.26	58.2	331.66	27
2012/6/21	144	77.89	12.71	7.26	47.2	202	669	145.06	56.5	333.15	25
2012/6/25	147	77.63	19.72	7.32	53.87	206	680	158.54	55	333.89	25
2012/6/29	146	78.43	17.53	7.33	45.1	201	602	148.39	22.8	330.2	25
2012/7/3	148.05	78.27	11.14	7.39	45.8	206.3	603.58	142.6	52.4	329.05	26

4.1.2. Grey Relational Analysis

The grey relational analysis results of the DGA data are shown in Table 2 and the grey relational degree of each factor ranged from 0.46 to 0.93. It can be seen from Table 2 that:

- (1) There was a strong correlation between the load and gas concentration of H₂ and CO and a less stronger correlation between load and CH₄, C₂H₂, C₂H₄, CO₂, and total hydrocarbon. The correlation between load and the gas concentration of C₂H₆ was weak.
- (2) The correlation between oil temperature and each gas concentration was strong and the grey relational degree ranged from 0.65 to 0.85.
- (3) The grey relational degree of ambient temperature and gas concentration was between 0.65 and 0.86, indicating that there was a strong correlation between ambient temperature and each gas.
- (4) The grey relational degree between oil temperature and ambient temperature was 0.85, which indicates that the ambient temperature was positively correlated with the oil temperature range.

- (5) The grey relational degree of the total hydrocarbon and the hydrocarbon gases was between 0.77 and 0.93, which verified that there was a strong correlation between the hydrocarbon gases and total hydrocarbon while the total hydrocarbon concentration was the sum of the four gas concentrations.

Table 2. Grey relational degree of each factor.

Factor	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11
A1	1.00	0.73	0.62	0.77	0.71	0.68	0.69	0.71	0.75	0.68	0.81
A2	0.73	1.00	0.73	0.83	0.89	0.67	0.71	0.92	0.79	0.57	0.79
A3	0.62	0.74	1.00	0.73	0.75	0.60	0.71	0.77	0.68	0.46	0.67
A4	0.77	0.84	0.73	1.00	0.80	0.68	0.74	0.84	0.85	0.56	0.86
A5	0.71	0.89	0.75	0.80	1.00	0.68	0.73	0.93	0.77	0.55	0.75
A6	0.68	0.68	0.60	0.68	0.68	1.00	0.69	0.67	0.65	0.66	0.65
A7	0.69	0.71	0.71	0.73	0.73	0.69	1.00	0.72	0.70	0.57	0.72
A8	0.71	0.92	0.77	0.84	0.93	0.67	0.73	1.00	0.78	0.54	0.77
A9	0.75	0.79	0.67	0.85	0.77	0.65	0.70	0.78	1.00	0.57	0.85
A10	0.67	0.57	0.46	0.56	0.54	0.66	0.57	0.54	0.57	1.00	0.61
A11	0.81	0.80	0.67	0.86	0.75	0.65	0.72	0.77	0.85	0.61	1.00

4.1.3. Empirical Mode Decomposition

EMD was performed on the time series data of dissolved gas concentration in Table 1 and the EMD results of H₂, CH₄, C₂H₆, C₂H₂, C₂H₄, CO, CO₂ and total hydrocarbon are shown in the Figure 2, respectively. As can be seen from the Figure 2a, the original H₂ nonlinear sequence was decomposed into three less volatile time series components IMF (IMF1-IMF3) and one residual component. Then the IMFs and the residual were predicted by the MGWO-LSSVM model, respectively.

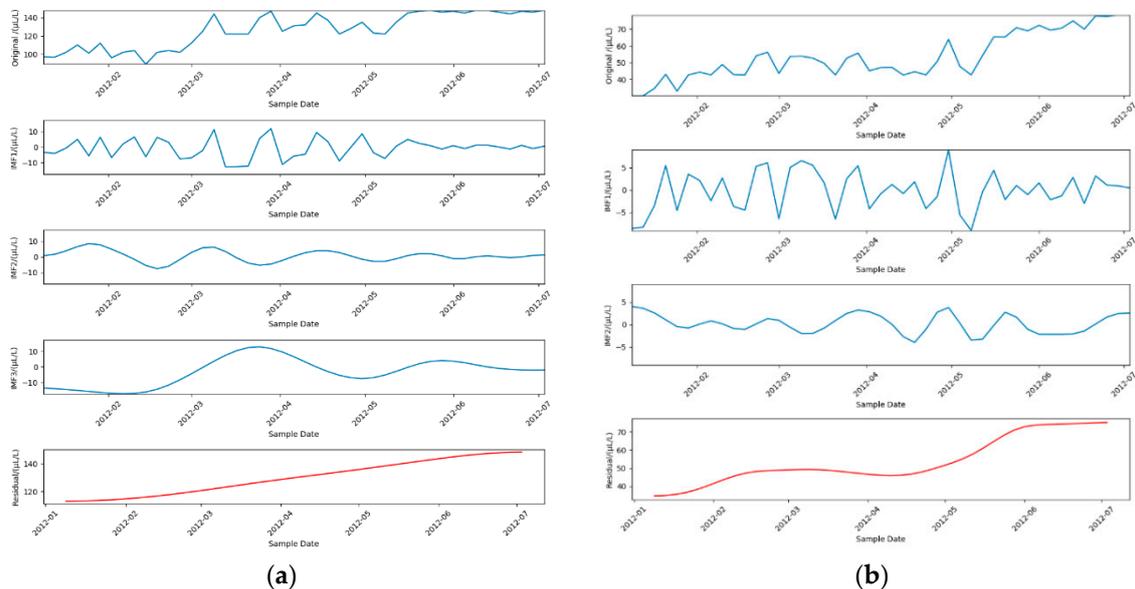


Figure 2. Cont.

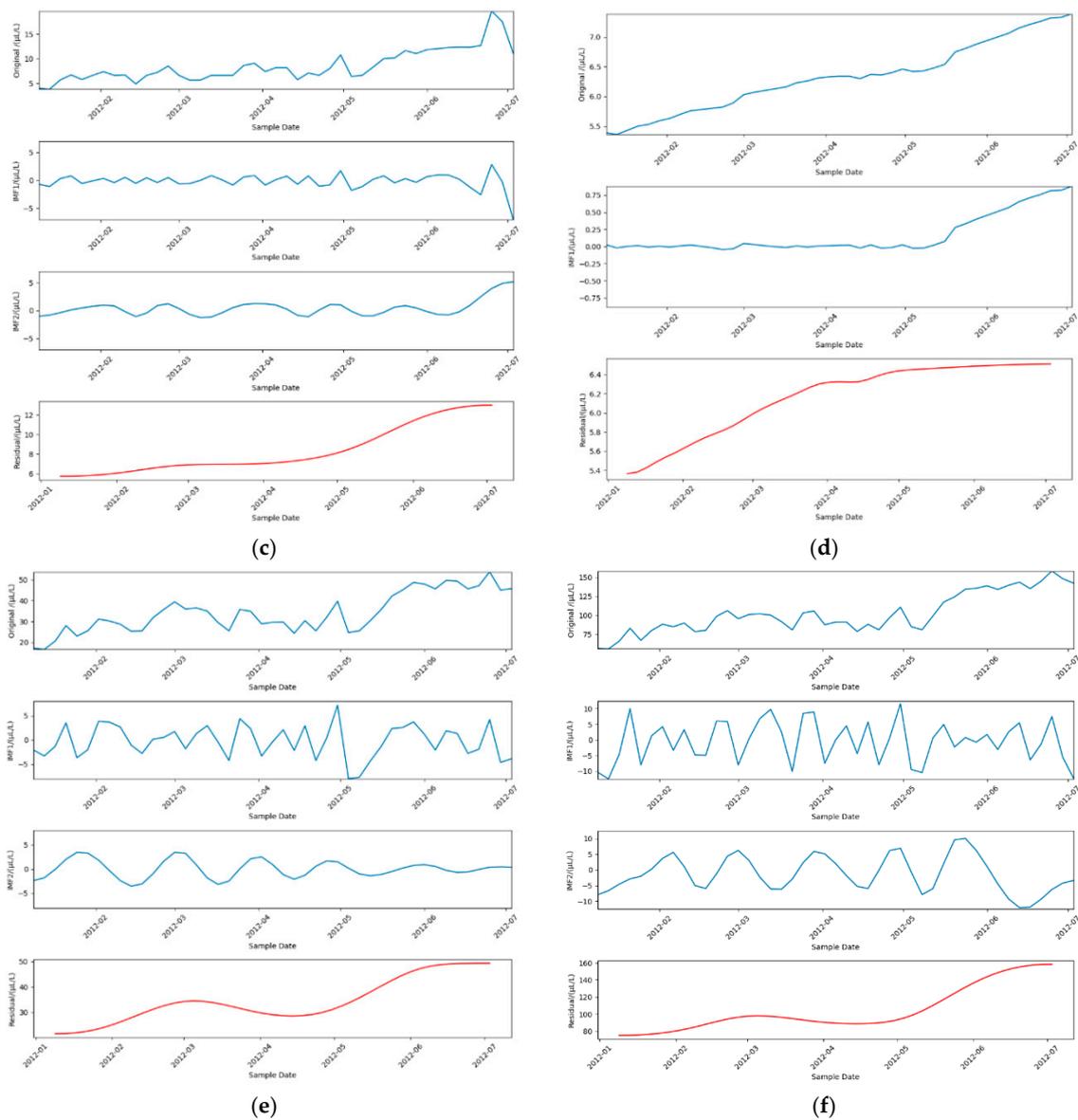


Figure 2. EMD results of the dissolved gas concentration. (a) H_2 ; (b) CH_4 ; (c) C_2H_6 ; (d) C_2H_2 ; (e) C_2H_4 ; and (f) total hydrocarbon.

4.1.4. Results and Discussion

The factors with grey relational degree greater than 0.50 were taken as the main factors to establish the MGWO-LSSVM prediction model. H_2 was taken as an example and the IMFs and residual of H_2 and transformer load, oil temperature and ambient temperature were used as the input of the model. The data from 9 January 2012 to 13 June 2012 were used as the training set, and 17 June 2012 to 3 July 2012 as the test set. The prediction results of H_2 by GRA-EMD-MGWO-LSSVM model are shown in Figure 3.

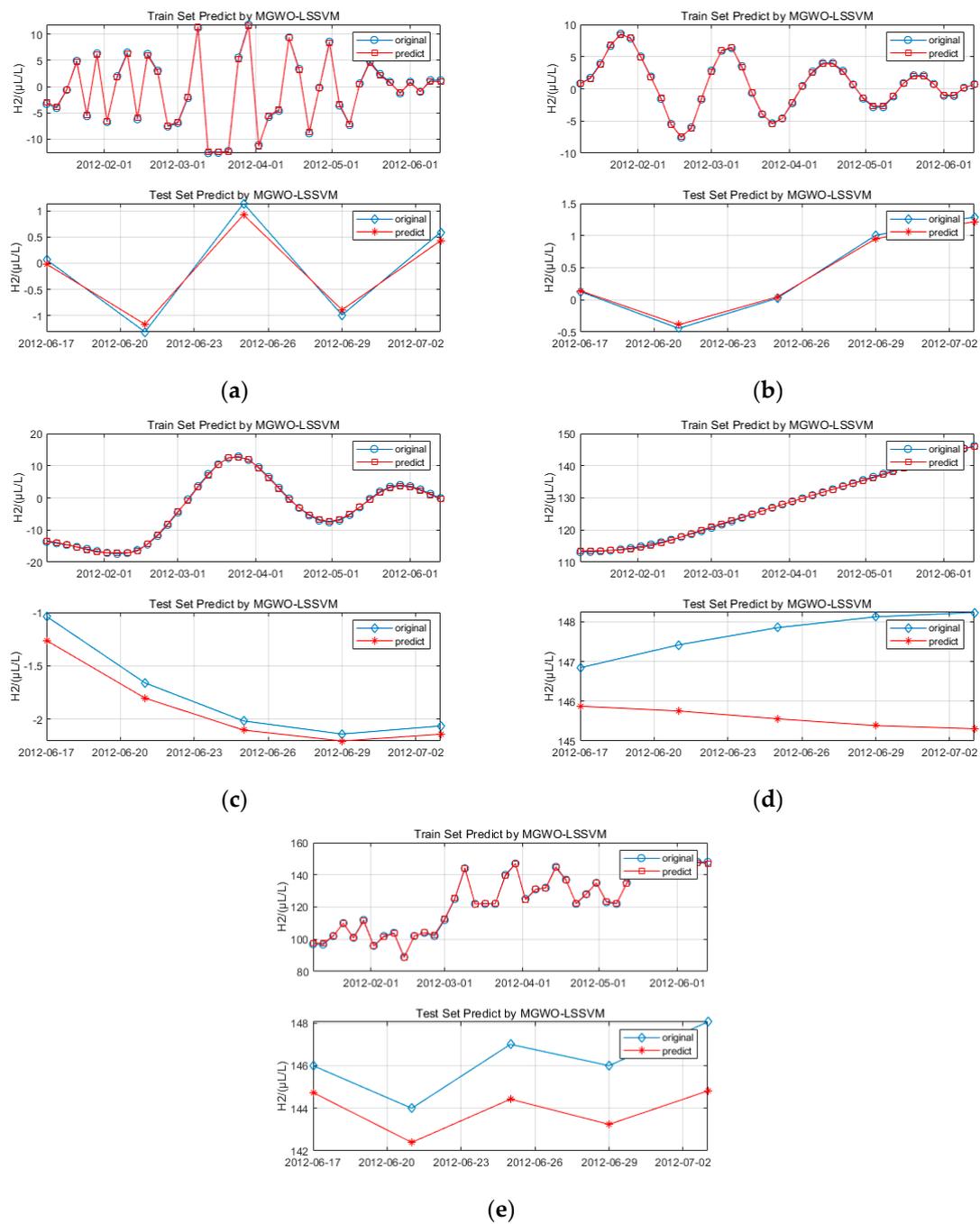


Figure 3. Prediction results of H_2 by the GRA-EMD-MGWO-LSSVM model. (a) intrinsic mode function 1 (IMF1); (b) IMF2; (c) IMF3; (d) residual and (e) H_2 .

In order to verify the validity of the proposed model, the prediction model was compared with several models. The comparison results were shown in Tables 3 and 4. For the prediction models in Table 3, only the dissolved gas concentration data were used as the input without processing by EMD. In Table 4, the dissolved gas concentration data processed by EMD were used as the input of EMD-PSO-LSSVM, EMD-GWO-LSSVM and EMD-MGWO-LSSVM, respectively, while for GRA-EMD-PSO-LSSVM, GRA-EMD-GWO-LSSVM and GRA-EMD-MGWO-LSSVM, the transformer load, oil temperature, ambient temperature and the EMD processing results were used as the input.

Table 3. Comparison of different prediction models.

Model	η_{MAPE} (%)	η_{RMSE}	η_{max} (%)
LSSVM	12.25	8.68	18.5
PSO-LSSVM	10.85	7.26	17.2
GWO-LSSVM	9.72	6.28	15.8
MGWO-LSSVM	8.64	5.58	14.6

Table 4. Comparison of different prediction models.

Model	η_{MAPE} (%)	η_{RMSE}	η_{max} (%)
EMD-PSO-LSSVM	8.54	6.49	12.06
EMD-GWO-LSSVM	7.62	5.76	10.02
EMD-MGWO-LSSVM	6.57	4.38	8.76
GRA-EMD-PSO-LSSVM	4.39	5.25	6.06
GRA-EMD-GWO-LSSVM	3.62	3.86	5.02
GRA-EMD-MGWO-LSSVM	1.3	2.29	2.18

It can be seen from Tables 3 and 4 that:

- (1) Compared with LSSVM, PSO-LSSVM and GWO-LSSVM, the MGWO-LSSVM model obtained higher prediction accuracy and showed good effectiveness.
- (2) Compared with PSO-LSSVM, GWO-LSSVM and MGWO-LSSVM, after the processing step by EMD, EMD-PSO-LSSVM, EMD-GWO-LSSVM and EMD-MGWO-LSSVM shows better performance in prediction, respectively. The EMD method could decompose the concentration time series of dissolved gas in oil to produce a series of stationary data sequences with the same scale, which reduced the influence of nonlinearity and non-stationary of dissolved gas concentration data on the prediction results and improved the accuracy of prediction model.
- (3) By using GRA to conduct a correlation analysis of dissolved gas concentration and transformer load, oil temperature, ambient temperature and the mutual influencing factors of gas concentration were extracted and irrelevant information were reduced, which could improve the prediction accuracy
- (4) Compared with other prediction models, the GRA-EMD-MGWO-LSSVM model proposed in this paper achieved higher prediction accuracy.

4.2. Prediction Example 2

In this part, the raw DGA data were collected from the main transformer of a 750 kV substation of State Grid Corporation of China to verify the proposed model in this paper. The transformer was put into operation in April 2010 and the equipment model was ODFPS-700000/750GY. Since April 21, 2011, trace amounts of acetylene appeared in the transformer and the data of gas concentration, oil temperature, load and ambient temperature were recorded by the workers in the follow-up test once every day.

4.2.1. Sample Collection

Table 5 shows the data of the transformer during April 22, 2011 and May 18, 2011, where A1–A11 represents H_2 ($\mu\text{L/L}$), CH_4 ($\mu\text{L/L}$), C_2H_6 ($\mu\text{L/L}$), C_2H_2 ($\mu\text{L/L}$), C_2H_4 ($\mu\text{L/L}$), CO ($\mu\text{L/L}$), CO_2 ($\mu\text{L/L}$), total hydrocarbon ($\mu\text{L/L}$), oil temperature ($^{\circ}\text{C}$), load (MW) and ambient temperature ($^{\circ}\text{C}$), respectively.

Table 5. Distribution of transformer sample data.

Date	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11
2011/4/22	4.23	1.7	0.9	0.1	2.1	32.07	574.62	4.8	40.5	328	10
2011/4/23	7	1.83	0.75	0.14	1.03	35.05	510.22	3.75	38.3	334.6	12
2011/4/24	13.56	1.88	1.49	0.19	1.6	39.82	509.36	5.16	44.4	327.5	13
2011/4/25	13.02	1.94	0.86	0.23	1.08	39.31	704.2	4.11	48.8	329.42	15
2011/4/26	12.5	2.39	0	0.22	1.5	44.62	178.43	4.11	46	335.01	14
2011/4/27	10.82	2.65	1.31	0.35	1.69	44.78	225.78	6	46.4	334.27	15
2011/4/28	6.14	1.97	0	0.42	1.47	41.13	137.16	3.86	47.1	330	17
2011/4/29	8.63	2.03	0.63	0.41	1.52	40.36	138.52	4.59	36.2	329	10
2011/4/30	8.12	2.01	0.87	0.39	1.45	49.25	168.17	4.72	40.7	330.54	11
2011/5/1	7.29	2.47	0.54	0.55	1.7	43.32	197.67	5.26	38.2	333.5	9
2011/5/2	7.2	2.4	0	0.62	1.5	42.3	304	4.52	43	327.9	12
2011/5/3	9.5	2.82	1.06	0.69	1.43	49.09	294.01	6	44	330	10
2011/5/4	8.16	2.91	0	0.91	2.03	52.92	245.02	5.85	43	333	12
2011/5/5	10.13	2.88	0.97	0.95	1.84	56.15	243.84	6.64	44	343	14
2011/5/6	8.13	2.71	0	0.91	1.85	49.9	581.9	5.47	48.1	309.05	17
2011/5/7	11.41	3.52	0	0.93	2.31	56.43	314.3	6.76	53.2	331.2	20
2011/5/8	11.39	3.44	0	0.97	3.23	54.33	255.8	7.64	33.8	331.66	11
2011/5/9	37.86	4.2	0	0.95	1.69	70.15	205.4	6.84	38.9	270.6	7
2011/5/10	11.58	4.02	2.56	0.78	3.01	67.97	525.9	10.37	41	248.6	12
2011/5/11	9.85	3.39	0.06	0.9	2.51	50.63	516.42	6.86	31.3	332.4	9
2011/5/12	10.17	3.48	0.16	0.91	2.28	54.36	348.39	6.83	30.3	326.4	10
2011/5/13	12.02	4.42	0	1.16	2.06	61.07	507.45	5.58	39.5	330.17	9
2011/5/14	18.57	2.57	0	1.07	2.17	43.28	355.42	5.81	42.1	333.87	11
2011/5/15	15.71	3.07	0	0.6	2.11	42.78	363.08	5.78	43.2	333.15	13
2011/5/16	15.05	4.24	0	1.28	1.15	42.4	311.26	6.67	50.1	331.29	15
2011/5/17	15.93	4.25	0	1.67	2.97	52.96	399.39	8.89	51.9	330.9	18
2011/5/18	15.7	4	0	1.71	3.33	50	307.6	9.04	34.8	333.5	17

4.2.2. Grey Relational Analysis

The grey relational degree of each factors were calculated according to GRA, as shown in Table 6 and the grey relational degree ranged from 0.43 to 0.82. It can be seen from Table 6 that

Table 6. Grey relational degree of each factor.

Factor	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11
A1	1.00	0.68	0.70	0.67	0.56	0.71	0.73	0.72	0.59	0.43	0.64
A2	0.68	1.00	0.60	0.80	0.68	0.81	0.66	0.82	0.61	0.54	0.66
A3	0.70	0.60	1.00	0.59	0.63	0.64	0.64	0.68	0.62	0.43	0.65
A4	0.67	0.80	0.59	1.00	0.70	0.79	0.68	0.82	0.65	0.54	0.68
A5	0.56	0.68	0.63	0.70	1.00	0.69	0.60	0.68	0.69	0.62	0.70
A6	0.71	0.81	0.64	0.79	0.69	1.00	0.67	0.80	0.62	0.50	0.64
A7	0.74	0.67	0.65	0.69	0.61	0.68	1.00	0.70	0.64	0.52	0.68
A8	0.72	0.82	0.68	0.82	0.68	0.80	0.70	1.00	0.60	0.50	0.69
A9	0.59	0.61	0.62	0.65	0.69	0.62	0.63	0.60	1.00	0.62	0.78
A10	0.43	0.54	0.43	0.54	0.62	0.50	0.51	0.50	0.62	1.00	0.58
A11	0.64	0.66	0.65	0.68	0.70	0.64	0.67	0.69	0.78	0.58	1.00

- (1) The grey relational degree of load and the gas concentration was between 0.43 and 0.62, indicating that there was a strong correlation between the load and gas concentration of CH₄, C₂H₂, C₂H₄ and CO₂, while the correlation between load and H₂, C₂H₆ was weak.
- (2) The grey relational degree between oil temperature and each gas concentration was between 0.58 and 0.69, which indicates that oil temperature had strong correlation with each gas concentration.
- (3) The grey relational degree of ambient temperature and gas concentration was between 0.64 and 0.70, indicating that there was also a strong correlation between ambient temperature and gas concentration.

- (4) The grey relational degree between oil temperature and ambient temperature was 0.78, which indicates that the ambient temperature was positively correlated with the oil temperature range. Lower ambient temperature was conducive to heat dissipation, resulting in lower oil temperature.
- (5) The grey relational degree of the total hydrocarbon and the hydrocarbon gases was between 0.68 and 0.82, which verified that there was a strong correlation between the hydrocarbon gases and total hydrocarbon while the total hydrocarbon concentration was the sum of the four gas concentrations.

4.2.3. Empirical Mode Decomposition

The EMD results of H₂, CH₄, C₂H₆, C₂H₂, C₂H₄, CO, CO₂ and total hydrocarbon are shown in the Figure 4, respectively. The original H₂ nonlinear sequence was decomposed into two less volatile time series components IMF (IMF1–IMF2) and one residual component shown in Figure 4a.

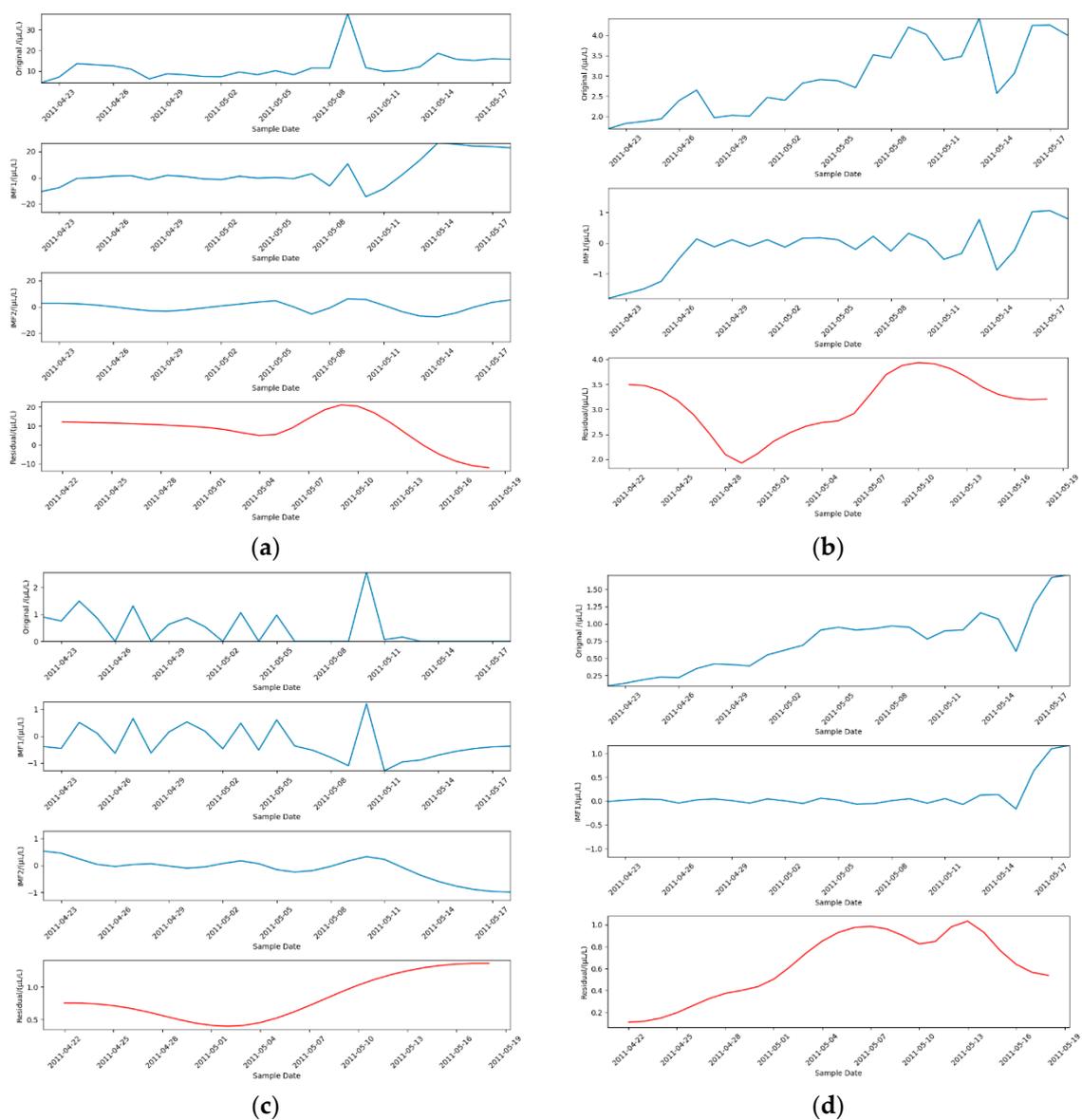


Figure 4. Cont.

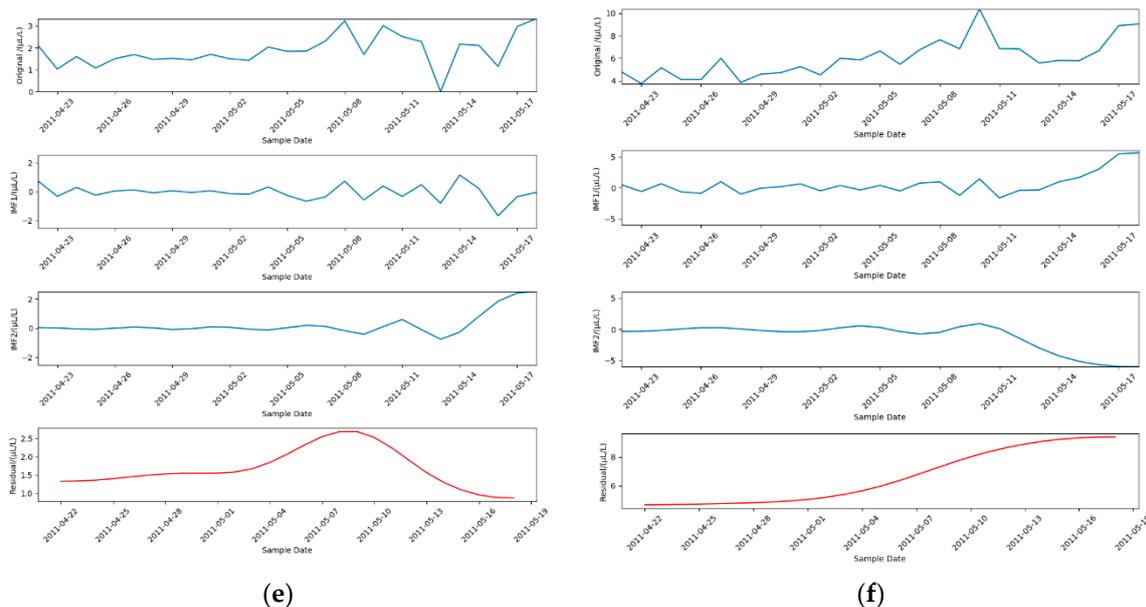


Figure 4. EMD results of the dissolved gas concentration. (a) H₂; (b) CH₄; (c) C₂H₆; (d) C₂H₂; (e) C₂H₄ and (f) total hydrocarbon.

4.2.4. Results and Discussion

The IMFs and the residual of each gas were predicted by the MGWO-LSSVM model, respectively. The data from June 7, 2010 to February 7, 2012 were used as the training set, and February 20, 2012 to July 17, 2012 as the test set. The prediction result of H₂ by the GRA-EMD-MGWO-LSSVM model is shown in Figure 5.

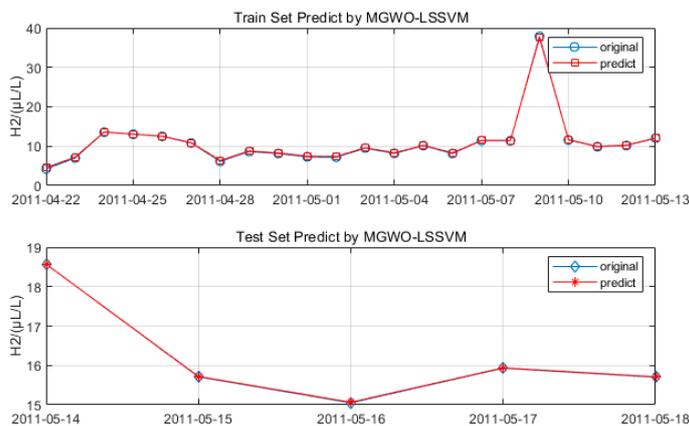


Figure 5. Prediction result of H₂ by the GRA-EMD-MGWO-LSSVM model. The.

The prediction results of different prediction models were compared with the proposed model in this paper, as shown in Tables 7 and 8.

Table 7. Comparison of different prediction models.

Model	η_{MAPE} (%)	η_{RMSE}	η_{max} (%)
LSSVM	13.65	15.58	20.6
PSO-LSSVM	10.39	12.49	18.06
GWO-LSSVM	9.62	10.76	17.02
MGWO-LSSVM	8.57	8.68	16.7

Table 8. Comparison of different prediction model.

Model	η_{MAPE} (%)	η_{RMSE}	η_{max} (%)
EMD-PSO-LSSVM	7.54	6.49	11.36
EMD-GWO-LSSVM	5.48	5.76	9.22
EMD-MGWO-LSSVM	3.63	4.56	8.75
GRA-EMD-PSO-LSSVM	4.38	3.76	6.26
GRA-EMD-GWO-LSSVM	2.68	3.23	4.02
GRA-EMD-MGWO-LSSVM	0.65	2.29	1.18

It can be seen from Tables 7 and 8 that:

- (1) The MGWO-LSSVM model had higher prediction accuracy, which verified the effectiveness of MGWO algorithm.
- (2) The prediction models with EMD method showed higher prediction accuracy, which was consistent with the results shown in Table 5.
- (3) The prediction models with GRA method also showed better performance, which was consistent with the results shown in Table 5.
- (4) Compared with other prediction models, the GRA-EMD-MGWO-LSSVM model proposed in this paper achieved higher prediction accuracy.

5. Conclusions

This paper presented a dissolved gas concentration prediction model based on GRA-EMD-MGWO-LSSVM for oil-immersed transformer oil. Firstly, the original time series data of dissolved gas concentration in the original oil were analyzed by the grey relational analysis method to extract the optimal feature set affecting gas concentration and the concentration data of each dissolved gas were decomposed by empirical mode decomposition. Then, the modified grey wolf optimized least squares support vector machine was used to predict the subsequences of each gas. Finally, the prediction sequences were reconstructed to obtain the final prediction results. The prediction model proposed in this paper was compared with LSSVM, PSO-LSSVM, GWO-LSSVM, MGWO-LSSVM, EMD-PSO-LSSVM, EMD-GWO-LSSVM, EMD-MGWO-LSSVM, GRA-EMD-PSO-LSSVM, and GRA-EMD-GWO-LSSVM by using DGA samples from two transformers. The main conclusions are listed as follows:

- (1) The modification strategy of GWO improved the performance of the original GWO, which promoted the accuracy of the prediction model.
- (2) The introduction of GRA and EMD into the prediction model could greatly improve the accuracy of the prediction model
- (3) The effect of transformer load, oil temperature and ambient temperature on the dissolved gas concentration was explored and these factors were considered in the model, which improved the performance of the prediction model.
- (4) The model proposed in this paper showed high prediction accuracy, stronger generalization ability and robustness.
- (5) The model proposed in this paper maintained high prediction accuracy and showed strong generalization ability for different DGA data samples.

Although the proposed model shows good performance in prediction, there were some differences in the prediction accuracy for different DGA data samples. For transformers in different operating states, oil temperature, load and ambient temperature had different effects on DGA data, which need more research in the future. In addition, the concentration of dissolved gas in oil is also related to the transformer operating life, operating state and other factors [31]. According to the further tests in the field, we know that for some transformers with long operation life, the dissolved gas concentration will increase with the increase of load due to equipment aging. Therefore, in future work, more running data will be collected and the relationship between dissolved gas concentration and the transformer operating life, operating state and other factors will be analyzed to develop a prediction model with stronger generalization ability and higher prediction accuracy, and further improve the practical ability of the prediction model.

Author Contributions: B.Z. collected DGA data and designed the algorithm with J.G., B.Z. and F.Z. test the example and write the manuscript. Z.X., W.Z., S.H. and P.F. helped design the algorithm and debug the code. All authors have approved the submitted manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the National Natural Science Foundation of China (51379160) and the State Grid Science and Technology Program of China.

Acknowledgments: The authors gratefully acknowledge the support of the National Natural Science Foundation of China (Grant.51379160), and the State Grid Science and Technology Program of China.

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

References

1. Wang, L.; Wang, Z.; Qu, H.; Liu, S. Optimal Forecast Combination Based on Neural Networks for Time Series Forecasting. *Appl. Soft Comput.* **2018**, *66*, 1–17. [[CrossRef](#)]
2. Wang, M.H.; Hung, C.P. Novel grey model for the prediction of trend of dissolved gases in oil-filled power apparatus. *Electr. Power Syst. Res.* **2003**, *67*, 53–58. [[CrossRef](#)]
3. Liao, R.J.; Zheng, H.B.; Grzybowski, S.; Yang, L.J.; Tang, C.; Zhang, Y.Y. Fuzzy information granulated particle swarm optimisation-support vector machine regression for the trend forecasting of dissolved gases in oil-filled transformers. *IET Electr. Power Appl.* **2011**, *5*, 230–237. [[CrossRef](#)]
4. Fei, S.W.; Wang, M.J.; Miao, Y.B.; Tu, J.; Liu, C.L. Particle swarm optimization-based support vector machine for forecasting dissolved gases content in power transformer oil. *Energy Convers. Manag.* **2009**, *50*, 1604–1609. [[CrossRef](#)]
5. Fei, S.; Sun, Y. Forecasting dissolved gases content in power transformer oil based on support vector machine with genetic algorithm. *Electr. Power Syst. Res.* **2008**, *78*, 507–514. [[CrossRef](#)]
6. Liao, R.J.; Bian, J.P.; Yang, L.J.; Grzybowski, S.; Wang, Y.Y.; Li, J. Forecasting dissolved gases content in power transformer oil based on weakening buffer operator and least square support vector machine–Markov. *IET Gener. Transm. Distrib.* **2012**, *6*, 142–151. [[CrossRef](#)]
7. Liao, R.; Zheng, H.; Grzybowski, S.; Yang, L. Particle swarm optimization-least squares support vector regression based forecasting model on dissolved gases in oil-filled power transformers. *Electr. Power Syst. Res.* **2011**, *81*, 2074–2080. [[CrossRef](#)]
8. Lu, S.X.; Lin, G.; Li, M.J.; Wei, C.H.; Wang, J.K. Grey relational analysis using Gaussian process regression method for dissolved gas concentration prediction. *Int. J. Mach. Learn. Cybern.* **2019**, *10*, 1313–1322. [[CrossRef](#)]
9. Lin, J.; Sheng, G.; Yan, Y.; Dai, J.; Jiang, X. Prediction of Dissolved Gas Concentrations in Transformer Oil Based on the KPCA-FFOA-GRNN Model. *Energies* **2018**, *11*, 225. [[CrossRef](#)]
10. Zheng, H.; Zhang, Y.; Liu, J.; Wei, H.; Zhao, J.; Liao, R. A novel model based on wavelet LS-SVM integrated improved PSO algorithm for forecasting of dissolved gas contents in power transformers. *Electr. Power Syst. Res.* **2018**, *155*, 196–205. [[CrossRef](#)]
11. Pereira, F.H.; Bezerra, F.E.; Junior, S.; Santos, J.; Chabu, I.; Souza, G.F.M.; Micerino, F.; Nabeta, S.I. Nonlinear Autoregressive Neural Network Models for Prediction of Transformer Oil-Dissolved Gas Concentrations. *Energies* **2018**, *11*, 1691. [[CrossRef](#)]

12. Lin, J.; Su, L.; Yan, Y.; Sheng, G.; Xie, D.; Jiang, X. Prediction Method for Power Transformer Running State Based on LSTM_DBN Network. *Energies* **2018**, *11*, 1880. [[CrossRef](#)]
13. Liu, C.; Zhang, H.; Xie, Z.; Wang, Z.; Chen, L.; Lin, X.; Ding, H.; Li, Z.; Xiong, W. Combined Forecasting Method of Dissolved Gases Concentration and Its Application in Condition-based Maintenance. *IEEE Trans. Power Deliv.* **2019**, *34*, 1269–1279. [[CrossRef](#)]
14. Peimankar, A.; Weddell, S.J.; Jalal, T.; Laphorn, A.C. Multi-Objective Ensemble Forecasting with an Application to Power Transformers. *Appl. Soft Comput.* **2018**, *68*, 233–248. [[CrossRef](#)]
15. Mirjalili, S.; Mirjalili, S.M.; Lewis, A. Grey Wolf Optimizer. *Adv. Eng. Softw.* **2014**, *69*, 46–61. [[CrossRef](#)]
16. Ibrahim, R.A.; Elaziz, M.A.; Lu, S. Chaotic opposition-based grey-wolf optimization algorithm based on differential evolution and disruption operator for global optimization. *Expert Syst. Appl.* **2018**, *108*, 1–27. [[CrossRef](#)]
17. Al-Betar, M.A.; Awadallah, M.A.; Faris, H.; Aljarah, I.; Hammouri, A.I. Natural selection methods for Grey Wolf Optimizer. *Expert Syst. Appl.* **2018**, *113*, 481–498. [[CrossRef](#)]
18. Singh, N.; Singh, S.B. A novel hybrid GWO-SCA approach for optimization problems. *Eng. Sci. Technol. Int. J.* **2017**, *20*, 1586–1601. [[CrossRef](#)]
19. Tzeng, G.H.; Tasur, S.H. The multiple criteria evaluation of grey relation model. *J. Grey Syst.* **1994**, *6*, 87–108.
20. Huang, N.E.; Shen, Z.; Long, S.R.; Wu, M.C.; Shih, H.H.; Zheng, Q.; Yen, N.C.; Tung, C.C.; Liu, H.H. The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. R. Soc. Soc. Lond.* **1998**, *454*, 903–995. [[CrossRef](#)]
21. Lima, E.R.D.; Andrade, A.O.; Pons, J.L.; Kyberd, P.; Nasuto, S.J. Empirical mode decomposition: A novel technique for the study of tremor time series. *Med. Biol. Eng. Comput.* **2006**, *44*, 569–582. [[CrossRef](#)] [[PubMed](#)]
22. Uygur, K.; Bayramoglu, I.; Nazikoglu, A.; Yilmaz, M.; Bayazit, Y.; Muftuoglu, S. Using empirical mode decomposition to correlate paleoclimatic time-series. *Nat. Hazards Earth Syst. Sci.* **2007**, *7*, 299–307.
23. Büyükşahin, Ü.Ç.; Ertekin, Ş. Improving forecasting accuracy of time series data using a new ARIMA-ANN hybrid method and empirical mode decomposition. *Neurocomputing* **2019**, *361*, 151–163. [[CrossRef](#)]
24. Wang, K.; Niu, D.; Sun, L.; Zhen, H.; Liu, J.; De, G.; Xu, X. Wind Power Short-Term Forecasting Hybrid Model Based on CEEMD-SE Method. *Processes* **2019**, *7*, 843. [[CrossRef](#)]
25. Cui, C.; Zhao, M.; Wong, K. An LSTM-Method-Based Availability Prediction for Optimized Offloading in Mobile Edges. *Sensors* **2019**, *19*, 4467. [[CrossRef](#)] [[PubMed](#)]
26. Padhy, S.; Panda, S.; Mahapatra, S. A modified GWO technique based cascade PI-PD controller for AGC of power systems in presence of Plug in Electric Vehicles. *Eng. Sci. Technol. Int. J.* **2017**, *20*, 427–442. [[CrossRef](#)]
27. Mittal, N.; Singh, U.; Sohi, B. Modified grey wolf optimizer for global engineering optimization. *Appl. Comput. Intell. Soft Comput.* **2016**, *2016*, 7950348. [[CrossRef](#)]
28. Vapnik, V.N. *The Nature of Statistical Learning Theory*; Springer: New York, NY, USA, 1995.
29. Domingos, S.D.; de Oliveira, J.F.; de Mattos Neto, P.S. An intelligent hybridization of ARIMA with machine learning models for time series forecasting. *Knowl. Based Syst.* **2019**, *175*, 72–86.
30. Fang, X.; Yuan, Z. Performance enhancing techniques for deep learning models in time series forecasting. *Eng. Appl. Artif. Intell.* **2019**, *85*, 533–542. [[CrossRef](#)]
31. De Faria, H., Jr.; Costa, J.G.S.; Olivas, J.L.M. A review of monitoring methods for predictive maintenance of electric power transformers based on dissolved gas analysis. *Renew. Sustain. Energy Rev.* **2015**, *46*, 201–209. [[CrossRef](#)]

