Article

An Interval Forecasting Model Based on Phase Space Reconstruction and Weighted Least Squares Support Vector Machine for Time Series of Dissolved Gas Content in Transformer Oil

Fang Yuan 1,2*, Jiang Guo 1,2*, Zhihuai Xiao 2, Bing Zeng 1,2*, Wenqiang Zhu 1,2 and Sixu Huang 1,2

1 Intelligent Power Equipment Technology Research Center, Wuhan University, Wuhan 430072, China; coderyuan@whu.edu.cn (F.Y.); zengbing_whu@whu.edu.cn (B.Z.); 200632650021@whu.edu.cn (W.Z.); huangsixu@whu.edu.cn (S.H.)
2 School of Power and Mechanical Engineering, Wuhan University, Wuhan 430072, China; xiaozhihuai@whu.edu.cn

* Correspondence: guo.river@whu.edu.cn; Tel.: +86-150-723-71659

Received: 27 January 2020; Accepted: 30 March 2020; Published: 3 April 2020

Abstract: Transformer state forecasting and fault forecasting are important for the stable operation of power equipment and the normal operation of power systems. Forecasting of the dissolved gas content in oil is widely conducted for transformer faults, but its accuracy is affected by data scale and data characteristics. Based on phase space reconstruction (PSR) and weighted least squares support vector machine (WLSSVM), a forecasting model of time series of dissolved gas content in transformer oil is proposed in this paper. The phase spaces of time series of the dissolved gas content sequence are reconstructed by chaos theory, and the delay time and dimension are obtained by the C-C method. The WLSSVM model is used to forecast time series of dissolved gas content, the chemical reaction optimization (CRO) algorithm is used to optimize training parameters, the bootstrap method is used to build forecasting intervals. Finally, the accuracy and generalization ability of the forecasting model are verified by the analysis of actual case and the comparison of different models.

Keywords: transformer; dissolved gas content in oil; forecasting; chaos theory; phase space reconstruction; weighted least-squares support vector machine; chemical reaction optimization; bootstrap

1. Introduction

A power transformer is among the key equipment of a power system. During the long operation of the transformer, due to equipment aging, discharge fault, thermal fault, and other reasons, a small amount of gas will be produced in the insulation oil, and the content of various components of dissolved gas and the proportion of components in the oil are closely related to the operation condition of the transformer. Through dissolved gas analysis (DGA) [1–5], some latent faults in the transformer and their development degree can be found. DGA is an internationally recognized and effective method of diagnosing early transformer faults that has been proven in practice by many fault diagnoses. According to the changing trend of dissolved gas content in the oil, the operation state of the transformer can be tracked, and any abnormality in the equipment can be further determined. Then, the fault type can be inferred, and the early faults in the transformer can be found in time. If the transformer is forecasted to fail, the maintenance plan can be arranged in advance to ensure reliable and stable operation of the power system.
How to establish a reasonable prediction model is the focus of oil dissolved gas prediction research. For a long time, improving the forecasting method has been regarded as the most important way to improve forecasting accuracy. Related scholars have done a lot of work. The existing methods of dissolved gas content forecasting in transformer oil can be summarized into three categories: methods based on statistics, intelligent models, and combined models. Forecasting methods based on statistics mainly include weighted average (WA), Kalman filter (KF), autoregressive moving average (ARMA), gray model (GM), etc. [6,7]. GM can reveal the law of development with a small amount of incomplete information. However, the model only describes the process of monotonic exponential increase or decrease with time, while data with large fluctuations will produce large deviations from the forecasting results, and the accuracy is mainly limited by the distribution of the sequence data itself. Intelligent forecasting methods are typically represented by artificial neural network (ANN) [8], recurrent neural network (RNN) with network structure of loop feedback [9], and long short-term memory (LSTM) [10]. These methods analyze and train a large amount of historical data to obtain forecasting models that can reflect the development trend of time series. However, data-driven models need a large amount of historical data to overcome the overfitting problem, which may lead to unacceptable performance in practical applications due to the limitation of key gas content data [11–14]. In order to improve the accuracy of model forecasting, scholars have used a variety of influencing factors as input parameters for correlation analysis and combined multiple models [15–17].

However, problems still exist in the actual transformer operation and maintenance process. First, the installation locations of transformers are scattered, data collected through the monitoring equipment are not centralized, and large and continuous time sequence data are difficult to obtain. At the same time, the time series data are not continuous with previous historical data after the transformer completes regular maintenance. Therefore, the data samples that can be obtained and used are usually small. Second, many factors affect dissolved gases in oil, such as operating environment, load, humidity in oil tank, oil temperature, and so on, which will directly or indirectly change the content of dissolved gases. Therefore, some relevant factors used in the forecasting model can give better results. However, as the core equipment of the power system, the number and scale of transformers is huge. When various online monitoring devices are installed on transformers, factors such as equipment safety and investment cost need to be considered according to different functions, types, and importance, so it is difficult to fully configure the different types of monitoring devices, and the associated parameters required by the forecasting model are often incomplete. As a result, it is difficult for forecasting models to work well because of the above problems.

Some studies show that time series of dissolved gases in oil are complex nonlinear dynamic systems closely related to the operation state of the transformer. According to chaos theory, time series of dissolved gases in oil satisfy the main characteristics of chaos, including internal randomness, fractal dimension, and universality [18]. By calculating the Lyapunov index of time series of dissolved gases in oil, it can be found that most have chaotic characteristics [19–21]. Therefore, the chaotic phase space reconstruction method can be used to reconstruct the nonlinear dynamic trajectories of crude time series of dissolved gases in oil, and on this basis, the feature space transformation can be carried out from the perspective of phase space to fully extract the time series characteristics.

Support vector machine (SVM) is based on the principle of minimizing structural risk, which is helpful solve the problem of regression prediction in the case of small samples. As an improvement of the SVM model, least squares SVM (LSSVM) has also been applied in many pattern recognition and regression problems [22–25], but LSSVM loses robustness.

Above all, a phase space reconstruction (PSR) chemical reaction optimization (CRO) weighted least squares SVM (WLSSVM) model combining chaos theory and weighted least squares support vector machine is proposed to forecast the time series of dissolved gases in power transformer oil. The phase space reconstruction method is applied to reconstruct the chaotic time series to fully explore the internal laws and characteristics contained in the historical data, and the delay time and dimension are obtained by the C-C method. WLSSVM is used for gas content forecasting, and the chemical
reaction optimization algorithm is used to optimize model parameters to ensure the accuracy of model forecasting in the case of small samples, bootstrap is used for forecasting interval constructing. Finally, the accuracy and generalization ability of the forecasting model are verified by comparing with different data samples and models.

The remainder of this paper is organized as follows: Section 2 summarizes the related work of forecasting technology for time series of dissolved gases in transformer oil. Section 3 introduces the basic principle of the bootstrap method and PSR-CRO-WLSSVM model. A transformer forecasting model based on bootstrap and PSR-CRO-WLSSVM is proposed in Section 4, and examples and effects of the model are given in Section 5. Finally, conclusions are drawn, and potential future work is discussed in Section 6.

2. Related Work

In recent years, with the rapid development of computer technology and artificial intelligence, scholars have introduced more machine learning methods to predict the time series of dissolved gases in oil, including gray model (GM) [26–28], artificial neural network (ANN) [29], support vector machine (SVM) [30], etc. The authors of [31] applied the gray method to forecast the gas content in transformer oil, and established the GM (1, 1) model based on historical data. During the modeling process, the non-equal interval data of the dissolved gas contents are processed mathematically and converted into equal interval data. Wang et al. [32] used a GM method that considered the development trend of dissolved gases in oil and had general forecasting performance. According to the results of thermal decomposition of insulating oil, the gas components produced by the pyrolysis reaction are related to each other. For example, when insulation oil overheats, the main gas components produced are methane and ethylene, and the produced gas has a strong correlation with the type of fault. At the same time, GMs forecasting results are related to the law of the data itself. When the data have a certain trend, they can achieve better forecasting performance; otherwise, it would be worse. Wang et al. [33] proposed an improved gray model to forecast the dissolved gas in transformer oil, and the model still achieved high forecasting accuracy in the case of small samples, which has certain reference and promotion significance. However, the gray forecasting model only describes the process of monotonic growth or decline with exponential law over time, and if there are large fluctuations and changes, the forecasting result of the gray model has a large deviation [6,7]. Fei [34] proposed a power transformer fault forecasting method combining rough set and gray theory. The model improved the three-ratio diagnosis decision table through the rough set method, and combined the fault diagnosis method with gas forecasting to improve the accuracy of fault forecasting. Some experts and scholars have introduced neural network (NN) into time series forecasting of dissolved gas content, which has been widely used. The authors of [35] proposed a deep belief network (DBN) approach to forecast transformer dissolved gas content. Pereira et al. [36] proposed a nonlinear autoregressive neural network model combined with discrete wavelet transform to forecast the content of dissolved gas in transformer oil, which showed better results compared with the current prediction model and commonly used time series technique.

Scholars have improved the forecasting accuracy of the model by introducing multifactor parameters. The authors of [37] proposed an optimal weighted combination forecasting model, which combined the four forecasting algorithms of gray theory, BP neural network, genetic algorithm, and Kalman forecasting algorithm to forecast the concentration and development trend of dissolved gases in oil. Grey relational analysis (GRA) was introduced into the forecasting model by [38,39]. Considering the correlation among the components, GRA was used for quantitative analysis of the correlation of input variables before forecasting, and factors with weak correlation were removed, then the above forecasting method was used for modeling and calculation. The sequence of dissolved gases in oil was analyzed by the GRA method [40]. The results showed that there was a certain coupling relationship between the seven common gases, so the gray multivariate forecasting model was established and further improved in later studies. These methods have done a lot of work and contributed to the
prediction of dissolved gas content in oil in terms of model optimization and data processing, and achieved good results, but ignored the sample size problem and the lack of related factors in the actual engineering application. Support vector machine (SVM) relies on the principle of minimizing structural risk, which considers both empirical risk and the complexity of the learning machine, and therefore it is good for solving small samples and optimal problems, and has good generalization ability [41,42], so it is an effective approach for solving forecasting problems [43]. Least squares support vector machine (LSSVM) was introduced in [44] as a reformulation of the standard SVM [45,46] that simplifies the model to a great extent by applying linear least squares criteria to the loss function instead of the traditional quadratic programming method. The simplicity and inherited advantages of SVM, such as its being based on the principle of minimizing structural risk and its kernel mapping, promote the application of LSSVM in many pattern recognition and regression problems [47–50]. Zheng et al. [51] introduced LSSVM to dissolved gas content forecasting and made forecasts for the five gases. However, while improving the standard SVM model, LSSVM has lost its robustness.

In terms of data preprocessing, some scholars decompose the original sequence based on the nonlinear and nonstationary characteristics of the time series. Zeng et al. [52] used the empirical mode decomposition (EMD) method to process the DGA data and decompose nonstationary signals into characteristic frequency function components of different frequencies, and modeled each subsequence component separately, reconstructing its prediction by superposition. As a result, the combined prediction results, which meet the accuracy requirements, were obtained. The method is simple, convenient, and easy to operate, but it is prone to modal aliasing problems [53–55]. Lin et al. [56] used the kernel principal component analysis (KPCA) method to select the characteristic parameters and used generalized regression neural network (GRNN) to forecast the gas concentration in transformer oil, which improved the accuracy compared with the model without pretreatment. According to the above research, we know that data preprocessing and feature extraction are necessary for forecasting of time series of dissolved gas content, because the signal contains a large amount of information effectively. Therefore, if more linear and nonlinear features are mined through the historical time series of dissolved gas content, it will be more helpful for the forecasting. Chaos theory has been applied in power load forecasting, wind speed forecasting, equipment fault diagnosis, and other fields, but it is seldom used in forecasting of dissolved gas content in transformer oil. Liu et al. [57] analyzed the chaotic characteristics of power load through phase space reconstruction to extract the effective information on power load, and applied the reconstructed results to the forecasting model. Sun et al. [58] applied phase space reconstruction technology to inverter fault diagnosis, in which the phase space reconstruction method was used to obtain current characteristic trajectories of inverters in various operating states. Zhang et al. [59] applied chaos theory to recognition of partial discharge of Geographic Information System (GIS) equipment, extracted the chaotic features of partial discharge signals with four typical defects as the feature quantity, and then applied these chaotic features to pattern recognition and obtained good recognition performance. Qi et al. [60] applied chaos theory to the selection of data length and collection period of dissolved gas content in oil monitoring equipment, and gave suggestions on the collection period, which was of great help to the life of monitoring equipment and effective use of storage space.

Above all, the current research on forecasting methods of time series of dissolved gas content focuses on the selection of artificial intelligence models and optimization methods, which directly or indirectly improve the effectiveness of power transformer fault forecasting. However, the problem of small sample size and the inability to obtain associated parameters have not been well solved. At the same time, the traditional point forecasting method has difficulty reflecting the uncertainty of forecasting results, and these problems limit the application of artificial intelligence algorithms or models in dissolved gas content forecasting. Therefore, a model based on bootstrap and PSR-CRO-WLSSVM is proposed to solve these problems.
3. Fundamentals of bootstrap and PSR-CRO-WLSSVM Model

3.1. Phase Space Reconstruction

A chaotic system is a nonlinear system under the action of certain evolution law. It is universal, and its long-term behavior is reflected in spatial distribution with certain regularity [61,62]. The components of a chaotic system interact with each other and develop together, so they show the phenomenon of partial information containing other components. Theoretically, due to the correlation between components, the analyzing of the time series of any variable can restore the basic dynamic characteristics of the system.

In the 1980s, Packard and Takens [63] proposed the theory of phase space reconstruction. Its main purpose is to restore time series of chaotic attractors, and its principle is to choose the appropriate delay time and embedding dimension. Using chaos theory with the phase space reconstruction method, low-dimensional time series can be mapped to high-dimensional phase space and keep the differential homeomorphism while using less information, extracting more abundant structural characteristics.

Phase space reconstruction is an effective method to analyze nonlinear time series. The basic idea is to take a time series as a component of a nonlinear dynamical system. The variation law of this component can be used to reconstruct the equivalent high-dimensional phase space of the dynamic system, and the time series can be projected into the variable point trajectories of the high-dimensional phase space. If there is a time series of dissolved gas content in oil \(\{x_i\}(i = 1, 2, \ldots, N)\), where \(N\) is the number of data points in the time series, then the time series set reconstructed by phase space can be expressed by Equation (1):

\[
\begin{bmatrix}
X_1 \\
X_2 \\
\vdots \\
X_M
\end{bmatrix} =
\begin{bmatrix}
x_1 & x_1+\tau & x_1+(m-1)\tau \\
x_2 & x_2+\tau & x_2+(m-1)\tau \\
\vdots & \vdots & \vdots \\
x_M & x_M+\tau & x_M+(m-1)\tau
\end{bmatrix}
\]

(1)

where, \(m\) is the embedding dimension of the time series, \(\tau\) is the delay time, \(M\) is the number of vectors that the time series embedded in the phase space and \(N = M + (m-1)\tau\).

For a certain time series, in theory the existence of an optimal \(\tau\) and \(m\), can neither be too big nor too small, otherwise they will be unable to accurately capture the dynamic characteristics of the original signal.

At present, there are mainly the following methods to obtain the delay time: autocorrelation, average displacement, complex autocorrelation, mutual information, and C-C method. The C-C method is a method that can not only keep its nonlinear characteristics, but also calculate a small amount of delay time. Besides, it is easy to operate and has strong anti-noise ability. In this paper, the optimal delay time is obtained by the C-C method. The detailed description is as follows:

Investigate a pair of phase points in phase space:

\[
X(i) = [x(i), x(i+\tau), \ldots, x(i+(m-1)\tau)]
\]

(2)

\[
X(j) = [x(j), x(j+\tau), \ldots, x(j+(m-1)\tau)]
\]

(3)

Assuming the distance between them is \(r_{ij}(m)\), it is obvious that \(r_{ij}(m)\) is a function of the embedding dimension \(m\) of phase space:

\[
r_{ij}(m) = ||X(i) - X(j)||
\]

(4)
Given a critical distance \( r \), the proportion of the logarithm of points whose distance is less than \( r \) in the logarithm of points is denoted as the associated integral:

\[
C(m, N, r, \tau) = \frac{2}{M(M-1)} \sum_{1 \leq i \leq j \leq M} H(r - r_{ij}(m))
\]

(5)

where \( r \) is the neighborhood radius, and \( H(\cdot) \) is the Heaviside function and is described by Equation (6):

\[
H(x) = \begin{cases} 
0 & x \leq 0 \\
1 & x > 0
\end{cases}
\]

(6)

The time series \([X_1, X_2, \ldots, X_N]\) is divided into \( \tau \) non-intersecting time series, whose length is \( \text{INT}(N/\tau) \), and \( \text{INT} \) is an integer. The statistics of each subsequence can be calculated by Equation (7):

\[
S(m, N, r, \tau) = \frac{1}{\tau} \sum_{l=1}^{\tau} \left[ C_l \left(m, \frac{N}{\tau} r, \tau\right) - C^m_l \left(m, \frac{N}{\tau} r, \tau\right) \right]
\]

(7)

where \( C_l \) is the correlation integral of the \( l \)-th subsequence.

For \( N \to \infty \), Equation (7) can be deformed to Equation (8):

\[
S(m, r, \tau) = \frac{1}{\tau} \sum_{l=1}^{\tau} \left[ C_l (m, r, \tau) - C^m_l (1, r, \tau) \right]
\]

(8)

The deviation is defined as Equation (9):

\[
\Delta S(m, \tau) = \max \left[ s(m, r_j, \tau) \right] - \min \left[ s(m, r_j, \tau) \right]
\]

(9)

\( \Delta S(m, \tau) \) measures the maximum deviation of \( \Delta S(m, \tau) \) from radius \( r \). When the maximum deviation \( \Delta S(m, \tau) \) is the minimum value, the points in the reconstructed phase space are closest to uniform distribution, the reconstructed dynamical system orbit is fully displayed in the phase space, and the time series correlation is closest to zero.

Therefore, the \( \Delta S(m, \tau) \sim \tau \) curve also reflects the autocorrelation of the original time series (parameter \( m \) is fixed). Since \( \Delta S(m, \tau) \) is always positive, the optimal delay time can be the time point corresponding to the first local minimum value of \( \Delta S(m, \tau) \).

According to the BDS statistical conclusion, when \( N \geq 3000 \), \( 2 \leq m \leq 5 \), and \( \sigma/2 \leq r \leq 2\sigma \), the \( S(m, r, \tau) \sim \tau \) curve better reflects the autocorrelation of the original time series, where \( \sigma \) is the standard deviation of the time series. Take \( m = 2, 3, 4, 5 \) \( r_i = i \times 0.5\sigma, i = 1, 2, 3, 4 \), \( \bar{S}(\tau) \) and \( \Delta \bar{S}(\tau) \) can be described by Equations (10) and (11):

\[
\bar{S}(\tau) = \frac{1}{16} \sum_{m=2}^{5} \sum_{i=1}^{4} S(m, r_i, \tau)
\]

(10)

\[
\Delta \bar{S}(\tau) = \frac{1}{4} \sum_{m=2}^{5} \Delta S(m, \tau)
\]

(11)

It can be seen from the above two expressions that \( \bar{S}(\tau), \Delta \bar{S}(\tau) \) reflect the autocorrelation characteristics of the original time series. Considering that \( \bar{S}(\tau) \) values can be positive or negative, \( \Delta \bar{S}(\tau) \) values are always positive; finding the first zero crossing of \( \bar{S}(\tau) \) or the first local minimum point of \( \Delta \bar{S}(\tau) \) is the optimal delay. The index \( S_{\text{cor}}(\tau) \) can be defined by Equation (12):

\[
S_{\text{cor}}(\tau) = \Delta \bar{S}(\tau) + \left| \bar{S}(\tau) \right|
\]

(12)
Looking for $S_{cor}(\tau)$ of the global minimum value of $\tau$ can get the best embedding window $\tau_w$, and according to $\tau_w = (m - 1) \tau$, embedded figures $m$ can be obtained.

3.2. Weighted Least Squares Support Vector Machine

3.2.1. Least Squares Support Vector Machine

Suppose there is a training set $\{x_k, y_k\}_{k=1}^N$, where $x_k \in \mathbb{R}^n$ is n-dimensional input data, and $y_k \in \mathbb{R}^m$ is output data. Therefore, in the original weight space, the prediction model can be considered as the following optimal problem in Equations (13) and (14):

$$
\min_{w, b, e} J(w, e) = \frac{1}{2} w^T w + \frac{1}{2} \gamma \sum_{k=1}^N e_k^2
$$

s.t. $y_k = w^T \phi(x_k) + b + e_k, k = 1, 2, \ldots, N$

where, $w \in \mathbb{R}^n$ is the weight vector of the original weight space, $e_k \in \mathbb{R}^n$ is an error variable, $\phi(\cdot) : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is a nonlinear mapping function that maps input spatial data to higher-dimensional (possibly infinite dimensional) feature spaces, $b \in \mathbb{R}^n$ is the offset value, and $\gamma > 0$ is the regularization parameter (also called penalty coefficient).

Thus, in the original weight space, there is the following nonlinear model:

$$
y(x) = w^T \phi(x) + b
$$

Lagrange multipliers are introduced for Equation (15); $a_k \in \mathbb{R}^n$, Lagrange function defined as Equation (16):

$$
L(w, b, e; a) = J(w, e) - \sum_{k=1}^N a_k \{w^T \phi(x_k) + b + e_k - y_k\}
$$

By taking partial derivatives of the variables in Equation (16) and sorting out and eliminating $w$ and $e$, the optimal problem is transformed into a linear system by Equation (17):

$$
\begin{bmatrix}
0 & 1^T \\
1 & \Omega + \frac{1}{\gamma} I
\end{bmatrix}
\begin{bmatrix}
b \\
a
\end{bmatrix}
= \begin{bmatrix}
o \\
y
\end{bmatrix}
$$

where, $y = [y_1, y_2, \ldots, y_N]^T$, $1 = [1, 1, \ldots, 1]^T$, $a = [a_1, a_2, \ldots, a_N]^T$, $\Omega \in \mathbb{R}^{N \times N}$, $\Omega_{ij} = \phi(x_i)^T \phi(x_j) = K(x_i, x_j)$, $i, j = 1, 2, \ldots, N$, and $K(\cdot, \cdot)$ is a kernel function satisfying Mercer’s theorem.

Commonly used kernel functions are linear kernel function $K(x_i, x_j) = x_i^T x_j$, polynomial kernel function $K(x_i, x_j) = (x_i^T x_j + 1)^d$, $d = 1, 2, \ldots$, radial basis function (RBF) kernel (Gaussian kernel function), $K(x_i, x_j) = \exp(-||x_i - x_j||^2/2\sigma^2)$, etc. The kernel function of RBF is used in this paper, $a$ and $b$ can be calculated by Equation (17), so the nonlinear prediction model of LSSVM can be obtained as Equation (18):

$$
y(x) = \sum_{k=1}^N a_k K(x, x_k) + b
$$

3.2.2. Weighted LSSVM

While improving the standard SVM model, LSSVM loses its robustness. As a result, therefore, the weight of all training data in the objective function is $\gamma$ and all samples play the same role in training, which is inconsistent with the actual situation.

For each piece of sample data, due to its position in the whole sample and the degree of influence by noise, its importance differs.
In order to regain robustness by treating different training data differently, weighted LSSVM is used [64]. Based on the LSSVM model, each error quantity $e_k = a_k / \gamma$ is given a different weight factor $v_k$, and the optimization problem is described as Equation (19):

$$
\min_{w^*, b^*, c^*} J(w^*, e^*) = \frac{1}{2} w^T w^* + \frac{1}{2\gamma} \sum_{k=1}^{N} v_k e_k^2 \quad \text{s.t.} \quad y_k = w^T \varphi(x_k) + b^* + e_k^*, k = 1, 2, \ldots, N
$$

The Lagrange function can be described as Equation (20):

$$
L(w^*, b^*, c^*, a^*; \gamma) = J(w^*, e^*) - \sum_{k=1}^{N} a_k^* \left[ w^T \varphi(x_k) + b^* + e_k^* - y_k \right]
$$

Similarly, according to the KKT condition, the system of linear Equation (21) can be obtained.

$$
\begin{bmatrix}
0 & 1^T \\
1 & \Omega + V_{\gamma}
\end{bmatrix}
\begin{bmatrix}
b^* \\
a^*
\end{bmatrix}
= \begin{bmatrix}
o \\
y
\end{bmatrix}
$$

where, $V_{\gamma} = diag \left( \frac{1}{\text{IQR}_1}, \ldots, \frac{1}{\text{IQR}_2} \right)$ is a diagonal matrix, and the weight $V_k$ is determined by the error variable $e_k = a_k / \gamma$.

$$
V_k = \begin{cases}
1 & \text{if } e_k / \hat{s} \leq c_1 \\
\frac{\hat{s} - e_k / \hat{s}}{e_k / \hat{s} - c_2} & \text{if } c_1 \leq e_k / \hat{s} \leq c_2 \\
10^{-4} & \text{otherwise}
\end{cases}
$$

where, the robust estimate is the standard deviation of the amount of error, $\hat{s} = \frac{\text{IQR}}{1.349}$, $\text{IQR}$ is the quartile spacing of error $e_k$, which is the difference between the 0.75$n$ value and the 0.25$n$ value if arranged in numerical order, and $\hat{s}$ measures how far $e_k$ deviates from the Gaussian. The constants $c_1$ and $c_2$ are generally 2.5 and 3 [65].

The specific steps of the weighted LSSVM algorithm are as follows:

Step 1: Given training data set $\{x_k, y_k\}_{k=1}^{N}$, find the optimal parameters (through the following chemical reaction optimization algorithm). For the optimal parameter, $e_k = a_k / \gamma$ is calculated by Equation (20).

Step 2: The robust estimate $\hat{s}$ is calculated based on the distribution of error $e_k$

Step 3: According to Equation (21), the corresponding weight $V_{\gamma}$ value is determined by $e_k$.

Step 4: According to Equation (21), $a^*$ and $b^*$ are solved, and the final nonlinear prediction model is given as Equation (23):

$$
y(x) = \sum_{k=i}^{N} a_k^* K(x, x_k + e^*)
$$

The LSSVM model solved by Equation (17) is the optimal solution under the assumption that error $e_k$ obeys Gaussian distribution, and WLSSVM corrects the deviation of $\epsilon$ in the case of non-Gaussian fractions by the weight defined in Equation (22), which makes the WLSSVM regression robust.

3.3. Chemical Reaction Optimization

Chemical reaction optimization (CRO) is a meta-heuristic algorithm proposed by Lam in 2010. Inspired by the interaction between molecules in chemical reactions to find the lowest potential energy in potential energy surface, the algorithm adopts four primary reactions and follows the law of energy conservation [66]. From the perspective of microscopic analysis of chemical reactions, we can see that at the initial stage of a chemical reaction, the state of the molecules in the container is unstable due to
excessive energy in the molecules. In order to reach a stable state, each molecule will be led to the lowest possible energy state by the collision between molecules and the chemical reaction after the collision. The result is the product of the chemical reaction, and the formation process is the process of gradual reduction of reaction potential energy. In simple terms, CRO is an optimization process to search for the minimum potential energy of the system [67–70].

The basic operation units involved in the CRO algorithm are composed of molecules (ω) and container walls (buffer), where the molecules possess both potential energy (PE) and kinetic energy (KE) as the container walls create the environment in which the reaction occurs. The molecular PE is the ultimate criterion for evaluating a chemical reaction and thus becomes the objective function of the question of interest while KE is a quantized value for determining whether the system can initiate a molecular reaction. In a chemical reaction, there are four basic reaction operators: single-molecule collision, single-molecule decomposition, intermolecular collision, and molecular synthesis.

Single-molecule collision is a process that changes the molecular KE and PE due to collisions between molecules. The energy change in the process can be described by Equation (24):

\[
KE_{\omega'} = (PE_\omega - PE_{\omega'} + KE_\omega) \times \delta
\]  

(24)

where \( \omega \) is the original molecule, \( \omega' \) is the new molecule after structural change, \( \delta \in [KELossRate, 1] \) is a random number, KELossRate is the upper limit (in percentage) of the monomolecular collision loss rate, a constant, and \( PE_\omega = f(\omega) \) is molecular potential energy, with \( f(\cdot) \) as the objective function of the problem considered. Monomolecular collision enables local searching in the problem space.

Single molecule decomposition is a reaction process in which a molecule collides with a wall, and breaks down or splits into two new molecules. The energy difference \( E_{\text{dec}} \) before and after the collision is passed to the two new molecules in a random manner. The energy change in the reaction process can be expressed by Equation (25):

\[
E_{\text{dec}} = (PE_\omega + KE_\omega + \delta_1 \times \delta_2 \times \text{buffer}) - (PE_{\omega'_1} + PE_{\omega'_2}),
\]

\[
KE_{\omega'_1} = E_{\text{dec}} \times \delta_3,
\]

\[
KE_{\omega'_2} = E_{\text{dec}} \times (1 - \delta_3)
\]

(25)

where \( \delta_1, \delta_2 \) are of uniform distribution in \( [0, 1] \) and \( \delta_3 \) is a random value in \( [0, 1] \). Compared to monomolecular collision, monomolecular decomposition is capable of local search in a larger scope.

Intermolecular collision describes the process of two new molecules after collision and energy exchange. A new molecule can be taken out of the original molecular structure domain. Since the molecule does not collide with the wall, there is no energy loss, so the total energy is unchanged after the reaction. The two new molecules have a total KE of \( E_{\text{inter}} \), which is distributed between them randomly, and the energy change can be described by Equation (26):

\[
E_{\text{inter}} = (PE_{\omega'_1} + PE_{\omega'_2} + KE_{\omega'_1} + KE_{\omega'_2}) - (PE_{\omega'_1} + PE_{\omega'_2}),
\]

\[
KE_{\omega'_1} = E_{\text{inter}} \times \delta_4,
\]

\[
KE_{\omega'_2} = E_{\text{inter}} \times (1 - \delta_4)
\]

(26)

where \( \delta_4 \) is a random value in the range of \( [0, 1] \).

Molecular synthesis is the phenomenon of two molecules colliding to produce a new molecule. It is a wall-free collision, so the energy remains constant before and after the collision. The energy change is shown in Equation (27):

\[
KE_{\omega'} = (PE_{\omega'_1} + PE_{\omega'_2} + KE_{\omega'_1} + KE_{\omega'_2}) - (PE_{\omega'})
\]

(27)
Molecular synthesis greatly increases the diversity of molecules, and the synthesized new molecules are obviously different from the originals, which usually have higher molecular activity. Molecular synthesis improves the searching ability of molecules in new regions, thus improving the global searching performance of CRO.

In this paper, the CRO algorithm is used to optimize the parameters of WLSSVM, and the optimal penalty coefficient $\gamma$ and kernel width $\sigma$ are obtained. The optimization process of CRO-WLSSVM is as follows:

Step 1: Initialize the chemical reaction optimization algorithm. It is necessary to determine the number of initial molecules in the container (PopSize), the upper limit (KELossRate) of the percentage of KE loss in the wall-hitting reaction, the determinants of molecular reaction type (MoleColl), the determinants of monomolecular reaction type ($\alpha$), the determinants of multi-molecular reaction type ($\beta$), the maximum iteration times (Iteration), etc.

Step 2: Calculate the initial potential energy of each molecule, and take the molecular KE initial value as the initial kinetic energy.

Step 3: Iteratively optimize the molecules in the container through four basic reaction operators. Only one basic reaction operator is executed in each iteration. The optimization process of each iteration consists of three judgment processes, which are reaction type, monomolecular reaction type, and intermolecular reaction type.

Step 4: Set the objective function. If the molecule meets the stop condition of the algorithm, the optimization calculation will be terminated. The smallest PE molecule is the global optimal solution, and the corresponding solution is the initial kernel width and penalty coefficient of the optimized WLSSVM, which can be assigned to WLSSVM to obtain the forecasting model of dissolved gas content.

3.4. Bootstrap

The bootstrap method is a statistical inference method of simulated sampling based on original data proposed by professor Bradley in 1979 [68] for statistical inference under small sample conditions. It belongs to one of the nonparametric estimation methods in statistics and can be used for statistical inference under small sample conditions.

Bootstrap is generally used for statistical inference when the model is difficult to obtain or difficult to assume. For example, when the data distribution is unknown, the bootstrap method does not need to make any distribution assumptions, and can indirectly obtain the distribution from the data generated by the original sample.

Therefore, the bootstrap method was used to resample the original test samples of dissolved gas in oil. Considering that the dissolved gas data in oil has time self-correlation, the resampling method of a single data point will destroy this dependent structure. However the block bootstrap method can preprocess the original time series data in blocks which are used as the basic unit for resampling to generate several new subsamples, thereby avoid the failure of the general self-service method to the greatest extent [69,70], and achieve a forecasting effect closer to the true distribution [71].

According to different block methods, nonoverlapping block bootstrap (NBB) method [72], moving block bootstrap (NBB) method [73], circular block bootstrap (CBB) method [74], and stationary bootstrap (SB) method [75] have been successively generated. Among them, the first three methods are relatively similar, they all use fixed-length block lengths and there are only slight differences in whether the subblocks overlap and whether the original sequence is processed end-to-end, meanwhile, compared with the simplest NBB method, MBB and CBB have higher estimation accuracy. The SB method uses the block length with geometric distribution and the randomized starting point of the block to ensure the stability of each subblock. However, some scholars have shown [76] that under the condition of similar block lengths, the variance estimated by the SB method is twice that of the NBB method, and three times that of the MBB method and the CBB method. Therefore, the MBB method is selected as the resampling method for the dissolved gas.
The resampling process of the MBB method is as follows:

Step 1: Assume that the original sample is \( X = \{ X_1, X_2, \ldots, X_N \} \) the original sample is divided into \( N - L + 1 \) partially overlapping blocks which can be expressed as \( Z_j = \{ X_{ji}, X_{ji+1}, \ldots, X_{ji+L-1} \} \), \( j = 1, 2, \ldots, N - L + 1 \), where the length of the block is \( L \), and the generated block set can be expressed as \( Z = \{ Z_1, Z_2, \ldots, Z_q \} \), \( q = N - L + 1 \). The selection of \( L \) will affect the estimation result. According to [77], \( L \) is related to the number of samples \( N \), and the empirical calculation method of \( L \) [78] is shown as Equation (28):

\[
L_{\text{thumb}} = \text{floor}(\sqrt[3]{N}) \tag{28}
\]

Step 2: Extract \( R \) blocks from \( Z \) with replacement, and construct a sample \( Z^* = \{ Z^*_1, Z^*_2, \ldots, Z^*_R \} \), where \( R = \text{floor}(N/L) \).

Step 3: Define the newly generated subsample as \( X^* = \{ X_1, X_2, \ldots, X_{N^*} \} = \{ Z^*_1, Z^*_2, \ldots, Z^*_R \} \), where the length of the subsample is \( N^* = R \times L \).

Step 4: Repeat steps 1 to 3 \( M \) times to construct \( M \) subsamples in sequence.

The correct choice of resampling number \( M \) is the key to ensure the accuracy of model prediction and the efficiency of equilibrium model calculation. A. Khosravi et al. [79] showed that there is no significant positive correlation between the size of the resampling number \( M \) and the width of the prediction interval, that is, an excessively large number of resampling \( M \) cannot significantly improve the quality of the prediction interval. E. Zio [80] believes that, in general, setting the resampling number \( M \) between 20 to 200 can meet the application requirements of most practical projects.

Based on the resampled subsamples, a prediction interval can be constructed. Many engineering examples have proved that the forecasting interval constructed by the bootstrap method can not only accurately represent the uncertainty degree of the forecasting results, but also analyze the range of future time series, which can effectively compensate for the limitation that the forecasting method of points can only obtain the deterministic forecasting value.

In view of the above advantages, bootstrap was introduced in this paper to construct the forecasting interval of dissolved gas in oil. In the process of forecasting, subjective factors such as the setting of forecasting model parameters and objective factors such as the noise of data collection are the main reasons affecting the uncertainty of forecasting results [81]. The bootstrap method takes into account the uncertainty caused by model error and data noise error, and constructs the forecasting interval based on the above factors.

The construction principle of forecasting interval is summarized as follows:

Suppose the sample set of dissolved gas in transformer oil is \( D_r = \{ x_i, t_i \}_{i=1}^{N} \): \( x_i \) is the input variable, \( t_i \) is the \( i \)th actual sample value, the error between results of forecasting model \( \hat{y}(x_i) \) and actual sampling value \( t_i \) can be described as Equation (29):

\[
t_i - \hat{y}(x_i) = [y(x_i) - \hat{y}(x_i)] + \varepsilon(x_i) \tag{29}
\]

where \( y(x_i) \) is the true regression value of the forecasting model, \( t_i - \hat{y}(x_i) \) is total error between actual values and forecasting values, \( (x_i) - \hat{y}(x_i) \) represents model error, and \( \varepsilon(x_i) \) represents data noise.

\( M \) subsamples were obtained by means of heavy sampling, and the subsamples were used for training in the corresponding forecasting model of dissolved gas.

The real value of output of forecasting model \( \hat{y}(x_i) \) is the average of all WLSSVM models forecasting results, which can be described by Equation (30):

\[
\hat{y}(x_i) = \frac{1}{M} \sum_{l=1}^{M} \hat{y}_l(x_i) \tag{30}
\]
where, $\hat{y}_l(x_i)$ is forecasting value obtained by the $l$th data set and the corresponding WLSSVM model, and $\sigma^2_{\hat{y}}(x_i)$ is the variance of the forecasting model error, which can be expressed by Equation (31):

$$\sigma^2_{\hat{y}}(x_i) = \frac{1}{M-1} \sum_{l=1}^{M} (\hat{y}_l(x_i) - \hat{y}(x_i))^2 \tag{31}$$

$\sigma^2_{\epsilon}(x_i)$ is the variance of data noise, which represents the uncertainty of samples caused by random noise in the measurement process, and can be described by Equation (32):

$$\sigma^2_{\epsilon}(x_i) \approx E\{ (t_i - \hat{y}(x_i))^2 \} - \sigma^2_{\hat{y}}(x_i) \tag{32}$$

In order to effectively forecast $\sigma^2_{\epsilon}(x_i)$, the square residual sequence of the model needs to be constructed and combined into a new sample set $D_{\text{new}} = \{x_i, r^2_i\}_{i=1}^{N}$. By training the new sample set with the corresponding model, the forecasting result of $\sigma^2_{\epsilon}(x_i)$ can be obtained. The square residual sequence $r^2_i$ can be obtained by Equation (33):

$$r^2_i = \max((t_i - \hat{y}(x_i))^2 - \sigma^2_{\hat{y}}(x_i), 0) \tag{33}$$

In order to maximize the probability of noise variance in the sample, the maximum likelihood estimation method is used to train the new model to ensure the confidence level of the forecasting [70]. The objective function is shown in Equation (34):

$$f_{\text{cost}} = \frac{1}{2} \sum_{i=1}^{N} \left[ \frac{r^2_i}{\sigma^2_{\hat{y}}(x_i)} + \ln(\sigma^2_{\hat{y}}(x_i)) \right] \tag{34}$$

After training of the model, the forecasting interval of dissolved gas in transformer oil with a significance level of $a$ can be constructed as Equation (35):

$$\hat{y}(x_i) \pm z_{1-a/2} \sqrt{\sigma^2_{\hat{y}}(x_i) + \sigma^2_{\epsilon}(x_i)} \tag{35}$$

where $z_{1-a/2}$ is the $1-a/2$ quantile in a standard Gaussian.

4. Transformer Forecasting Model Based on bootstrap and PSR-CRO-WLSSVM

4.1. Parameter Selection

The purpose of forecasting dissolved gases is to support fault warning and fault forecasting of transformers. Transformer fault forecasting usually combines the results of gas forecasting with fault diagnosis theory or fault diagnosis model to realize fault forecasting at a certain time in the future. Therefore, the selection of forecasting parameters depends on the parameters needed for transformer fault diagnosis [82].

During normal operation of the transformer, due to aging and cracking of insulation oil and solid insulation, a very small amount of gas will be decomposed, mainly including hydrogen ($H_2$), methane ($CH_4$), ethane ($C_2H_6$), ethylene ($C_2H_4$), acetylene ($C_2H_2$), carbon monoxide (CO), carbon dioxide (CO$_2$), etc. [2–4]. When a fault or abnormality occurs inside the transformer, the contents of some components in these gases will increase rapidly. For example, when the insulating oil is overheated, CH$_4$ and C$_2$H$_4$ are the main increased gas components and show a strong correlation. In the case of high energy discharge, the content of H$_2$ and C$_2$H$_2$ increases and shows a strong correlation. Based on the correlation between dissolved gas contents and variation and transformer faults, fault diagnosis methods or models recommended by International Electrotechnical Commission (IEC) and Institute of Electrical and Electronics Engineers (IEEE) have been produced, which are divided into two categories.
The input parameters of the IEC method, Roger method, and Doernenburg method are gas ratios and of the key gas method and David triangle method are gas content, and the input parameters of these fault diagnosis methods are all based on the above gas parameters.

Therefore, H₂, CH₄, C₂H₆, C₂H₄, C₂H₂, CO, and CO₂ were selected as the forecasting objects in this paper.

4.2. Evaluation Index

In order to verify the forecasting accuracy of the model, the mean absolute scale error (MASE), was used to evaluate the forecasting results of the proposed algorithm. The MASE was proposed by Hyndman and Koehler (2006) as a generally applicable measurement of forecast accuracy without the problems seen in the other measurements. The MASE can be used to compare forecast methods on a single series, and, because it is scale-free, to compare forecast accuracy across series [83,84].

\[
MASE = \frac{1}{n} \sum_{i=1}^{n} \frac{|y_i - \hat{y}_i|}{\frac{1}{n-2} \sum_{i=2}^{n} |y_{i-1} - y_{i-2}|}
\]

Prediction interval coverage probability (PICP), prediction intervals normalized averaged width (PINAW), and coverage width-based criterion (CWC) were used to evaluate the results of the forecasting interval [85,86].

a. PICP

PICP can be used represent the accuracy of interval forecasting and effectively avoid gas exceeding the forecasted upper and lower limits. The higher the value of PICP becomes, the more real values the predicted interval contains, and the more reliable the constructed interval will be.

\[
PICP = \frac{1}{N} \sum_{i=1}^{N} c_i
\]

\[
c_i = \begin{cases} 
1, & t_i \in [L_i, U_i] \\
0, & t_i \notin [L_i, U_i]
\end{cases}
\]

where \( N \) is the number of target values, \( t_i \) is the \( ith \) target value, and \( L_i \) and \( U_i \) are the lower bound and upper bound, respectively, of the forecasting interval corresponding to the \( ith \) target value.

b. PINAW

The quality of the forecasting interval is generally evaluated by PICP. If the target values are within the forecasting interval, the corresponding PICP will reach 100% coverage. The width of prediction intervals determines the amount of information they contain. In practice, it makes no sense to set a too wide prediction interval. PINAW is used to represent the degree of uncertainty of the prediction interval. The narrower the width, the lower the uncertainty of the interval and the better the accuracy of the model.

\[
PINAW = \frac{1}{R \times N} \sum_{i=1}^{N} (U_i - L_i)
\]

where \( R \) is the range of the target value and represents the difference between the maximum and minimum value of test sample of time series of the dissolved gas.

c. CWC

PICP and PINAW can only evaluate one aspect of the prediction interval, and these two indices are gain and loss indices, respectively. CWC is a comprehensive evaluation index, which contradicts
the two monotonicity evaluation indices. Therefore, CWC is considered to comprehensively evaluate the quality of the forecast interval and can be expressed by Equations (40) and (41):

\[
CWC = PINAW(1 + \gamma(PICP)e^{-\eta(PICP-\mu)})
\]  

\[
\gamma(PICP) = \begin{cases} 
0, & PICP \geq \mu \\
1, & PICP < \mu 
\end{cases}
\]  

where, \( \mu \) is the confidence interval and \( \eta \) is a super parameter, which is used to enlarge the difference between PICP and \( \mu \), and \( \eta \) can be assigned 30 [87].

4.3. Modeling Process of bootstrap and PSR-CRO-WLSSVM

The construction process of the bootstrap and PSR-CRO-WLSSVM interval forecasting model constructed in this paper is shown in Figure 1.

The algorithms involved in the model are summarized as follows: the bootstrap method was used to construct the data set and the prediction interval, the PSR method was used to transform the feature space of the sample data, and the CRO algorithm was used to optimize the penalty coefficient and the core width of WLSSVM, and the optimal parameters were obtained through iterative optimization of the four reaction operators. The samples were divided into training samples and test samples. The training samples were used for model training, to generate the global optimal interval forecasting model, and the test samples were used for verification, to obtain the forecasting results and ultimately point forecasting and interval forecasting of the dissolved gas in oil were realized finally.

Step 1: Normalize the data. The original DGA samples were normalized to convert the content of dissolved gas into the relative content within the range of [0,1], which is conducive to reducing the mutual exclusion between gases and avoiding the order of magnitude difference of input parameter values. The normalized treatment is shown in Equation (42):

\[
x'_{ij} = \frac{x_{ij}}{\sum_{j=1}^{k} x_{ij}}, i = 1, \ldots, n
\]  

Meanwhile, the original sample set was divided into a training set and test set, and M pseudo-sample sets were constructed by the bootstrap method.

Step 2: Construct and train the PSR-CRO-WLSSVM model, as shown in Figure 2.

PSR was used to preprocess the sample data, and the C-C method was used to calculate the optimal delay time \( t \) and embedded dimension \( m \) of the dissolved gas time series, as shown in Equations (9) and (10). Then the sample time series was reconstructed according to the embedded dimension and the delay time was obtained, as shown in Equation (12).

Then the phase space matrix was used as the training set and CRO was used as the optimization algorithm to train WLSSVM. Global optimization of kernel width and penalty factor is obtained by iterative optimization of the four reaction operators. The details of parameter optimization are introduced in Section 3.2.

In order to avoid model overfitting, the five-fold cross-validation method was adopted, and the predicted mean root mean square error under the five-fold cross-validation method was taken as the objective function of the model [6,38].
4.3. Modeling Process of bootstrap and PSR-CRO-WLSSVM

Figure 1. Interval forecasting process based on bootstrap and phase space reconstruction (PSR)-chemical reaction optimization (CRO)-weighted least squares support vector machine (WLSSVM).
The algorithms involved in the model are summarized as follows: the bootstrap method was used to construct the data set and the prediction interval, the PSR method was used to transform the feature space of the sample data, and the CRO algorithm was used to optimize the penalty coefficient and the core width of WLSSVM, and the optimal parameters were obtained through iterative optimization of the four reaction operators. The samples were divided into training samples and test samples. The training samples were used for model training, to generate the global optimal interval forecasting model, and the test samples were used for verification, to obtain the forecasting results and ultimately point forecasting and interval forecasting of the dissolved gas in oil were realized.

Step 1: Normalize the data. The original DGA samples were normalized to convert the content of dissolved gas into the relative content within the range of \([0,1]\), which is conducive to reducing the mutual exclusion between gases and avoiding the order of magnitude difference of input parameter values. The normalized treatment is shown in Equation (42):

\[
\tilde{x}_{ij} = \frac{x_{ij}}{\sum_{j=1}^{n} x_{ij}}, \quad i = 1, \ldots, M, \quad j = 1, \ldots, D
\]

Meanwhile, the original sample set was divided into a training set and test set, and M pseudo-sample sets were constructed by the bootstrap method.

Step 2: Construct and train the PSR-CRO-WLSSVM model, as shown in Figure 2.

![Figure 2. Construction process of PSR-CRO-WLSSVM model.](image)

Step 3: Calculate the variance of model error. The WLSSVM model was constructed with optimal super parameters, and the model was imported into M pseudo-sample sets for training. Equation (27) was used to construct the point forecasting results of the model with the expected value of the pseudo-sample set. According to Equation (28), the error variance sequence of the model was constructed.

Step 4: Calculate the variance of data noise error. Equation (29) was used to construct the variance sequence and variance data set of data noise. Equation (31) was used as the fitness function of noise variance, and a new forecasting model of noise variance was trained by the CRO algorithm.

Step 5: Construct the forecasting interval. According to Equation (32), the forecasting interval of dissolved gas in transformer oil was constructed.

Step 6: Evaluate the results. Finally, the forecasting model was used to forecast the test set, and the forecasting values were compared with actual values to analyze the accuracy and uncertainty of the model by the evaluation indices of point forecasting and interval forecasting.

5. Experimental study

5.1. Forecasting Examples

The training and test sample set used in the bootstrap and PSR-CRO-WLSSVM model were taken from the DGA data of a transformer with the voltage of 750 kV from a substation of the state grid of China, and the transformer model was ODFPS-700000/750GY. The time series of the data samples was from 23 May 2011 to 8 August 2012 and the monitoring period was once a day. Monitoring data from 23 May 2011 to 9 July 2012 were selected as training samples, and monitoring data from 10 July 2012 to
8 August 2012 were selected as test samples. Considering the similarity of each gas, the forecasting results of H2 are shown as an example, the special treatment of each gas is also described below. The original data of H2 is shown in Figure 3.

A test sample of H2 with a length of 415 from 23 May 2011 to 9 July 2012 was selected as the original sample, and based on the MBB resampling method, a subsample set was set, the number of subsamples $M$ was set to 100, and the block length $L$ was 7, the number of resampling $R$ was 59, and 100 subsamples with a length of 413 were formed. The construction method of the sample set is in Section 3.4.

Before model training, the data were normalized and transformed by PSR. The C-C method was used to calculate the delay time and embedding dimension of the reconstructed sample phase space. The results are shown in Figure 4. $\Delta S(\tau)$ of the first minimum point corresponding $\tau$ is 11, so the delay time is 11. Figure 5 shows that when $t_w = 20$, the $S_{cor}(\tau)$ is equal to the global minimum, because $t_w = (m - 1) \tau$, $m = 3$ can be obtained.

The calculation results of reconstruction parameters of different gases are shown in Table 1, where S1 to S3 represent $\Delta S(\tau)$, $S(\tau)$, $S_{cor}(\tau)$, respectively. The results of optimal values of $\tau_w$, $\tau$, and $m$ are summarized in Table 2. It is important to note that the phase space reconstruction of the optimal embedding dimension and delay time of dissolved gases are different, because although all gases are produced by the transformer oil, but the rate of change, tendency of change, influence factors and chaos characteristic are not the same, so the dissolved gases cannot use the same set of parameters for the phase space reconstruction, instead, they need to be calculated separately.
The preliminary analysis result of the case is as follows:

(1) The forecasting model proposed in this paper can effectively predict the gas change process and development trend of dissolved gas. The test samples selected in this paper were from 10 July 2012 to 8 August 2012, and the content of the H2 was in a state of great overall fluctuation, with the minimum and maximum values ranging from 36.32 to 48.41 μL/L. Figure 5 shows that the point forecasting results of the model for this time period are basically consistent with the actual gas change trend, which can effectively forecast the future change of dissolved gas in transformer oil.

(2) The forecasting interval constructed by the model proposed in this paper can effectively reflect the degree of uncertainty of the forecasting results. The gas trend and change in the two periods from 10 July 2012 to 15 July 2012 and 26 July 2012 to 29 July 2012 were relatively stable and in a slow and stable change process. The uncertainty of gas change was small, and the range width of the forecasting interval was generally small, with the minimum value of 0.88 μL/L and the average value of 1.17 μL/L. Since 24 July 2012 and 2 August 2012, the gas has been in a state of great fluctuation, and the gas value continues to rise rapidly and fluctuates repeatedly, this phenomenon indicates that in this time period, the uncertainty of gas change is large, because there might be some external disturbance, such as temperature mutation or internal change in harsh environment, partial discharge, or local high temperature. Meanwhile, the average width of the forecasting interval for these two periods was 1.94 μL/L, exceeding the overall average width of 1.63 μL/L and reaching a maximum of 2.64 μL/L. Therefore, the decision maker should be prompted to pay more attention to the transformer status.

It can be seen that the change process of interval width is related to the uncertainty of gas prediction and the uncertainty of the result can be analyzed according to the interval. When the interval width increases, it means that the forecasting value of the dissolved gas of the transformer oil indicates the risk, and the reason for the change should be analyzed in time as external interference, latent faults of the transformer, problems with the model itself, or other conditions. The correlation does not directly diagnose and locate faults, but it provides a new way of thinking for the early warning and analysis of transformer fault risk.

(3) The forecasting results of this model can be used as input of fault diagnosis models to support transformer fault forecasting.

---

**Figure 4.** Results of PSR for H2 time series samples from 23 May 2011 to 8 August 2012.

**Figure 5.** Interval forecasting result of time series of H2 from 10 July 2012 to 8 August 2012.
Table 1. Results of the C-C method for the dissolved gas samples from 23 May 2011 to 8 August 2012.

| Gas | Index | Delay time (τ) | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 20 |
|-----|-------|---------------|---|---|---|---|---|---|---|---|---|----|----|----|----|
| H₂  | S1    |               | 0.403 | 0.338 | 0.290 | 0.248 | 0.223 | 0.201 | 0.173 | 0.147 | 0.136 | 0.112 | 0.074 | 0.079 | 0.012 |
|     | S2    |               | 0.178 | 0.165 | 0.150 | 0.131 | 0.121 | 0.112 | 0.097 | 0.081 | 0.079 | 0.065 | 0.043 | 0.046 | 0.012 |
|     | S3    |               | 0.225 | 0.173 | 0.140 | 0.116 | 0.102 | 0.089 | 0.076 | 0.066 | 0.056 | 0.046 | 0.031 | 0.033 | 0.000 |

Table 2. The optimal value of τₓ, τ, and m for the dissolved gas samples from 23 May 2011 to 8 August 2012 after reconstruction.

| Gases | τ | τₓ | m |
|-------|---|----|---|
| H₂    | 11 | 20 | 3 |
| C₂H₂ | 5 | 65 | 12 |
| C₂H₆ | 6 | 37 | 8 |
| CH₄  | 5 | 59 | 13 |
| CO₂  | 6 | 32 | 7 |
| CO   | 12 | 34 | 4 |
| O₂   | 9 | 37 | 6 |

The reconstructed data set was input into the WLSSVM model for prediction, and the CRO optimization algorithm was used for optimization to obtain the optimal hyperparameters of the WLSSVM. The key parameter settings of the prediction model are shown in Table 3. The optimal parameter arrangement of WLSSVM is shown in Table 4, where D₁ represents the original training...
data set of H\textsubscript{2}. The results of its optimal parameters were used to forecast the M subsample sets. D\textsubscript{2} represents the squared residual set, and its optimal parameters were used for the forecasting of data noise in M + 1th model.

**Table 3.** Key parameters of PSR-CRO-WLSSVM forecasting model.

| Key Parameters         | Value       |
|------------------------|-------------|
| \(\tau\)               | 11          |
| \(m\)                  | 3           |
| Penalty coefficient \(\gamma\) | 0.1–100    |
| Kernel width \(\sigma\) | 0.01–30     |
| Kernel function        | RBF         |
| Epochs                 | 4000        |
| Initial number of molecules | 80         |
| Upper limit of KE loss | 0.3         |
| MoleColl               | 0.3         |
| \(\alpha\)             | 500         |
| \(\beta\)              | 15          |
| k fold cross validation| 5           |
| Iteration              | 5000        |

**Table 4.** Optimal parameters of the forecasting model for M subsample sets and one squared residual set.

| Parameters | \(D_1\) | \(D_2\) |
|------------|---------|---------|
| \(\gamma\) | 50.848  | 16.876  |
| \(\sigma\) | 0.0945  | 0.0313  |

For transformer condition monitoring, high reliability information is usually used for decision-making and analysis to ensure accurate judgment of the results. According to the principles of security and reliability, a 95% confidence level was selected for modeling. The results of point forecasting and interval forecasting of H\textsubscript{2} are shown in Figure 5.

The preliminary analysis result of the case is as follows:

1. The forecasting model proposed in this paper can effectively predict the gas change process and development trend of dissolved gas.

The test samples selected in this paper were from 10 July 2012 to 8 August 2012, and the content of the H\textsubscript{2} was in a state of great overall fluctuation, with the minimum and maximum values ranging from 36.32 to 48.41 \(\mu\)L/L. Figure 5 shows that the point forecasting results of the model for this time period are basically consistent with the actual gas change trend, which can effectively forecast the future change of dissolved gas in transformer oil.

2. The forecasting interval constructed by the model proposed in this paper can effectively reflect the degree of uncertainty of the forecasting results.

The gas trend and change in the two periods from 10 July 2012 to 15 July 2012 and 26 July 2012 to 29 July 2012 were relatively stable and in a slow and stable change process. The uncertainty of gas change was small, and the range width of the forecasting interval was generally small, with the minimum value of 0.88 \(\mu\)L/L and the average value of 1.17 \(\mu\)L/L. Since 24 July 2012 and 2 August 2012, the gas has been in a state of great fluctuation, and the gas value continues to rise rapidly and fluctuates repeatedly, this phenomenon indicates that in this time period, the uncertainty of gas change is large, because there might be some external disturbance, such as temperature mutation or internal change in harsh environment, partial discharge, or local high temperature. Meanwhile, the average width of the forecasting interval for these two periods was 1.94 \(\mu\)L/L, exceeding the overall average width of 1.63 \(\mu\)L/L and reaching a maximum of 2.64 \(\mu\)L/L. Therefore, the decision maker should be prompted to pay more attention to the transformer status.
It can be seen that the change process of interval width is related to the uncertainty of gas prediction and the uncertainty of the result can be analyzed according to the interval. When the interval width increases, it means that the forecasting value of the dissolved gas of the transformer oil indicates the risk, and the reason for the change should be analyzed in time as external interference, latent faults of the transformer, problems with the model itself, or other conditions. The correlation does not directly diagnose and locate faults, but it provides a new way of thinking for the early warning and analysis of transformer fault risk.

(3) The forecasting results of this model can be used as input of fault diagnosis models to support transformer fault forecasting.

The 30 sets of forecasted results of dissolved gas in oil were input into the transformer fault diagnosis model in [88] for fault diagnosis. As a result, the diagnosis results of 24 sets of data were low-energy discharge fault, and the other six sets of data were normal state, in which five of these six sets of normal data appear in the first eight sets of data at an earlier time, and this shows that the fault state of the transformer is still in the incubation period and the fault characteristics are not obvious.

According to the analysis report after the transformer is disassembled, it is found that the cause of the fault was the intermittent discharge caused by the looseness between the fasteners of the grounding system caused by vibration, which is consistent with the phenomenon reflected in the prediction of this paper.

5.2. Comparison Results of Different Models

In order to further test the accuracy and generalization performance of the forecasting model in this paper, BPNN, LSSVM, WLSSVM, PSO-WLSSVM [12], CRO-WLSSVM were selected for testing and comparing.

Back propagation neural network (BPNN) and LSSVM are classical and widely used machine learning methods, which can be used as the basis and reference for prediction accuracy and generalization ability indexes. The WLSSVM model was selected to verify the effect of the model on the performance improvement of LSSVM. The CRO-WLSSVM model was selected for comparison with PSO-CRO-WLSSVM to verify the performance of PSR, and the parameters of the model are shown in Table 3. The particle swarm optimization (PSO)-WLSSVM was selected to compared with CRO-WLSSVM to verify the performance of CRO, in which the population is 40, the maximum speed is 1.0, the learning factors $C_1$ and $C_2$ are 1.5, and the number of iterations is 5000.

The above six models were constructed by bootstrap method and trained by five-fold cross validation method. Data samples were selected from H$_2$ samples in Section 5.1 for forecasting analysis, and corresponding point forecasting and interval forecasting results were obtained as shown below.

The comparison results of the performance indicators predicted by the six models are shown in Figure 7. The MASE of the model in this paper is 1.861, which is lower than the other five models, indicating that the model in this paper has better prediction accuracy. For further analysis and comparison, Figure 6 shows the predicted results of six models. For example, at two time points of 24 July 2012 and 2 August 2012, when the dissolved gas content of H$_2$ showed a large fluctuation, compared with other models, the model in this paper could change the trend rapidly and maintain a small degree of deviation, which verified that the model in this paper could effectively predict the dynamic change process of dissolved gas.
optimal number of iterations is 1125, training time of PSO is 4817 ms and the optimal number of iterations is 1407. The performance of CRO is better on the convergence speed and convergence accuracy.

According to the comparison results between the model in this paper and the CRO-WLSSVM model in point forecasting and interval forecasting results, the model in this paper improved by 18.2% and 39% on two indices of MASE and CWC, which indicates that for the time series of dissolved gas in oil with chaotic characteristics, the phase space reconstruction method can effectively extract features to improve the prediction performance of the model.

Comparing the six models as a whole, in terms of point forecasting, it can be seen from Figure 6 that all the six models can accurately forecast the variation trend of dissolved gas in transformer oil.
and according to Figure 7, most of the models have good performance on MASE. In terms of interval forecasting, Figure 8 shows that the interval coverage of the six models all reached more than 90%, which was close to the performance requirements of the forecasting interval and among them, the model in this paper reached more than 95%, which met with the performance requirements of the forecasting interval. This shows that the model in this paper is consistent with the five comparison models in the prediction of dissolved gas in oil. At the same time, in the aspect of point forecasting and interval forecasting, the model presented in this paper has the best performance in the five indexes such as MASE, CWC, etc., which indicates that the model presented in this paper has certain advantages for the forecasting of dissolved gas with small samples.

![Figure 7](image_url)  
**Figure 7.** Mean absolute scale error (MASE) value for point forecasts of six models based on time series of H2 from 10 July 2012 to 8 August 2012.

![Figure 8](image_url)  
**Figure 8.** Value of prediction interval coverage probability (PICP), prediction intervals normalized averaged width (PINAW), and coverage width-based criterion (CWC) for forecasting intervals of six models based on time series of H2 from 10 July 2012 to 8 August 2012.

### 6. Conclusions

In order to overcome the problem that the data volume of transformer samples is small, and the uncertainty of prediction results cannot be represented by the traditional point forecasting models, a forecasting model of dissolved gas content in transformer oil based on bootstrap and PSR-CRO-WLSSVM was proposed and can be summarized as follows:
(1) WLSSVM is used as a forecasting model to predict small samples, and PSR method is introduced into the forecasting of oil-dissolved gas of transformer. The PSR method based on chaos theory considers the autocorrelation of gas time, fully excavates the inherent laws and characteristics contained in historical data, and realizes the preprocessing and feature extraction of gas data. Meanwhile, the global search advantage of CRO is used to optimize the forecasting model. The results show that the above method can effectively help the WLSSVM model to improve the forecasting accuracy of dissolved gas in oil.

(2) By combining the bootstrap method with the PSR-CRO-WLLSVM, a model for both point forecasting and interval forecasting was constructed. This method considers the data noise error and model error, which can describe the accuracy of the forecasting and the uncertainty of the forecasting. Compared with BPNN and LSSVM in the aspect of point forecasting and interval forecasting, the model presented in this paper has the best performance in five indexes such as MASE, CWC, etc.

(3) The actual case analysis proves that by combining the results of point forecasting and interval forecasting, the model in this paper can closely follow the change trend of dissolved gas, and discover potential risks through the change of uncertainty of interval. At the same time, the output result of the model can be used as the input parameter of fault diagnosis method for real-time fault forecasting, which provides more comprehensive decision support for the development trend, hidden risks, and fault analysis of dissolved gas.

There are still shortcomings and new opportunities in the current research. Our future work will focus on further study on the application of the forecasting interval in transformer fault forecasting. Meanwhile, the influence of the selection of the oil chromatographic data acquisition interval and the selection of prediction data length on the forecasting accuracy will also be the next research focus.

**Author Contributions:** Data curation, W.Z. and S.H.; Methodology, J.G. and S.H.; Resources, W.Z.; Software, F.Y. and Z.X.; Validation, J.G. and Z.X.; Writing—original draft, F.Y.; Writing—review and editing, J.G., B.Z., and W.Z. 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 (51979204) 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. 51979204), and the State Grid Science and Technology Program of China.

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

**References**

1. Singh, S.; Bandyopadhyay, M.N. Dissolved gas analysis technique for incipient fault diagnosis in power transformers: A bibliographic survey. IEEE Electr. Insul. Mag. 2010, 26, 41–46. [CrossRef]
2. Cruz, V.G.M.; Costa, A.L.H.; Paredes, M.L.L. Development and evaluation of a new DGA diagnostic method based on thermodynamics fundamentals. IEEE Trans. Dielectr. Electr. Insul. 2015, 22, 888–894. [CrossRef]
3. Gouda, O.S.; El-Hoshy, S.H.; El-Tamaly, H.H. Proposed heptagon graph for DGA interpretation of oil transformers. IET Gener. Transm. Distrib. 2018, 12, 490–498. [CrossRef]
4. Ghoneim, S.S.M. Intelligent Prediction of Transformer Faults and Severities Based on Dissolved Gas Analysis Integrated with Thermodynamics Theory. IET Sci. Meas. Technol. 2018, 12, 388–394. [CrossRef]
5. Souahlia, S.; Bacha, K.; Chaari, A. SVM-based decision for power transformers fault diagnosis using Rogers and Doernenburg ratios DGA. In Proceedings of the 10th International Multi-Conferences on Systems, Signals & Devices 2013 (SSD13), Hammamet, Tunisia, 18–21 March 2013; pp. 1–6.
6. Deng, J. Introduction to grey system. J. Grey Syst. 1989, 1, 1–24.
7. Leung, M.T.; Chen, A.S.; Daouk, H. Forecasting exchange rates using general regression neural networks. Comput. Oper. Res. 2000, 27, 1093–1110. [CrossRef]
8. Shaban, K.; El-Hag, A.; Matveev, A. A cascade of artificial neural networks to predict transformers oil parameters. IEEE Trans. Dielectr. Electr. Insul. 2009, 16, 516–523. [CrossRef]
9. Guardado, J.L.; Naredo, J.L.; Moreno, P.; Fuerte, C.R. A comparative study of neural network efficiency in power transformers diagnosis using dissolved gas analysis. *IEEE Trans. Power Deliv.* 2001, 16, 643–647. [CrossRef]

10. Hippert, H.S.; Pedreira, C.E.; Souza, R.C. Neural networks for short-term load forecasting: A review and evaluation. *IEEE Trans. Power Syst.* 2001, 16, 44–55. [CrossRef]

11. Liang, Z.; Wang, L.; Fu, D. Electric power system short-term load forecasting using Lyapunov exponents technique. *Proc. CSEE* 1998, 18, 368–371.

12. 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]

13. Jardine, A.K.S.; Lin, D.; Banjevic, D. A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mech. Syst. Signal Process.* 2006, 20, 1483–1510. [CrossRef]

14. Ganyun, L.; Haozhong, C.; Haibao, Z.; Lixin, D. Fault diagnosis of power transformer based on multi-layer SVM classifier. *Electr. Power Syst. Res.* 2005, 74, 1–7. [CrossRef]

15. Khemchandani, R.; Chandra, S. Twin support vector machines for pattern classification. *IEEE Trans. Pattern Anal.* 2007, 29, 905–910.

16. Mohamed, E.; Abdelaziz, A.; Mostafa, A. A neural network-based scheme for fault diagnosis of power transformers. *Electr. Power Syst. Res.* 2005, 75, 29–39. [CrossRef]

17. Shah, A.M.; Bhalja, B.R. Fault discrimination scheme for power transformer using random forest technique. *IET Gener. Transm. Distrib.* 2015, 10, 1431–1439. [CrossRef]

18. Zhang, X. Research of Chaos Synchronization and Its Application in Communication; Harbin Engineering University: Harbin, China, 2002.

19. Li, T.; Liu, Z. The chaotic property of power load and its forecasting. *Proc. CSEE* 2000, 20, 36–40.

20. Mori, H.; Urano, S. Short-term load forecasting with chaos time series analysis. In Proceedings of the 1996 International Conference on Intelligent Systems Applications to Power Systems, Orlando, FL, USA, 28 January–2 February 1996; pp. 133–137.

21. De Kruif, B.J.; De Vries, T.J.A. Pruning error minimization in least squares support vector machines. *IEEE Trans. Neural Netw.* 2003, 14, 696–702. [CrossRef]

22. 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]

23. Tay, F.E.H.; Cao, L.J. Application of support vector machines in financial time series forecasting. *Omega-Int. J. Manag. Sci.* 2001, 29, 309–317. [CrossRef]

24. Wu, Q. The forecasting model based on wavelet v-support vector machine. *Expert Syst. Appl.* 2009, 36, 7604–7610. [CrossRef]

25. Witzak, M. Modelling and Estimation Strategies for Fault Diagnosis of Non-Linear Systems: From Analytical to Soft Computing Approaches (Lecture Notes in Control and Information Sciences); Springer: New York, NY, USA, 2007.

26. Zheng, R.; Zhao, J.; Zhao, T. Prediction of power transformer oil dissolved gas concentration based on modified gray model. In Proceedings of the 2010 IEEE International Conference on Electrical and Control Engineering (ICECE), Wuhan, China, 25–27 June 2010; pp. 1499–1502.

27. Zheng, R.; Zhao, J.; Wu, B. Transformer oil dissolved gas concentration prediction based on genetic algorithm and improved gray verhulst model. In Proceedings of the 2009 AIC’09 International Conference on Artificial Intelligence and Computational Intelligence, Shanghai, China, 7–8 November 2009; Volume 4, pp. 575–579.

28. Zhao, W.; Zhu, Y. A prediction model for dissolved gas in transformer oil based on improved verhulst grey theory. In Proceedings of the 2008 3rd IEEE Conference on Industrial Electronics and Applications (ICIEA), Singapore, 3–5 June 2008; pp. 2042–2044.

29. Shaban, K.B.; El-Hag, A.H. Benhmed, K. Prediction of Transformer Furan Levels. *IEEE Trans. Power Deliv.* 2016, 31, 1778–1779. [CrossRef]

30. Zhang, S.; Bai, Y.; Wu, G.; Yao, Q. The forecasting model for time series of transformer DGA data based on WNN-GNN-SVM combined algorithm. In Proceedings of the 2017 1st International Conference on Electrical Materials and Power Equipment (ICEMPE), Xi’an, China, 14–17 May 2017; pp. 292–295.
31. Bin, S.; Ping, Y.; Yunbai, L.; Xishan, W. Study on the fault diagnosis of transformer based on the grey relational analysis. In Proceedings of the International Conference on Power System Technology, Kunming, China, 13–7 October 2002; pp. 2231–2234.

32. Wang, M.H. Grey-extension method for incipient fault forecasting of oil-immersed power transformer. *Electr. Power Compon. Syst.* 2004, 32, 950–975. [CrossRef]

33. 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]

34. Fei, S.W.; 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]

35. Dai, J.; Song, H.; Yang, Y.; Chen, Y.; Sheng, G.; Jiang, X. Concentration Prediction of Dissolved Gases in Transformer Oil Based on Deep Belief Networks. *Power Syst. Technol.* 2017, 41, 2737–2742.

36. 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]

37. Yang, T.-F.; Liu, P.; Li, Z.; Zeng, X.-J. A New Combination Forecasting Model for Concentration Prediction of Dissolved Gases in Transformer Oil. *Proc. CSEE* 2008, 28, 108–113.

38. Xiao, Y.; Zhu, H.; Chen, X. Concentration prediction of dissolved gas-in-oil of a power transformer with the multivariable grey model. *Autom. Electr. Power Syst.* 2006, 30, 64–67.

39. Sima, L.; Shu, N.; Zuo, J. Concentration prediction of dissolved gases in transformer oil based on grey relational analysis and fuzzy support vector machines. *Power Syst. Prot. Control* 2012, 40, 41–46.

40. Lin, X.; Huang, J.; Xiong, W. Interval prediction of dissolved-gas concentration in transformer oil. *Electr. Power Autom. Equip.* 2016, 36, 73–77.

41. Ruppert, D.; Wand, M.P.; Carroll, R.J. *Semiparametric Regression*; Cambridge University Press: Cambridge, UK, 2003.

42. Pai, P.F.; Hong, W.C. Forecasting regional electricity load based on recurrent support vector machines with genetic algorithms. *Electr. Power Syst. Res.* 2005, 74, 417–425. [CrossRef]

43. Yang, Z.; Gu, X.S.; Liang, X.Y.; Ling, L.C. Genetic algorithm-least squares support vector regression based predicting and optimizing model on carbon fiber composite integrated conductivity. *Mater. Des.* 2010, 31, 1042–1049. [CrossRef]

44. Vapnik, V.N.; Vapnik, V. *Statistical Learning Theory*; Wiley: New York, NY, USA, 1998.

45. Vapnik, V. *The Nature of Statistical Learning Theory*; Springer: Berlin/Heidelberg, Germany, 2013.

46. Suykens, J.A.; van Gestel, T.; de Brabanter, J. *Least Squares Support Vector Machines*; World Scientific: Singapore, 2002.

47. van Gestel, T.; Suykens, J.A.; Baestaens, D.-E.; Lambrechts, A.; Lanckriet, G.; Vandaele, B.; de Moor, B.; Vandewalle, J. Financial time series prediction using least squares support vector machines within the evidence framework. *IEEE Trans. Neural Netw.* 2001, 12, 809–821. [CrossRef] [PubMed]

48. van Gestel, T.; Suykens, J.A.; Baesens, B.; Viaene, S.; Vanthienen, J.; Dedene, G.; de Moor, B.; Vandewalle, J. Benchmarking least squares support vector machine classifiers. *Mach. Learn.* 2004, 54, 5–32. [CrossRef]

49. Zhang, Y.; Liu, Y. Traffic forecasting using least squares support vector machines. *Transp. Metr.* 2009, 5, 193–213. [CrossRef]

50. Zendehboudi, A. Implementation of GA-LSSVM modelling approach for estimating the performance of solid desiccant wheels. *Energy Convers. Manag.* 2016, 127, 245–255. [CrossRef]

51. Zhang, X.; Wang, J.; Zhang, K. Short-term electric load forecasting based on singular spectrum analysis and support vector machine optimized by Cuckoo search algorithm. *Electr. Power Syst. Res.* 2017, 146, 270–285. [CrossRef]

52. Zeng, B.; Guo, J.; Zhang, F.; Zhu, W.; Xiao, Z.; Huang, S.; 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]

53. Zhang, M.; Wang, C.; Cao, Q. *Improved EEMD on the Application Research of Signal Trend Analysis*; Trans Tech Publications Ltd.: Stafa-Zurich, Switzerland, 2012; pp. 2020–2023.

54. Wang, X.; Meng, L. Ultra-short-term load forecasting based on EEMD-LSSVM. *Power Syst. Prot. Control* 2015, 1, 61–66.
55. Mao, M.; Gong, W.; Zhang, L. Short-term photovoltaic generation forecasting based on EEMD-SVM combined method. Proc. CSEE 2013, 33, 17–24.
56. 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]
57. Liu, C.W.; Thorp, J.S.; Lu. J. Detection of transiently chaotic swings in power systems using real-time phasor measurements. IEEE Trans. Power Syst. 1994, 9, 1285–1292.
58. Sun, D.; Meng, J.; Guan, Y. Inverter faults diagnosis in PMSM DTC drive using reconstructive phase space and fuzzy clustering. Proc. CSEE 2007, 27, 49–53.
59. Zhang, X.; Xiao, S.; Shu, N. GIS partial discharge pattern recognition based on the chaos theory. IEEE Trans. Dielectr. Electr. Insul. 2014, 21, 783–790. [CrossRef]
60. Qi, B.; Zhang, P.; Rong, Z.; Li, C.; Yang, Y.; Chen, Y. Optimal Length Selection Method of DGA Data Based on Phase Space Reconstruction. Proc. CSEE 2018, 38, 2504–2511.
61. Kim, H.S.; Eykholt, R.; Salas, J.D. Nonlinear dynamics, delay times, and embedding windows. Phys. D Nonlinear Phenom. 1999, 127, 48–60. [CrossRef]
62. Durlauf, S.N. Nonlinear dynamics, chaos, and instability—Statistical-theory and economic evidence. J. Econ. Lit. 1993, 31, 232–234.
63. Takens, F. Detecting strange attractors in turbulence. In Dynamical Systems and Turbulence, Warwick 1980; Rand, D., Young, L.S., Eds.; Springer: Berlin/Heidelberg, Germany, 1981; Volume 898, pp. 366–381.
64. Suykens, J.A.K.; de Brabanter, J.; Lukas, L.; Vandewalle, J. Weighted least squares support vector machines: Robustness and sparse approximation. Neurocomputing 2002, 48, 85–105. [CrossRef]
65. Fan, Y.-G.; Li, P.; Song, Z. Dynamic weighted least squares support vector machines. Control Decis. 2006, 21, 1129–1133.
66. Bechikh, S.; Chaabani, A.; Said, L.B. An efficient chemical reaction optimization algorithm for multiobjective optimization. IEEE Trans. Cybern. 2015, 45, 2051–2064. [CrossRef]
67. Eldos, T.; Khreishah, A. Maximally distant codes allocation using chemical reaction optimization with enhanced exploration. Int. J. Adv. Comput. Sci. Appl. 2016, 7, 235–243. [CrossRef]
68. Wan, C.; Xu, Z.; Pinson, P.; Dong, Z.Y.; Wong, K.P. Probabilistic Forecasting of Wind Power Generation Using Extreme Learning Machine. IEEE Trans. Sustain. Energy 2014, 29, 1033–1044. [CrossRef]
69. Beran, R. Discussion of “Jackknife Bootstrap and Other Resampling Methods in Regression analysis”. Ann. Stat. 1986, 14, 1295–1298. [CrossRef]
70. Xie, Y.; Zhu, Y. Bootstrap method: Developments and frontiers. Stat. Inf. Forum 2008, 23, 91–96.
71. Khosravi, A.; Nahavandi, S.; Creighton, D. Prediction intervals for short-term wind power generation forecasts. IEEE Trans. Sustain. Energy 2013, 4, 602–610. [CrossRef]
72. Carlstein, E. The use of subseries methods for estimating the variance of a general statistic from stationary time series. Ann. Stat. 1986, 14, 1171–1179. [CrossRef]
73. Künsch, H.R. The jackknife and the bootstrap for general stationary observations. Ann. Stat. 1992, 17, 1217–1261. [CrossRef]
74. Politis, D.; Roman, J.P. A Circular Block Resampling Procedure for Stationary Data; Wiley: New York, NY, USA, 1993.
75. Politis, D.; Roman, J.P. The stationary bootstrap. J. Am. Stat. Assoc. 1994, 89, 1303–1313. [CrossRef]
76. Clements, M.P.; Kim, J.H. Bootstrap prediction intervals for autoregressive time series. Comput. Stat. Data Anal. 2007, 51, 3580–3594.
77. De Brabanter, K. Approximate Confidence and Prediction Intervals for Least Squares Support Vector Regression. IEEE Trans. Neural Netw. 2011, 22, 110–120. [CrossRef]
78. Cong, N.; Shang, J.; Ren, Y. Unstructured Road Spectrum a-stable Distribution Parameters Interval Estimation and Reconstruction Based on Moving Block Bootstrap Method. J. Mech. Eng. 2013, 49, 106–113. [CrossRef]
79. Khosravi, A.; Nahavandi, S.; Creighton, D. Comprehensive review of neural network-based prediction intervals and new advances. IEEE Trans. Neural Netw. 2011, 22, 341–356. [CrossRef] [PubMed]
80. Efron, B.; Tibshirani, R.J. An Introduction to the Bootstrap; Chapman and Hall: New York, NY, USA, 1993.
82. Zhang, K.; Yuan, F.; Guo, J.; Wang, G. A Novel Neural Network Approach to Transformer Fault Diagnosis Based on Momentum-Embedded BP Neural Network Optimized by Genetic Algorithm and Fuzzy c-Means. *Arab. J. Sci. Eng.* 2015, 41, 3451–3461. [CrossRef]

83. Shi, T.; Mei, F.; Lu, J.; Lu, J.; Pan, Y.; Zhou, C.; Zheng, J. Phase Space Reconstruction Algorithm and Deep Learning-Based Very Short-Term Bus Load Forecasting. *Energies* 2019, 12, 4349. [CrossRef]

84. Hyndman, R.J. Another look at forecast accuracy metrics for intermittent demand. *Foresight Int. J. Appl. Forecast.* 2006, 4, 43–46.

85. Li, R.; Jin, Y. A wind speed interval prediction system based on multi-objective optimization for machine learning method. *Appl. Energy* 2018, 228, 2207–2220. [CrossRef]

86. Li, K.; Wang, R.; Lei, H.; Zhang, T.; Liu, Y.; Zheng, X. Interval prediction of solar power using an Improved Bootstrap method. *Sol. Energy* 2018, 159, 97–112. [CrossRef]

87. Zhang, Y.; Hao, S.; Qian, X. Interval Prediction of Wind Power Based on Error ecomposition and Bootstrap Method. *Power Syst. Technol.* 2019, 43, 1941–1947.

88. Yuan, F.; Guo, J.; Xiao, Z.; Zeng, B.; Zhu, W.; Huang, S. A Transformer Fault Diagnosis Model Based on Chemical Reaction Optimization and Twin Support Vector Machine. *Energies* 2019, 12, 960. [CrossRef]