Article

Signal-Noise Identification for Wide Field Electromagnetic Method Data Using Multi-Domain Features and IGWO-SVM

Xian Zhang [1], Diquan Li [1,*], Jin Li [2,*], Bei Liu [3], Qiyun Jiang [1] and Jinhai Wang [1,4]

1   Key Laboratory of Metallogenic Prediction of Nonferrous Metals and Geological Environment, Monitoring Ministry of Education, School of Geosciences and Info-Physics, Central South University, Changsha 410083, China; zhangxian128@csu.edu.cn (X.Z.); jiangqiyun@csu.edu.cn (Q.J.); 205002061@csu.edu.cn (J.W.)
2   College of Information Science and Engineering, Hunan Normal University, Changsha 410081, China
3   College of Mathematics and Physics, Hunan University Arts and Science, Changde 415000, China; liubei@huas.edu.cn
4   The Third Geological Exploration Institute of Qinghai Province, Xining 810029, China
*   Correspondence: lidiquan@csu.edu.cn (D.L.); lijin@hunnu.edu.cn (J.L.); Tel.: +86-731-8883-6153 (D.L. & J.L.)

Abstract: Noise tends to limit the quality of wide field electromagnetic method (WFEM) data and exploration results. The existing WFEM denoising methods lack the signal identification process and are only able to filter or eliminate abnormalities in the time or frequency domain, which easily leads to the loss of more abundant real data and to low data quality. Thus, we built the WFEM data sample library to extract the multi-domain features. Then, neighborhood search and location sharing were used to improve the grey wolf optimizer (IGWO) algorithm. The support vector machine (SVM) parameters were optimized by IGWO to train multi-domain features, and an IGWO-SVM data model was generated. We used the data model to quantitatively test the WFEM signal and noise in the simulation and measured data. This method can effectively identify the WFEM signal and noise, eliminate the identified noise, and use the identified signal to reconstruct the effective data. Finally, the digital coherence technique was used to extract the spectrum amplitude of the effective frequency points. The experiments demonstrated the advantage of the convergence of IGWO algorithms and the comparison of the SVM parameters optimization techniques. The proposed method can quickly and effectively search the optimal SVM parameters, significantly improve the identification effect of WFEM signal noise, and completely remove the abnormal noise waveform in the reconstructed data. The more stable electric field curves in the results verify the effectiveness of the algorithm design and optimized identification method.

Keywords: wide field electromagnetic method (WFEM); multi-domain features; improved grey wolf optimizer (IGWO); support vector machine (SVM); signal-noise identification

1. Introduction

The wide field electromagnetic method (WFEM) is an important geophysical method, which is a controlled source frequency domain electromagnetic method with completely independent intellectual property rights in China [1]. With a complete theoretical system and mature instruments, WFEM overcomes the shortcoming of weak and random signals of the natural source electromagnetic method and improves the signal-to-noise ratio and resolution when the field sources are periodic signals and pseudo-random signals [2]. WFEM improves the work efficiency and anti-interference ability in the field, eliminates the weak signal caused by observing only in the "far region", organically integrates the "transition region" and the "far region", and significantly increases the observation scope and detection depth. Geoelectric information of multiple frequencies can be sent and received at one time. WFEM defines the wide field apparent resistivity for the whole

region by retaining the high-order term in the calculation formula and by observing one component that can obtain the electric field curve and the apparent resistivity curve [3,4].

Electromagnetic noise is more intense in a modern city and working area. To a certain extent, noise limits the further development of the electromagnetic method. Therefore, most electromagnetic method researchers are concerned with the denoising technique. However, WFEM data are also disturbed by noise, and the denoising technique is still the first prerequisite for the collected data. In addition, the high-quality of WFEM data are the foundation of inversion calculation and geological interpretation. To improve the longitudinal resolution and exploration effect of WFEM detection technology, it is necessary to strengthen the research on the WFEM data denoising method. The traditional WFEM data processing method uses frequency-domain processing, which can improve data quality with less interference. When most frequency points are distorted by persistent strong noise, these frequency-domain denoising methods rely on the selection of power spectrum, resulting in an unreliable data processing effect [5–7]. However, the signal processing methods in the time domain are all integrated processing of electromagnetic data collected [8–10], which improves the data quality to a certain extent but lacks the identification link of signal and noise. Therefore, how to effectively eliminate WFEM noise by using the new method is one of the key technical problems to be solved urgently. The comprehensive evaluation method is the smoothness of the curve and the frequency points without abnormal change. The WFEM 7 frequency wave data include 7-0/7-1/7-2/7-3/7-4/7-5 frequency groups, respectively. In this paper, we focus on the data analysis of 7-2/7-3 frequency groups, which are 64, 32, 16, 8, 4, 2, 1 Hz corresponding to 7-2 frequency group data and 48, 24, 12, 6, 3, 1.5, 0.75 Hz corresponding to 7-3 frequency group data [11].

Here, the time domain feature information is the time variable used to describe the signal waveform. Frequency domain feature analysis can observe signal features by frequency spectrum. Time-frequency domain features can represent multiple statistical values of the sample data to be tested. The multi-domain features' fusion can accurately characterize the details of WFEM data and quantitatively describe the difference between signal and noise. Support vector machine (SVM) is a generalized linear classifier, which classifies binary data through supervised learning [12]. Its decision boundary is the maximum margin hyperplane to be solved by learning samples. SVM usually solves the two-class problems by establishing a hyperplane and distinguishing positive and negative examples as much as possible [13]. When the WFEM data only need to be divided into signal and noise, the SVM is a suitable classification algorithm. However, the penalty factor c and kernel parameter g are the main factors that affect the SVM classification results [14,15]. Grey wolf optimizer (GWO) is a new type of swarm intelligence optimization algorithm [16]. GWO optimizes searches by simulating the social hierarchy relationship and hunting behavior of grey wolves in nature. The algorithm divides a population into four social levels, and the individuals in the population represent the solution of the optimization problem [17]. However, the convergence speed of the GWO algorithm is slow, and it is easy to fall into the local optimal solution but difficult to obtain the global optimal solution. In this paper, we propose a new method to improve the GWO algorithm, which uses neighborhood search and location sharing to enhance the balance between local sand global searches, maintain diversity and improve convergence speed.

In this paper, we propose a novel WFEM signal-noise identification method, which is based on multi-domain features and an improved grey wolf optimizer support vector machine (IGWO-SVM). We constructed a WFEM sample library, extracted the peak-to-peak value and pulse factor in time domain features, the mean frequency in frequency domain features, the wavelet singular entropy in time-frequency domain features, and analyzed the signal and noise feature of WFEM data. An IGWO algorithm was used to search the best parameters of the SVM, which learned the sample library's feature and trained data model. The results were compared with those of K-means clustering, Fuzzy C mean (FCM) clustering and the K nearest neighbor (KNN) algorithm. Then, the IGWO-SVM data model was used to directly remove the identified WFEM noise and retain the WFEM signal

for data reconstruction. Finally, the digital coherence technique was used to extract the reconstructed data spectrum amplitude of the effective frequency points. We compared the convergence of multiple intelligent optimization algorithms and optimization SVM models. The proposed method confirms that the fusion of multi-domain features and the IGWO-SVM can accurately and quickly recognize WFEM signals. We applied the proposed method to the simulation experiment and measured WFEM data for validation. The electric field curves were more stable, and data quality was improved. The satisfactory performance in the application and discussion verifies the effectiveness of the design and optimization method.

Note that the aim of this paper is to achieve the high precision WFEM signal-noise identification. The contributions of this paper are summarized as follows:

(1) The principle of multi-domain features and improved grey wolf optimizer are introduced. And the convergence of various optimization methods is illustrated.
(2) Four optimized SVM algorithm are quantified to demonstrate the advantages of the proposed method. Meanwhile, the K-means clustering, FCM clustering, KNN classification method and PSO-SVM method are compared.
(3) The validity of the proposed method is verified in many experiments and measured WFEM data.

The remainder of this paper is arranged as follows: Section 2 introduces the multi-domain features, and the principle and convergence of the grey wolf optimizer and the improved grey wolf optimizer. Section 3 presents the experiments and results that illustrate the effectiveness of the proposed method. Sections 4 and 5 show the applications and discussions in the measured WFEM data, respectively. Section 6 summarizes and highlights the major contributions of this paper.

2. Methodology

Based on a pseudo-random signal as the transmitting source, WFEM data will inevitably be affected by electromagnetic noise, resulting in abnormal waveform of the signal, and changing the electric field value. Considering that the normal WFEM signal should be a pseudo-random signal waveform, the signal is simple, regular and easy to identify. The feature extraction and intelligent identification are beneficial to WFEM signals and noise processing. Therefore, the fusion of multi-domain features and the IGWO-SVM were applied to WFEM signal identification. The proposed method was processed for the time-series waveform. Firstly, we extracted peak-to-peak values and the pulse factor feature in the time domain, the mean frequency feature in the frequency domain and the wavelet singular entropy in the time-frequency domain and introduced them for analyzing the WFEM signal and noise feature. The comparison of the convergence of several intelligent optimization algorithms and SVM parameters optimization methods followed. Finally, the multi-domain features and the IGWO-SVM was used in WFEM signal-noise identification. Next, multi-domain features and the IGWO algorithm were mainly introduced.

2.1. Multi-Domain Features

Feature extraction is a method and process of extracting object feature information by computer. It is mainly used for images, signal processing and machine learning to describe information. Multi-domain features are extracted from the time domain, the frequency domain and the time-frequency domain. In this paper, we focus on the peak-to-peak values and pulse factor features in the time domain, the mean frequency feature in the frequency domain, and the wavelet singular entropy in the time-frequency domain, respectively.

The peak-to-peak value is the difference between the maximum and minimum value of the signal, which is expressed as follows:

$$F_{pp} = \max(x(i)) - \min(x(i)) \tag{1}$$

The pulse factor is the peak signal divided by the mean absolute value and can also indicate whether the signal contains instantaneous spike, which is calculated as follows:

$$F_{pf} = \frac{\max|x(i)|}{\left(\frac{1}{N}\sqrt{\sum_{i=1}^{N}|x(i)|}\right)^2} \tag{2}$$

where x is time domain signal and N is the length of signal. According to the difference of WFEM signal and noise, the dimension feature value changes correspondingly, and the dimensionless index can show the noisy state of electromagnetic data more directly. Therefore, the dimension and dimensionless feature are used together.

Frequency domain analysis performs the Fourier transform through the time domain signal. The signal component and the time domain signal are interrelated and complement each other. Thus, the frequency domain is more concise. Frequency domain feature is extracted from the signal frequency spectrum feature by FFT. Among them, the mean frequency is expressed as follows:

$$F_{mf} = \frac{1}{N}\sum_{i=1}^{N} u(i) \tag{3}$$

Wavelet singular entropy is the most typical feature in the time-frequency domain [18]. Based on the theory of singular value decomposition, the wavelet singular entropy of the signal by wavelet transform method is decomposed into a series of singular values, which can reflect the basic feature of the original coefficient matrix. The uncertainty of the singular value set is analyzed by the statistical feature of information entropy, and a definite measure of the complexity of the original signal is given.

The singular value decomposition (SVD) of any $m \times n$ order matrix B can be expressed as follows:

$$B = U\Lambda V^T \tag{4}$$

where U and V are orthogonal matrices of $m \times m$ order and $n \times n$ order, respectively. $\Lambda = diag(\lambda_1, \lambda_2, \lambda_3, \ldots, \lambda_p)$ is the diagonal matrix, among them, $p = \min(m, n)$, its non-negative diagonal elements are arranged in descending order and are the singular eigenvalues of matrix A. SVD can represent the $m \times n$ order matrix A of rank K as the sum of K $m \times n$ order submatrices of rank 1. At this moment, the wavelet transform coefficient matrix of the signal can reflect the time-frequency distribution feature of the signal by SVD.

To quantitatively describe the frequency components and distribution feature of the signal, the wavelet singular entropy is defined as follows:

$$WSE = \sum_{i=1}^{N} \Delta p_i \tag{5}$$

where $\Delta p_i = -\left(\lambda_i / \sum_{i=1}^{N} \lambda_i\right) \log\left(\lambda_i / \sum_{i=1}^{N} \lambda_i\right)$ is the incremental wavelet singular entropy of the ith nonzero singular value λ_i. The simpler the signal being analyzed, the more concentrated the energy is in a few modes, and the smaller the wavelet singular entropy. Conversely, the more complex the signal, the more dispersed the energy, and the larger the wavelet singular entropy.

2.2. Grey Wolf Optimizer

Inspired by the predation behavior of grey wolves, Mirjalili et al. proposed the grey wolf optimizer (GWO) algorithm [16]. By simulating the predation behavior of grey wolves, the GWO was optimized based on the mechanism of pack cooperation [19]. The GWO algorithm is characterized by its simple structure in which few parameters need to be adjusted, its ease of implementation, its adaptive convergence factors and its information

feedback mechanism. It can achieve the balance between local optimization and global search, so it has good performance in precision and convergence speed to solve the problem.

Grey wolves encircle prey during a hunt, and the encircling behavior can be modeled as follows:

$$D = \left| C \cdot X_p(t) - X(t) \right| \tag{6}$$

$$X(t+1) = X_p(t) - A \cdot D \tag{7}$$

Equation (6) is the distance between an individual and the prey, and Equation (7) is the location update of the grey wolf, where t is the current iteration, C and A denote coefficient vectors, X_p is the position vector of the prey, and X indicates the position vector of the grey wolf. The vectors A and C are calculated as follows:

$$A = 2c \cdot r_1 - c \tag{8}$$

$$C = 2 \cdot r_2 \tag{9}$$

where the components of c are linearly decreased from 2 to 0 over the course of iterations and r_1 and r_2 are random vectors in [0,1]. The hunt is usually guided by α wolves, that is leaders, followed by β and δ wolves, which can also occasionally participate in hunting. However, in the search space, we have no idea about the location of the optimum solution.

Thus, the hunt, in which the hunters are moving toward the prey or solution over the provided search space, is the main approach of the GWO algorithm. To simulate the hunting behavior of grey wolves, we assume that α (best candidate solution), β and δ have better knowledge about the potential location of prey. Thus, we save the three best solutions obtained so far and oblige the other search agents to update their position by the position of the best search agent. The mathematical representation of such hunts is as follows:

$$\begin{cases} D_\alpha = \left| C_1 \cdot X_\alpha - X \right| \\ D_\beta = \left| C_2 \cdot X_\beta - X \right| \\ D_\delta = \left| C_3 \cdot X_\delta - X \right| \end{cases} \tag{10}$$

$$\begin{cases} X_1 = X_\alpha - A_1 \cdot D_\alpha \\ X_2 = X_\beta - A_2 \cdot D_\beta \\ X_3 = X_\delta - A_3 \cdot D_\delta \end{cases} \tag{11}$$

$$X(t+1) = \frac{X_1 + X_2 + X_3}{3} \tag{12}$$

where D_α, D_β and D_δ represent the distance between the current candidate grey wolf and the α, β and δ wolves, respectively. X_α, X_β and X_δ are the position of α, β and δ, respectively. C_1, C_2 and C_3 are random vectors, and X is the position of the current grey wolves. A_1, A_2 and A_3 are random vectors. The ω wolves, considered to be the remaining possible solutions in the pack, follow other solutions and update themselves with the other three best solutions expressed with Equation (11). The $X(t+1)$ is the final position of the ω wolves.

2.3. Improved Grey Wolf Optimizer

In the GWO, the search process is guided by three best wolves in each iteration, which shows a strong convergence toward these wolves [20]. In contrast, it suffers from a lack of the population diversity, an imbalance between the exploitation and exploration, and premature convergence [21]. Neighborhood search and location sharing are used to improve grey wolf optimization, namely IGWO, and enhance the ability of global optimization to avoid premature convergence [22].

The IGWO algorithm benefits from a new movement strategy, namely a dimensional-learning based hunting (DLH) search strategy, which is inherited from the individual hunting behavior of wolves in nature. DLH uses different methods to construct a neighborhood for each wolf, and neighboring information can be shared among wolves. Dimension

learning used for the DLH search strategy enhances the balance between local and global searches and maintains diversity.

In the DLH search strategy, each dimension of the new position of each wolf is calculated. This individual wolf is learned by its different neighbors, and a wolf from the top 3 wolves (α, β and δ) is randomly selected. First, a radius $R_i(t)$ is calculated using Euclidean distance between the current position of $X_i(t)$ and the candidate position $X_{i-GWO}(t+1)$ as follows:

$$R_i(t) = \|X_i(t) - X_{i-GWO}(t+1)\| \tag{13}$$

The neighbors of $X_i(t)$ defined by $N_i(t)$ is constructed as follows:

$$N_i(t) = \{X_j(t) | D_i(X_i(t), X_j(t)) \le R_i(t), X_j(t) \in (\alpha, \beta, \delta)\} \tag{14}$$

where $N_i(t)$ respected to radius $R_i(t)$, D_i is Euclidean distance between $X_i(t)$ and $X_j(t)$.

The neighborhood of $X_i(t)$ is constructed, multi-neighbors learning is performed as follows:

$$X_{i-DLH,d}(t+1) = X_{i,d}(t) + rand \times (X_{n,d}(t) - X_{r,d}(t)) \tag{15}$$

where $X_{i-DLH,d}(t+1)$ is the dth dimension of a random neighbor $X_{n,d}(t)$ selected from $N_i(t)$, and a random wolf $X_{r,d}(t)$ from α, β and δ wolf.

Selecting and updating the new position of $X_i(t+1)$ as follows:

$$X_i(t+1) = \begin{cases} X_{i-GWO}(t+1) & if \quad f(X_{i-GWO}) < f(X_{i-DLH}) \\ X_{i-DLH}(t+1) & otherwise \end{cases} \tag{16}$$

To verify the optimization performance of the IGWO algorithm. We give four benchmark functions to compare their convergence accuracy. Among them, the comparison of methods such as the GWO, the particle swarm optimization (PSO) [23], the multi-verse optimizer (MVO) [24], the moth-flame optimization (MFO) [25], the artificial bee colony (ABC) algorithm [26], the sine cosine algorithm (SCA) [27] and the imperialist competitive algorithm (ICA) [28]. Note that the population size is 10, and the maximum number of iterations is 100. Figure 1 shows convergence comparison of the four benchmark functions.

From Figure 1, we can see that the solution accuracy and convergence speed are better than other intelligent optimization algorithms at the same population and the iteration number. Through the convergence of the IGWO algorithm, we can see that this algorithm has obvious advantages in the optimization ability and stability of benchmark function and can better jump out of local optimization and obtain higher global optimization ability.
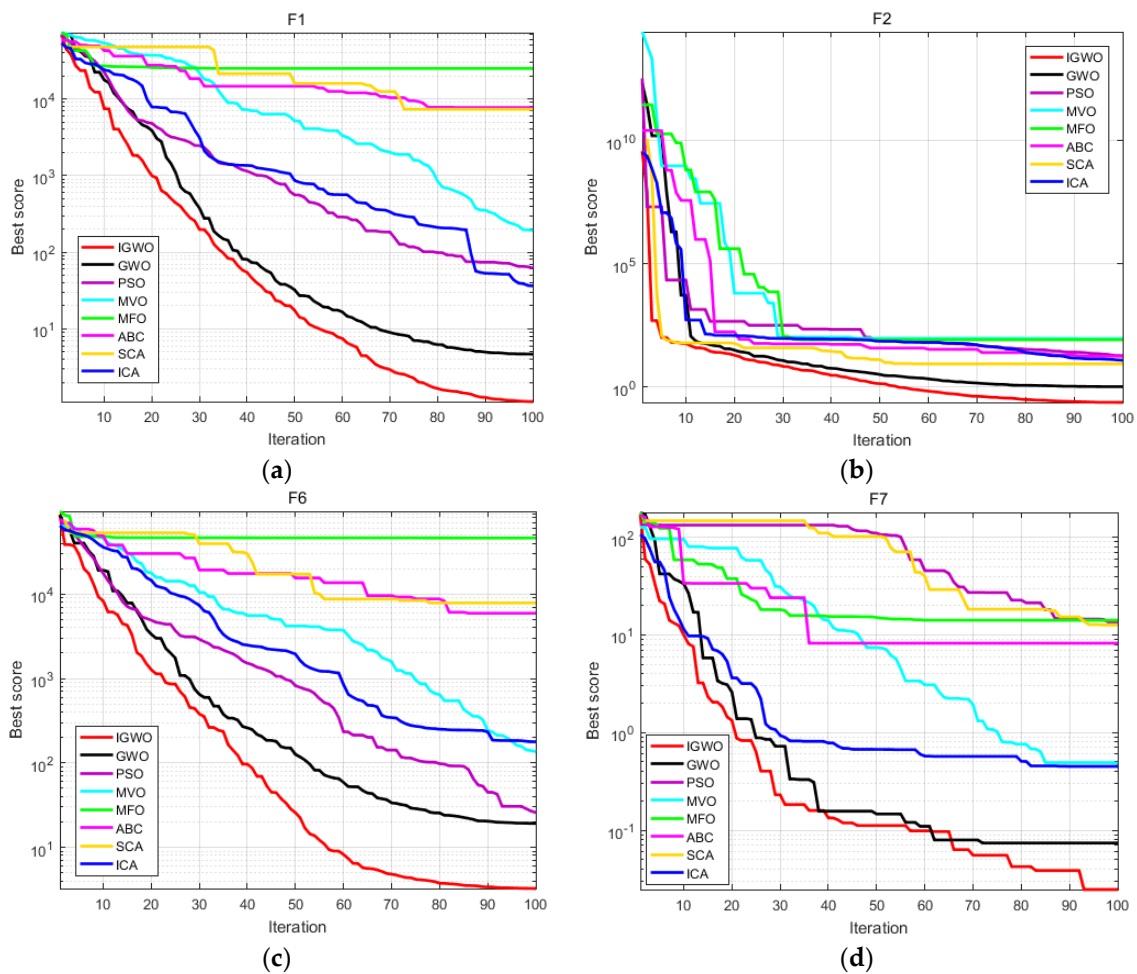
Figure 1. Convergence comparison of the benchmark functions. (**a**): F1; (**b**): F2; (**c**): F6; (**d**): F7.

3. Experiments and Results

In this section, we validate the proposed method by analyzing the feature extraction and the SVM parameters optimization. We also present our comprehensive comparison with the k-means clustering algorithm, the fuzzy c means (FCM) clustering algorithm, the k nearest neighbor (KNN) algorithm, the PSO-SVM method and the IGWO-SVM method in the sample library signals and the simulated analysis.

3.1. Sample Library Analysis

In order to analyze the quantitative identification relationship between pseudo-random signals and abnormal noise waveforms in the WFEM 7 frequency wave data, we built a data sample library of typical noise types and pseudo-random signals.

As shown in Figure 2, a group of time-domain waveforms of five types of signals and their corresponding spectra were randomly selected from the sample library. Among them, the sample library contained 30 pseudo-random signals, 30 impulse noises, 30 attenuation noises, 30 triangle wave noises and 30 square wave noises. The sampling length of each sample signal was 1200, and the sampling rate was 400 Hz.

We observed in the group of sample library signals, the time domain signal with noise results in abnormal mutation of the original pseudo-random signals, disorder of the signal and an increase of waveform amplitude (Figure 2). The signal was also seriously chaotic in the frequency domain, which could not reflect the inherent features of the original pseudo-random 7 frequency wave signal. However, the pseudo-random signal had the characteristics of periodicity, stable amplitude, relatively stable spectrum, and its frequency point information could be completely retained.
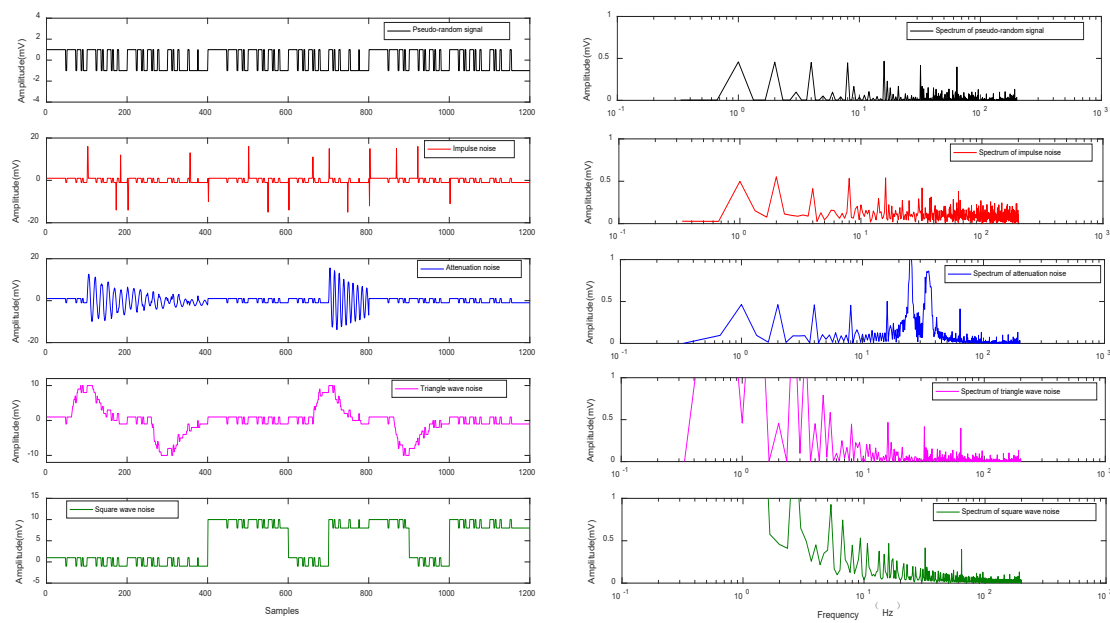
Figure 2. A set of sample library signals and frequency spectrum.

To test the performance of the IGWO optimization of SVM parameters, we set the upper bound of the parameters as 100, the lower bound of the parameters as 0.01, the maximum iteration times as 100, and the population size as 10, respectively. To optimize the penalty factor and kernel function parameters of the SVM, we compared with the sample library signals of SVM parameters optimized by the PSO, the cuckoo search (CS) [29] and the GWO, respectively. Note that the learning factors c1 and c2 in the PSO algorithm are 1.5. In the CS algorithm, the probability of being discovered by the host is 0.25. The performance comparison of the four optimization algorithms is shown in Tables 1 and 2. Among them, the best parameters value (c and g), mean square error (MSE), square correlation coefficient (SCC), algorithm iteration number and model accuracy were used for quantitative analysis.

$$MSE = \frac{1}{n}\sum_{j=1}^{n}(p_j)^2 \tag{17}$$

$$SCC = \left(1 - \frac{\sum\limits_{j}(p_j)^2}{\sum\limits_{j}(P_j - p_{jave})^2}\right) \tag{18}$$

where n is predicted sample size, p_j is absolute prediction error, P_j is real value and P_{jave} is the average of the real value. In the evaluation, the smaller the MSE is, the better, and the larger the SCC is, the better.

Table 1. The mean square error as the objective function of optimization.

Method/Parameter	c	g	MSE	SCC	Iteration	Accuracy	Time (s)
PSO-SVM	64.3989	13.7455	0.0096	0.9996	214	100	137
CS-SVM	55.0840	0.4135	0.0059	0.9953	45	100	34
GWO-SVM	95.2429	0.4137	0.0058	0.9950	46	100	37
IGWO-SVM	24.4877	0.4138	0.0057	0.9955	43	100	29

Table 2. The prediction error rata as the objective function of optimization.

| Method/Parameter	c	g	MSE	SCC	Iteration	Accuracy	Time (s)
PSO-SVM	15.2018	52.7497	0.0096	0.9996	278	100	174
CS-SVM	0.01	100	0.1552	0.8247	124	80	97
GWO-SVM	0.01	0.01	0.1529	0.8516	30	80	36
IGWO-SVM	28.1274	93.9011	0.0096	0.9996	275	100	166

From Table 1, although four type of the optimization algorithm can optimize SVM parameters, their performance and efficiency are lower than the IGWO algorithm. We can see that the IGWO algorithm optimized SVM parameters can obtain the smallest MSE and the minimum iteration and running time by using the MSE as the objective function for optimization. As can be seen from Table 2, when the prediction error rate is taken as the objective function, the IGWO algorithm can obtain the best SVM parameters with the minimum iterations on the premise of ensuring the accuracy of the model. However, the CS and GWO algorithms cannot guarantee the model accuracy in Table 2, resulting in unreliable optimization results. Therefore, the MSE was subsequently used as the objective function to optimize SVM parameters, that more accurately classified WFEM signals and noises.

Furthermore, we extracted the multi-domain features of the sample library signal for clustering and classification analysis. Figure 3 shows the signal noise classification of the sample library.

As shown in Figure 3, the multi-domain features are extracted from the sample library signals; the K-means clustering method can classify the 150 samples into two types in the sample library. When the multi-domain feature values of the WFEM data are close and chaotic, the K-means clustering method cannot select an appropriate clustering center, resulting in an unsatisfactory clustering effect. By calculating the Euclidean distance between each sample point and the cluster center [30], the FCM clustering method automatically divides the sample library signals and classifies the noisy signals and pseudo-random signals into different types. When the feature values of signal and noise in the sample library are similar, the Euclidean distance is only used to divide the sample library signal, which leads to the wrong division results. The KNN is one of the simplest methods in the supervised learning [31]. KNN is a kind of lazy learning without an explicit learning process or training process [32]. KNN is classified by measuring the distance between different feature values. Although the sample library can be effectively classified, it is impossible to accurately determine the K value for the recognition of the measured data, and the misjudgment phenomenon will also exist. The proposed method is completely adaptive to optimize parameters without artificial settings. By using the feature extraction, training and testing of the sample library, the signals and noises can be accurately divided in the sample library. Then the accuracy of the optimized SVM can be verified through the prediction effect. As a result, the multi-domain features and parameters of the optimized SVM method are suitable for WFEM data samples and for nonlinear and high-dimensional classification problems.

3.2. Simulation Analysis

To verify the identification effect of the proposed method, the simulated sample library noise types were used to analyze the synthesized WFEM signals. Figure 4 shows the signal-noise identification and spectrum results of the synthesized 7-2 frequency group signal and a comparison of the K-means method, the KNN method and the PSO-SVM method.
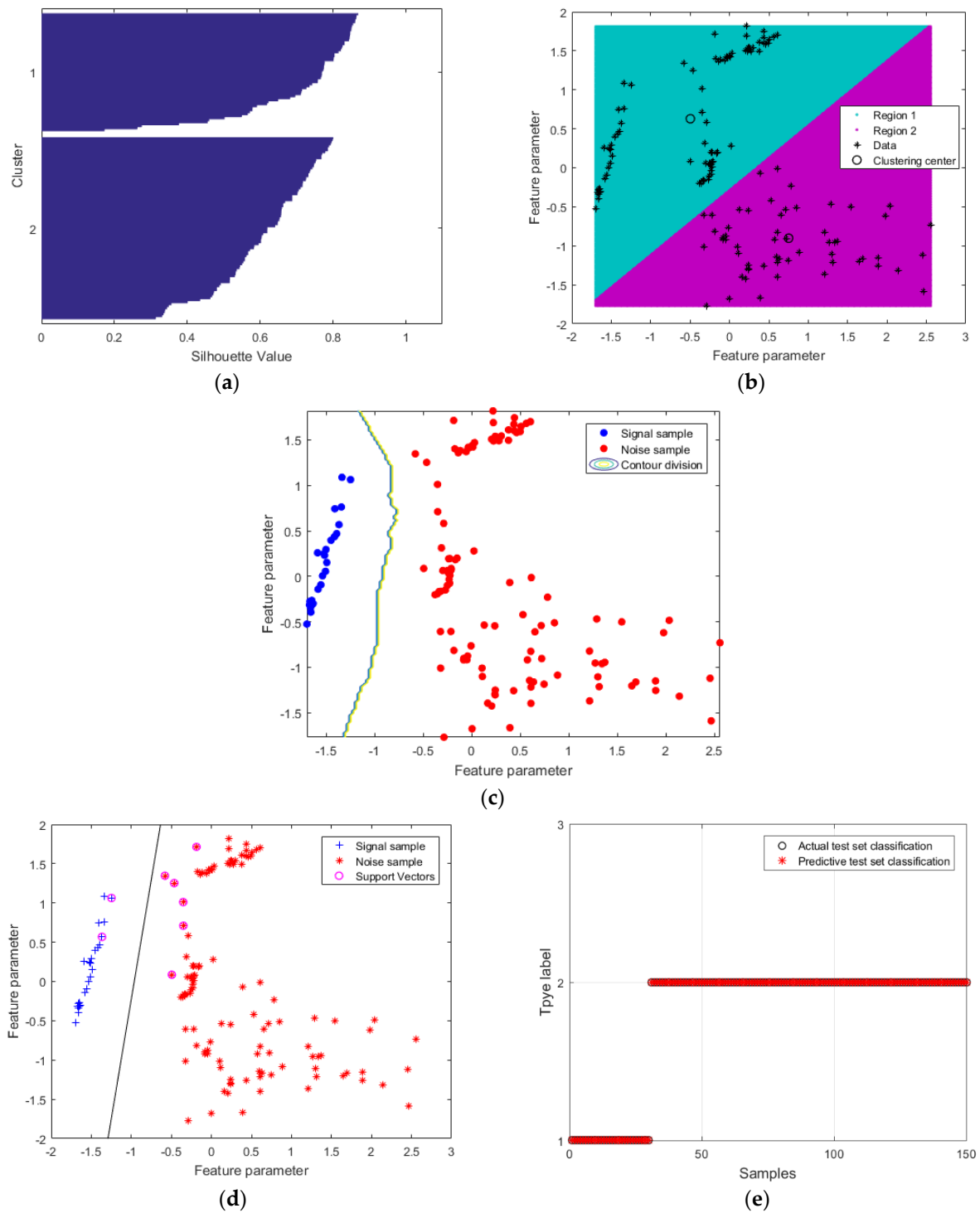
Figure 3. The results of the sample library signal noise classification: (**a**) K-means clustering; (**b**) Fuzzy C mean clustering; (**c**) K nearest neighbor classification; (**d**) Support vector machine and (**e**) Predictive effect.
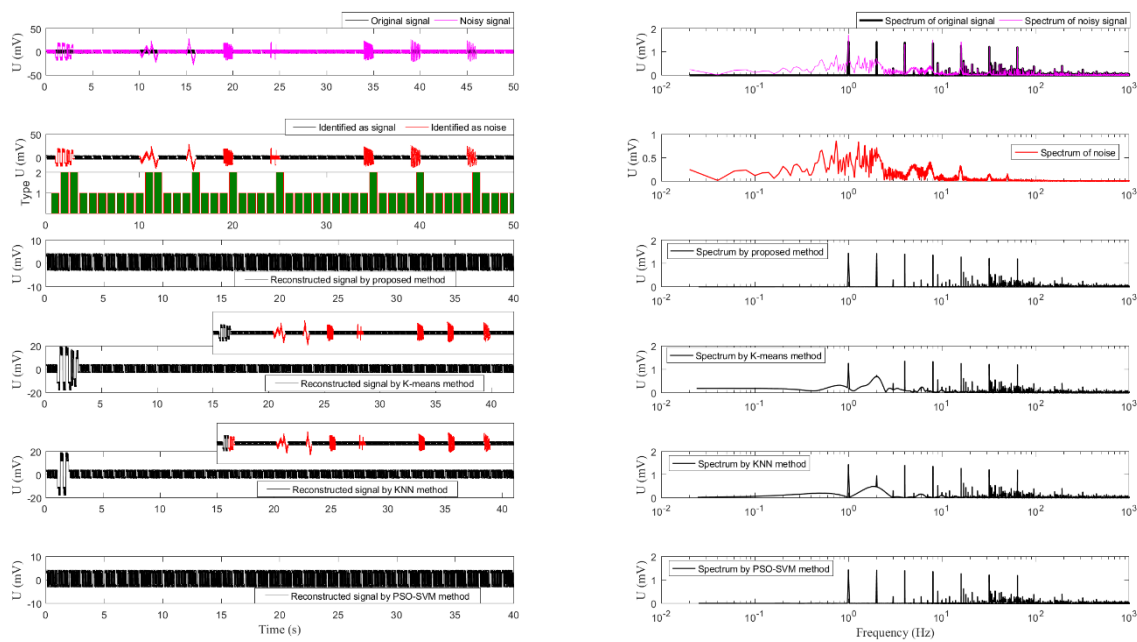
Figure 4. Signal-noise identification and spectrum analysis of simulation 7-2 frequency group signal.

Figure 4 shows that different noise types and abnormal waveforms appear in the time domain waveforms, that the frequency spectrum affected by noise is chaotic, and that the main frequency information changes to varying degrees with unstable main frequency value. Compared with the K-means clustering in the unsupervised learning method, due to the inconspicuous difference of the feature parameter values at the different noise types, some noise types cannot be identified, which makes it impossible to restore the original pseudo-random signal. In addition, different frequency points in its spectrum appear distorted. Compared with the KNN method, this method also has misidentification in the time domain, and part of the square wave noise is retained in the reconstructed signal, and the spectrum also produces corresponding distortion. Although the PSO-SVM method can effectively identify signal noise in the simulation 7-2 frequency group signal, we combined it with the convergence of the above optimization algorithm (Figure 1), which showed that the efficiency of the PSO algorithm is much lower than the IGWO algorithm (Tables 1 and 2). Even the final optimization parameters are not reliable in the next measured WFEM data. After processing by the proposed method, multi-domain feature extraction and the IGWO-SVM analysis can accurately identify the abnormal interference part of the signal, retain the part of the pseudo-random signal that is not affected by noise, filter out the noise spectrum and reconstruct the effective waveform of the pseudo-random signal and its original spectrum feature.

To quantitatively analyze the effectiveness of the proposed method, the electric field values and error of synthetic 7-2 frequency group signals at different frequencies are shown in Table 3. Among them, the error is calculated as follows:

$$error = \left(\frac{U_{noise} - U_{real}}{U_{real}} \right) \times 100\% \tag{19}$$

where U_{noise} is the noisy signal of electric field value and U_{real} is the real electric field value.

Table 3 shows that when the noise is added at different moments, the noisy signal shows mutation and chaos in the time domain and the frequency domain, the electric field values exceed the true values at 16-1 Hz frequencies, the distortion of the electric field value is 0.7154 mV at 2 Hz and the corresponding error increases to 49.85%. The reason is that noise affects the pseudo-random signal waveform in the time domain and the main frequency values and the harmonic component in the frequency domain. Compared with

the K-means clustering method, the main frequency value is alleviated, but the error is still large. The KNN method especially is only severely affected by noise at 2 Hz, resulting in a decrease of the main frequency value and an error increase to 32.59%, while the other main frequencies are relatively stable. The electric field value obtained by the PSO-SVM method and the proposed method is closer to the real electric field value, and the maximum error is reduced to 0.40% and 0.38%, respectively. Therefore, the proposed method can provide an effective way for WFEM signal processing.

Table 3. The electric field values and errors at different frequencies.

| Frequency F (Hz)	Real U (mV)	Noisy U (mV)	Error (%)	K-means U (mV)	Error (%)	KNN U (mV)	Error (%)	PSO-SVM U (mV)	Error (%)	Proposed Method U (mV)	Error (%)
1	1.4465	1.7278	19.44	1.2578	13.04	1.4446	0.13	1.4523	0.40	1.4521	0.38
2	1.4264	0.7154	49.85	0.7459	47.70	0.9614	32.59	1.4298	0.24	1.4297	0.23
4	1.4026	1.3388	4.54	1.3487	3.84	1.3987	0.27	1.4050	0.17	1.4051	0.17
8	1.3607	1.5062	10.69	1.3286	2.35	1.3631	0.17	1.3645	0.28	1.3644	0.27
16	1.2678	1.4314	12.90	1.2455	17.58	1.2709	0.24	1.2719	0.32	1.2720	0.33
32	1.2175	1.2285	0.90	1.2095	0.65	1.2164	0.09	1.2152	0.18	1.2152	0.18
64	1.1922	1.1955	0.27	1.1927	0.04	1.1929	0.05	1.1923	0.008	1.1923	0.008

4. Applications

4.1. Measured Data Analysis

Based on the analysis of noise types in the measured data, we added four kinds of noise to the measured data without abnormal waveform that analyzed and compared the signal noise identification effect and the electric field curve effect as shown in Figure 5.

As can be seen from Figure 5, the measured data without abnormal waveform shows periodicity, and the regularity in each cycle (among them, a cycle length is 19,200) belongs to the pseudo-random signal. When we observed the corresponding main frequency information, we noted that the data were almost unaffected by noise. By artificially adding noise, the time domain waveform appeared abnormal and frequency domain information became confused. Compared with the existing FCM clustering method based on feature extraction in the time domain, the FCM clustering method is an unsupervised learning method [33], which cannot effectively carry out feature learning and fine classification of signals and noises. When calculating the Euclidean distance between the feature parameter values and the clustering center, the signal and noise cannot be studied and classified effectively due to the high similarity or the large difference. The KNN method needs to obtain an effective K value through many experiments to identify the noise in the measured data. Therefore, the KNN method is only suitable for signal noise identification of small sample data and cannot effectively divide massive datasets. The proposed method is a completely adaptive optimization technique that uses peak-to-peak values and a pulse factor feature in the time domain, a mean frequency feature in the frequency domain, and wavelet singular entropy in the time-frequency domain for feature extraction, to effectively identify and remove the noise. Then, the identified signals without abnormal waveform are reconstructed, and the frequency spectrum information is highly restored. Furthermore, Figure 5b shows that the original electric field curve is stable and without abnormal changes. After adding noise, the electric field curve fluctuates at different frequency points. The electric field curve obtained by the FCM clustering method is still unstable and fluctuating because noise cannot be accurately identified. Although the KNN method can obtain the stable electric field curve, it cannot adaptively obtain the optimal parameter due to the influence of K value. This method will also affect the results and efficiency. The proposed method can accurately identify noise and restore the shape of the original electric field curve.

Figure 6 shows the signal-noise identification and reconstructed effect of measured signals. We can see that the measured signals are affected by a variety of the abnormal

mutations. The multi-domain features are extracted and the IGWO-SVM classification is combined to identify the noise with high precision, and the identified signal is retained and reconstructed.
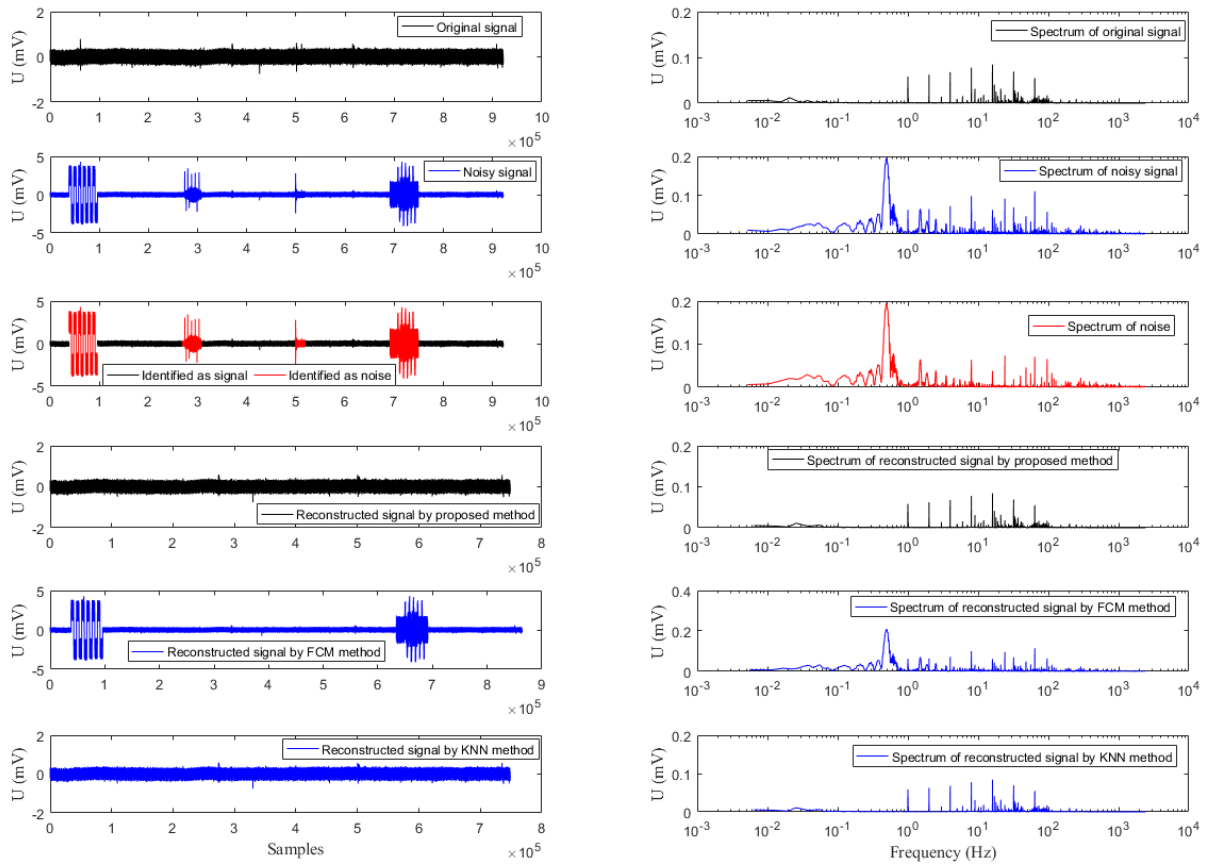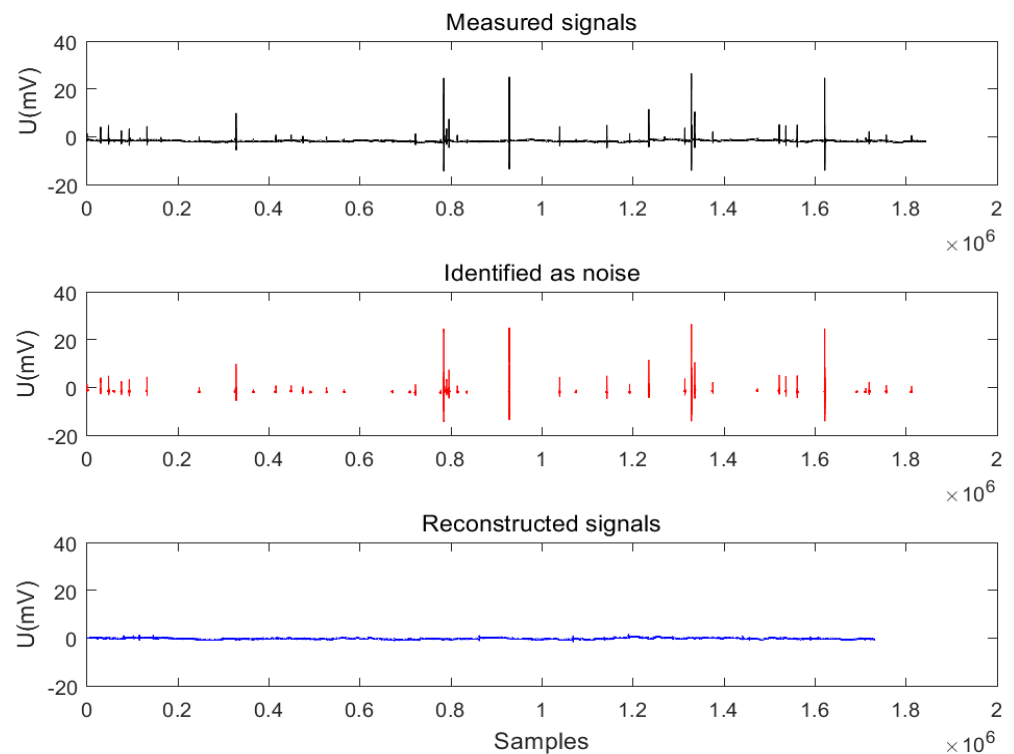

(**a**)
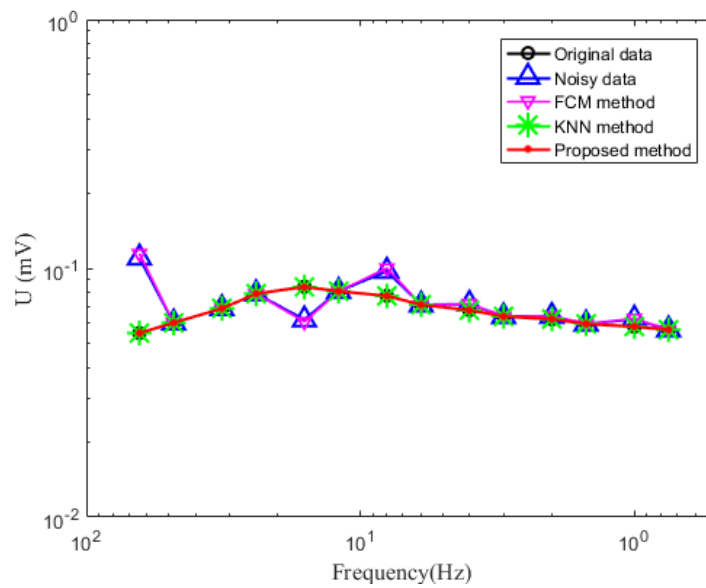
(**b**)

Figure 5. The process and results of the measured WFEM data, with comparison to the results from the FCM clustering method and the KNN method: (**a**) signal identification and data reconstruction; (**b**) electric field curve analysis.

Figure 6. The example of the measured signal noise identification.

4.2. Applied to Electric Field Curve

We conducted the electric field curve analysis of the measured sites. We used the digital coherence technique to extract the spectrum amplitude of the effective frequency points and analyze the medium and low frequency data of WFEM. The measured data of the 7-2/7-3/7-4/7-5 frequency groups were analyzed and processed in detail. Among them, 1, 0.5, 0.25, 0.125, 0.0625, 0.03125, 0.015625 Hz corresponded to 7-4 frequency group data, and 0.75, 0.375, 0.1875, 0.09375, 0.046875, 0.0234375, 0.01171875 Hz corresponded to 7-5 frequency group data. Figure 7 shows the electric field curve effect of the measured sites (S_1, S_2, S_3) by the proposed method in a certain industrial area in China.

From Figure 7a, the trend of the original electric field curve can be maintained after the processing of the high-quality WFEM data by the proposed method. Figure 7b,c show the electric field curve of the medium frequency of 7-2/7-3 frequency groups data and the low frequency of 7-4/7-5 frequency groups data, respectively. The original time domain data contained typical noise types that resulted in several frequency points falling or rising at the corresponding frequencies band. The high-precision identification technology not only eliminates the noise, but also improves the data quality of WFEM, and the electric field curve also eliminates the abnormal change of frequency points. The completely adaptive optimization method is proposed that can provide a novel technique for future inversion interpretation using the high-quality of WFEM data.
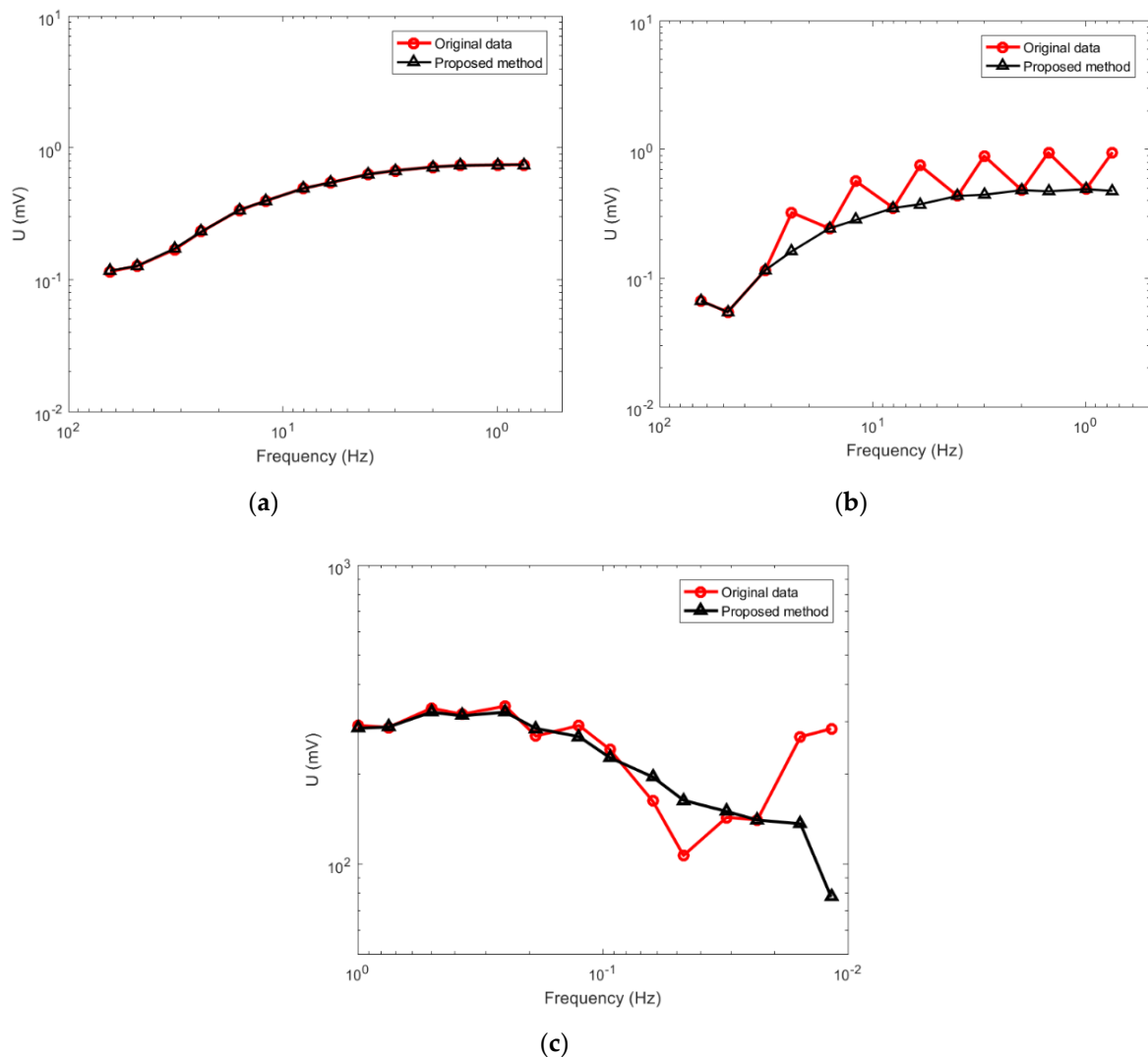
Figure 7. The electric field curve analysis of measured site. (**a**): S$_1$; (**b**): S$_2$; (**c**): S$_3$.

5. Discussions

The electromagnetic method is an important geophysical exploration method, which mainly includes the natural source electromagnetic method and the artificial source electromagnetic method. Compared to the natural source electromagnetic method, the artificial source electromagnetic method overcomes the weak and random signal in the natural field source and further improves the signal-to-noise ratio and resolution of signal. Incidentally, the field source is mainly composed of periodic square waves and pseudo-random signals. With the development of modern industry and technology, electromagnetic interference has become stronger and stronger, and noise suppression has always been a key problem for many electromagnetic workers, restricting the development of technical methods to a certain extent. The WFEM uses an artificial field source with a very powerful signal transmitter. In actual experiments, the observed signals were inevitably affected by various types of strong interference. To improve the longitudinal resolution and exploration effect of the WFEM detection technology, it is necessary to strengthen the research on the denoising method of the WFEM data. Therefore, how to effectively eliminate the noisy data of WFEM by using the new method is one of the key technical problems that urgently needs to be solved.

In recent years, time domain and frequency domain processing methods have been proposed for WFEM data processing, but the WFEM signal and noise recognition technology of time domain waveform is rarely proposed. Therefore, a WFEM signal-noise

identification method based on multi-domain features and the IGWO-SVM has been proposed in this paper. We first introduced the characteristic parameters to describe the WFEM signal-noise and the improved intelligent optimization algorithm and compared the convergence of multiple intelligent algorithms (Figure 1) to provide effective parameters for optimizing SVM classification. In the experiment, we introduced the signal and noise types and their frequency spectrums in the WFEM sample library (Figure 2). At the same time, we conducted parameter selection and performance comparison of the four optimized SVM algorithms to compare the sample library signals, highlighting the advantages of the IGWO-SVM (Tables 1 and 2). We further used clustering and classification algorithms to divide the sample library (Figure 3). To verify the identification effect, we conducted a comparison and quantitative analysis of the simulated synthetic data (Figure 4 and Table 3). In the applications, a noiseless measured site was selected for artificial noise-added processing, and the electric field curve compared results are shown (Figure 5). When the measured data were affected by noise, the proposed method identified the noise with high accuracy and reconstructed a high-quality effective signal (Figure 6). The effectiveness of the proposed method was further verified by comparing the electric field curves before and after data processing (Figure 7).

In a word, the proposed method improves the process of insufficient signal-noise identification in existing methods, reduces the excessive denoising processing of valid signals, and improves the data quality. However, applying the results of this paper to the inversion and interpretation of geophysical data will be the focus of further research.

6. Conclusions

A novel WFEM signal identification method, which uses multi-domain feature parameters to analyze the WFEM signal noise feature and applies the IGWO-SVM to identify signal and noise, while reconstructing the high-quality of WFEM data, has been developed.

The proposed method has been proven in the feature extraction of a sample library signal, IGWO convergence performance, optimal parameter of IGWO-SVM search ability and optimal classification effect, as well as analysis of the simulated and measured WFEM data. The results show that the WFEM signal-noise can be accurately identified. The reconstructed signal and its spectral information completely conform to the essential feature of WFEM pseudo-random data, and electric field curve is also more stable. The proposed method lays the foundation for feature extraction, improved intelligent optimization and WFEM signal processing. However, when the distinction between signal and noise is gradually fuzzy or complex or is subject to persistent strong interference, how to identify and denoise with high precision will be the focus of the future research.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used during the current study are available from the corresponding authors on reasonable request.

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

References

1. He, J.S. Wide field electromagnetic sounding methods. *J. Cent. South Univ. (Sci. Technol.)* **2010**, *41*, 1065–1072.
2. He, J.S.; Tong, T.G.; Liu, J.X. Mathematical analysis and realization of an sequence pseudo-random multi-frequencies signal. *J. Cent. South Univ. (Sci. Technol.)* **2009**, *40*, 1666–1671.
3. Ward, S.H. Electrical, electromagnetic and magnetotelluric methods. *Geophysics* **1980**, *45*, 1659–1666. [CrossRef]
4. Tang, J.T.; He, J.S. A new method to define the full-zone resistivity in horizontal electric dipole frequency soundings on a layered earth. *Chinese J. Geophys.* **1994**, *37*, 543–552.
5. Zhang, B.M.; Jiang, Q.Y.; Mo, D.; Xiao, L.Y. A new method for handling gross errors in electromagnetic prospecting data. *Chin. J. Geophys.* **2015**, *58*, 2087–2102.
6. Mo, D.; Jiang, Q.Y.; Li, D.Q.; Chen, C.J.; Zhang, B.M.; Liu, J.W. Controlled-source electromagnetic data processing based on gray system theory and robust estimation. *Appl. Geophys.* **2017**, *14*, 570–580. [CrossRef]
7. Yang, Y.; He, J.S.; Li, D.Q. A noise evaluation method for CSEM in the frequency domain based on wavelet transform and analytic envelope. *Chinese J. Geophys.* **2018**, *61*, 344–357.
8. Yang, Y.; Li, D.Q.; Tong, T.G.; Zhang, D.; Zhou, Y.T.; Chen, Y.K. Denoising controlled-source electromagnetic data using least-squares inversion. *Geophysics* **2018**, *83*, E229–E244. [CrossRef]
9. Li, G.; He, Z.S.; Tang, J.T.; Deng, J.Z.; Liu, X.Q.; Zhu, H.J. Dictionary learning and shift-invariant sparse coding denoising for controlled-source electromagnetic data combined with complementary ensemble empirical mode decomposition. *Geophysics* **2021**, *86*, A27–WB97. [CrossRef]
10. Li, J.; Peng, Y.Q.; Tang, J.T.; Li, Y. Denoising of magnetotelluric data using K-SVD dictionary training. *Geophys. Prospect.* **2021**, *69*, 448–473. [CrossRef]
11. Yang, Y.; He, J.S.; Li, D.Q. Energy distribution and effective components analysis of 2n sequence pseudo-random signal. *Tran. Nonferrous Met. Soc. China.* **2021**, *31*, 2102–2115. [CrossRef]
12. Kim, H.C.; Pang, S.N.; Je, H.M.; Kim, D.J.; Bang, S.Y. Constructing support vector machine ensemble. *Pattern Recogn.* **2003**, *36*, 2757–2767. [CrossRef]
13. Li, J.; Zhang, X.; Tang, J.T.; Cai, J.; Liu, X.Q. Audio magnetotelluric signal-noise identification and separation based on multifractal spectrum and matching pursuit. *Fractals* **2019**, *27*, 1940007. [CrossRef]
14. Zhu, W.Q.; Bao, H.X.; Zeng, Z.G.; Wen, Z.Q.; Zhu, Y.H.; Xiang, H.Z. Support Vector Machine Optimized Using the Improved Fish Swarm Optimization Algorithm and Its Application to Face Recognition. *Int. J. Pattern Recogn.* **2019**, *33*, 1956010. [CrossRef]
15. Tang, X.Z.; Hong, H.Y.; Shu, Y.Q.; Tang, H.J.; Li, J.F.; Liu, W. Urban waterlogging susceptibility assessment based on a PSO-SVM method using a novel repeatedly random sampling idea to select negative samples. *J. Hydrol.* **2019**, *576*, 583–595. [CrossRef]
16. Mirjalili, S.; Mirjalili, S.M.; Lewis, A. Grey Wolf Optimizer. *Adv. Eng. Softw.* **2014**, *69*, 46–61. [CrossRef]
17. Kohli, M.; Arora, S. Chaotic grey wolf optimization algorithm for constrained optimization problem. *J. Comput. Des. Eng.* **2017**, *5*, 458–472. [CrossRef]
18. He, Z.Y.; Ling, F.; Lin, S.; Bo, Z.Q. Fault detection and classification in EHV transmission line based on wavelet singular entropy. *IEEE T. Power Deliver.* **2010**, *25*, 2156–2163. [CrossRef]
19. Heidari, A.A.; Abbaspour, R.A.; Chen, H.L. Efficient boosted grey wolf optimizers for global search and kernel extreme learning machine training. *Appl. Soft. Comput.* **2019**, *81*, 105521. [CrossRef]
20. Zhang, X.; Li, D.Q.; Li, J.; Li, Y. Grey wolf optimization-based variational mode decomposition for magnetotelluric data combined with detrended fluctuation analysis. *Acta Geophys.* **2022**. [CrossRef]
21. Heidari, A.A.; Pahlavani, P. An efficient modified grey wolf optimizer with Lévy flight for optimization tasks. *Appl. Soft. Comput.* **2017**, *60*, 115–134. [CrossRef]
22. Nadimi-Shahraki, M.H.; Taghian, S.; Mirjalili, S. An improved grey wolf optimizer for solving engineering problems. *Expert Syst. Appl.* **2021**, *166*, 113917. [CrossRef]
23. Kennedy, J.; Eberhart, R. Particle Swarm Optimization. In Proceedings of the ICNN'95-International Conference on Neural Networks, Perth, WA, Australia, 27 November–1 December 1995.
24. Mirjalili, S.; Mirjalili, S.M.; Hatamlou, A. Multi-verse optimizer: A nature-inspired algorithm for global optimization. *Neural Comput. Appl.* **2015**, *27*, 495–513. [CrossRef]
25. Mirjalili, S. Moth-flame optimization algorithm: A novel nature-inspired heuristic paradigm. *Knowl-Based Syst.* **2015**, *89*, 228–249. [CrossRef]
26. Karaboga, D.; Basturk, B. A powerful and efficient algorithm for numerical function optimization: Artificial bee colony (ABC) algorithm. *J. Glob. Optim.* **2007**, *39*, 459–471. [CrossRef]
27. Mirjalili, S. SCA: A Sine Cosine Algorithm for Solving Optimization Problems. *Knowl Based Syst.* **2016**, *96*, 120–133. [CrossRef]
28. Kaveh, A.; Talatahari, S. Optimum design of skeletal structures using imperialist competitive algorithm. *Comput. Struct.* **2010**, *88*, 1220–1229. [CrossRef]
29. Yang, X.S.; Yang, S. Cuckoo Search: Recent Advances and Applications. *Neural Comput. Appl.* **2014**, *24*, 169–174. [CrossRef]
30. Zhang, X.; Li, J.; Li, D.Q.; Li, Y.; Liu, B.; Hu, Y.F. Separation of magnetotelluric signals based on refined composite multiscale dispersion entropy and orthogonal matching pursuit. *Earth Planets Space* **2021**, *73*, 76. [CrossRef]
31. Omiotek, Z.; Dzierzak, R.; Kepa, A. Fractal analysis as a method for feature extraction in detecting osteoporotic bone destruction. *Fractals* **2021**, *29*, 2150095. [CrossRef]

32. Hu, M.; Tsang, E.C.C.; Guo, Y.T.; Chen, D.G.; Xu, W.H. Attribute reduction based on overlap degree and k-nearest-neighbor rough sets in decision information systems. *Inf. Sci.* **2022**, *584*, 301–324. [CrossRef]
33. Li, G.; He, Z.S.; Deng, J.Z.; Tang, J.T.; Fu, Y.Y.; Liu, X.Q.; Shen, C.M. Robust CSEM data processing by unsupervised machine learning. *J. Appl. Geophy.* **2021**, *186*, 104262. [CrossRef]