IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received April 27, 2019, accepted May 15, 2019, date of publication May 22, 2019, date of current version July 18, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2918156

A Novel Hybrid Prediction Model for Hourly Gas Consumption in Supply Side Based on Improved Whale Optimization Algorithm and Relevance Vector Machine

WEIBIAO QIAO[1], KUN HUANG[2], MOHAMMADAMIN AZIMI[3], AND SHUAI HAN[3]
[1]School of Environmental and Municipal Engineering, North China University of Water Resources and Electric Power, Zhengzhou 450046, China
[2]State Key Laboratory of Oil and Gas Reservoir Geology and Exploitation, Southwest Petroleum University, Chengdu 610500, China
[3]Trenchless Technology Center, Louisiana Tech University, Ruston, LA 71270, USA

Corresponding author: Kun Huang (swpuhk_yjs@126.com)

ABSTRACT Accurate short-term prediction of the natural gas load is of great significance to the operation and allocation of the pipeline network. Because the short-term natural gas load has obvious nonlinearity and randomness, the traditional regression model is difficult to predict accurately. Therefore, this paper proposes a hybrid prediction model that integrates an improved whale swarm algorithm (IWOA) and relevance vector machine (RVM). In addition, empirical mode decomposition (EMD), approximate entropy (ApEn), and C-C method are introduced to aid the calculation. In this paper, the IWOA is used to test the four functions and compared with the other five algorithms. The results show that the convergence accuracy and convergence speed of the new algorithm are higher than other algorithms, indicating that it has better global optimization ability. Second, the IWOA-RVM model is used to predict the supply data of two natural gas stations in Anhui Province, China. The prediction results are compared with the five algorithms including RBFNN, GRNN, ELMANNN, LSSVM, and SMOSVM. The results show that: 1) through the test of four functions, IWOA has better ability to jump out of local optimum, has higher optimization performance, and the calculation speed is at a medium level and 2) compared with other models, the IOWA-RVM model has higher prediction accuracy when the amount of data is larger or smaller, but the calculation time is relatively long, but the calculation time is acceptable in engineering.

INDEX TERMS Short-term, natural gas demand, prediction, relevance vector machine, improved whale swarm algorithm.

I. INTRODUCTION

According to BP's World Energy Outlook in 2018, global energy demand is still showing a clear upward trend, and renewable energy is growing at the fastest rate. However, by 2040, although the proportion of primary energy demand has declined, oil, natural gas and coal still account for about 70% of the energy market, of which natural gas is growing much faster than oil and coal. At present, the global energy structure is undergoing a transition period. Natural gas, as a clean and efficient energy source, plays an important role in the energy system.

The associate editor coordinating the review of this manuscript and approving it for publication was Haruna Chiroma.

Nowadays, with the popularity of big data, many enterprises use the collected data to forecast and analyze. Especially for natural gas system, accurate demand forecasting can not only make the national energy macro-control more reasonable, but also make the management of pipeline network more effective. Especially in the context of rapid development of blockchain technology, accurate prediction is helpful to implement scheduling [1]. Moreover, accurate prediction is also crucial for the management of long-distance pipelines [2]. Natural gas demand forecasting can be divided into long-term forecasting, medium-term forecasting, short-term forecasting and ultra-short-term forecasting according to the time length. Because gas load forecasting has a high degree of nonlinear characteristics and prediction

is more difficult, more scholars pay more attention to related research.

Natural gas load forecasting models can be divided into two types: traditional models and intelligent models. There are many well-known statistical models, including RW model, TC model, ARX model [3], OLS model, ARIMA model, SARIMAX model [4], ARIMAX model [5], and hybrid prediction model based on ARMA and GA [6]. Although these statistical models are easily to realize, they still have inherent disadvantages, that is, they are theoretically linear. They are not only difficult to deal with complex non-linear problems, but also difficult to obtain satisfactory prediction performance.

However, in recent years, more and more scholars began to study intelligent algorithms, such as fuzzy theory models [7], support vector machine (SVM) models [8], [9], and neural network (NN) [10]–[15]. Among them, SVM model has remarkable prediction performance, especially in the case of small amount of data [16]. However, the SVM model also faces many challenges, which greatly affect the forecasting accuracy, such as the selection of the kernel function and its embedded parameters, the determination of appropriate penalty parameters and the suitable cross-validation process [9]. Therefore, in [17], an improved SVM algorithm is proposed, which is based on the combination of FNF (false neighbors filtered) and SVRLP (support vector regression local predictor). Compared with FNF-SVRLP model, the prediction accuracy of SVRLP model, autoregressive moving average model and artificial neural network model is lower. In [18], a structure-calibrated support vector regression (SC-SVR) model is proposed to forecast the gas demand. Compared with least squares SVM model and dynamic back propagation neural network (BPNN), this model has the lower error. Reference [19] presents a gas demand model which combines multi-neural network and multi-wavelet transform. Compared with the combination of three NN forecasting model without MT and any single NN forecast model, the proposed forecasting model has higher accuracy. In [20], a coupled gas demand forecasting model combines genetic algorithm and NN is developed. The input variable parameters of this coupling model are rainfall, relative humidity, temperature and wind speed. The results show that this prediction model has good accuracy in indexes of absolute deviation and correlation coefficient. In [21], authors proposed an appropriate combinational method of improved BPNN and real-coded GA for Shanghai's gas demand prediction. The prediction results based on this combination method are superior to several different combination forecasting models. In [22], authors proposed a hybrid model which is a wavelet BPNN optimized by GA and overcomes the problems of the traditional BP algorithm. In [23], authors take the gas demand in Szczecin (Poland) as an example, and presents the forecasting results using multilayer perceptron model (MLP) among ANNs. In the process of forecasting, some major factors including weather and calendar are input into the prediction model. The prediction results indicate that the MLP 22-36-1 model has good prediction accuracy. In [24], authors uses a novel fractional time delayed grey model with grey wolf optimization algorithm to forecast the natural gas demand in Chongqing, China, it shows this new model has better prediction than other reference models.

These research papers abovementioned enrich the contents of STNGLP, not only including new forecasting methodology proposed, but also containing novel hybrid or combined models proposed. However, these proposed artificial intelligent models and relevant hybrid or combined models also suffer from several embedded shortcomings, such as difficult to determine suitable network structure [25], time consuming, low convergent speed, and trapping into local optimal solution [26]. Therefore, looking for novel forecasting models is still an important issue. Although intelligent algorithms have more advantages than traditional algorithms, they also have some unavoidable disadvantages, such as long computation time, slow convergence speed, easy to converge prematurely, etc. Therefore, the optimization prediction model is still a hot research topic.

In order to obtain higher accuracy and stability at the same time, this paper proposes a novel hybrid prediction system which includes EMD, AE and PSR, RVM, and IWOA. This system consists of three modules: the whole algorithm module, the improved whale optimization algorithm module, and the evaluation module. EMD technique is applied to decompose data containing imfs and residue. In order to shorten the calculation time and improve the calculation efficiency, the entropy values of imf and residual error are calculated by acoustic emission method. CC method is used to calculate the ED and ODT of each Nimf in the data. According to the ED and ODT of each Nimf in the data, the phase space of each Nimf is reconstructed. After verifying the validity of the proposed IWOA model, a hybrid prediction RVM-IWOA (i.e. RVM optimized by IWOA) model is established and applied to the prediction of natural gas demand in the whole algorithm module. IWOA is tested on four test functions and compared with five optimization algorithms in the improved whale optimization module: COA, FOA, IGOA, IPSOA and WOA. Finally, the evaluation module including evaluation criteria is used to evaluate the prediction results.

The rest of this paper is arranged as follows. Sections 2 describe in detail the methods used in this study. Sections 3 give a novel hybrid forecast system detailly. Moreover, effective model checking method is described in Section 4. Three experimental results and their corresponding analysis are given in Section 5. The discussions are introduced in Section 6. Finally, the last Section 7 presents the conclusions of this paper.

II. THEORIES

A. RELEVANCE VECTOR MACHINE (RVM)

In 2000, Micnacl E. Tipping proposed a sparse probability model similar to SVM (Support Vector Machine), called relevance vector machine, which is a new supervised learning

method. Compared with SVM, the biggest advantage of RVM is that it greatly reduces the computational complexity of the kernel function and overcomes the shortcoming that the selected kernel function must satisfy Mercer condition.

For a given input set $\{x_n\}_{n=1}^N$ and corresponding output set $\{t_n\}_{n=1}^N$, the RVM model can be defined as follows:

$$t_n = \sum_{i=1}^{N} w_i \varphi_i(x) + w_0 + \xi_n \tag{1}$$

where w is the weight vector, $w = (w_0, w_1, \ldots w_N)$; ζ_n is assumed to be independently sampled from a zero-mean Gaussian distribution; $\varphi_i(x)$ is the nonlinear basis function.

Assuming that t_n is independent of each other, the likelihood estimation of training sample set $\{x_n, t_n\}_{n=1}^N$ can be written as follows:

$$p\left(t|w,\sigma^2\right) = \left(2\pi\sigma^2\right)^{-0.5N} \exp\left(-\frac{\|t-\psi(x)w\|^2}{2\sigma^2}\right) \tag{2}$$

where $t = (t_1, \ldots, t_N)^T$;

$$\psi(x) = \begin{bmatrix} \varphi(x_1) \\ \varphi(x_2) \\ \cdots \\ \varphi(x_N) \end{bmatrix} = \begin{bmatrix} 1 & K(x_1,x_1) & \cdots & K(x_1,x_N) \\ 1 & K(x_2,x_1) & \cdots & K(x_2,x_N) \\ 1 & \cdots & \cdots & \cdots \\ 1 & K(x_N,x_1) & \cdots & K(x_N,x_N) \end{bmatrix}$$

It will be analytically convenient to introduce a Gaussian prior distribution for w:

$$p(w|\alpha) = \prod_{i=0}^{N} N\left(w_i|0,\alpha_i^{-1}\right) \tag{3}$$

where α is the N +1 dimensional hyperparameter vector.

According to the Bayesian principle, the posterior probability distribution of all unknown parameters is:

$$p\left(w|t,\alpha,\sigma^2\right) = (2\pi)^{\frac{-(N+1)}{2}} \left|\sum\right|^{-0.5} \text{gexp}\left\{-(w-\mu)^T \sum^{-1} (w-\mu)\right\} \tag{4}$$

where the posterior covariance matrix is:

$$\begin{aligned} \sum &= \left(\sigma^{-2}\psi^T\psi + A\right)^{-1} \\ \mu &= \sigma^{-2} \sum \psi^T t \\ A &= diag\left(\alpha_0, \alpha_1, \llcorner, \alpha_N\right) \end{aligned} \tag{5}$$

The formulas for calculating the hyperparameter α and variance σ^2 are:

$$\alpha_i^{new} = \frac{\gamma_i}{\mu_i^2} \tag{6}$$

$$\left(\sigma^2\right)^{new} = \frac{\|t-\psi\mu\|^2}{N - \sum_{i=0}^{N} \gamma_i} \tag{7}$$

where μ_i is the i-th posterior average weight; N_{ii} is the i-th diagonal element of the posterior covariance matrix; N is the number of sample data; $\gamma_i = 1 - \alpha_i \psi_{i,i}$.

Given a new input value x_*, the probability distribution of the corresponding output follows a Gaussian distribution, i.e.:

$$p\left(t_*|t,\alpha_{MP},\sigma_{MP}^2\right) = N\left(t_*|y_*,\sigma_*^2\right) \tag{8}$$

The average predicted value is:

$$y_* = \mu^T \varphi(x_*) \tag{9}$$

B. IMPROVED WHALE OPTIMIZATION ALGORITHM (IWOA)

In 2016, Mirjalili proposed WOA [27], a new heuristic optimization that mimics humpback whale hunting. In WOA, each humpback whale represents a viable solution. In marine activities, humpback whales have a special way of hunting called bubble-net hunting. The traditional WOA algorithm includes three models: encircling prey, bubble-net attacking method and search for prey. The specific theories are as shown in literature [28]–[30].

Compared to other optimization algorithms, WOA is simple to operate and requires fewer parameters to adjust, only **A** and **C**. The adaptive change strategy of search vector **A** enables WOA to have a good balance of development and exploration capabilities and better local optimal avoidance. However, it is precisely because of the random mechanism of **A** that WOA has the disadvantages of slow convergence speed and low convergence precision. Moreover, in WOA, it is unreasonable to set the probability of contraction surrounding mechanism and the spiral position update (**A** and **C**) to 0.5. Therefore, this paper optimizes the traditional WOA to overcome the original shortcomings. The steps of the improved WOA algorithm are:

Step 1: Adaptive search surrounding mechanism and spiral position

The search surrounding mechanism is a rough search, and the spiral position is a detailed search within the current range. In the original WOA, the values of both mechanisms were 50%. However, in the early stage of the search, a broader rough search should be performed to cover the entire search space as much as possible to determine the overall direction of the optimal solution. In the later stage, a more precise search should be performed on a smaller scale and as close as possible to the true optimal solution to improve search efficiency. The adaptive search surrounding mechanism and the spiral position are defined as:

$$p_{t+1} = \begin{cases} p_0 & t = 1 \\ p_t a + p_{\min} & t > 1 \end{cases} \tag{10}$$

$$p_t^{'} = 1 - p_t \tag{11}$$

$$a = e\left[-30 \times (t/T_{\max})^S\right] \tag{12}$$

where p_0 is the initial probability of the adaptive contraction surrounding mechanism; p_t and p_{t+1} are probabilities of contraction surrounding mechanism of generation t-th and generation $(t+1)$-th, respectively; $p_{\min}$ is the minimal probability of contraction surrounding mechanism, p_t' is the probability

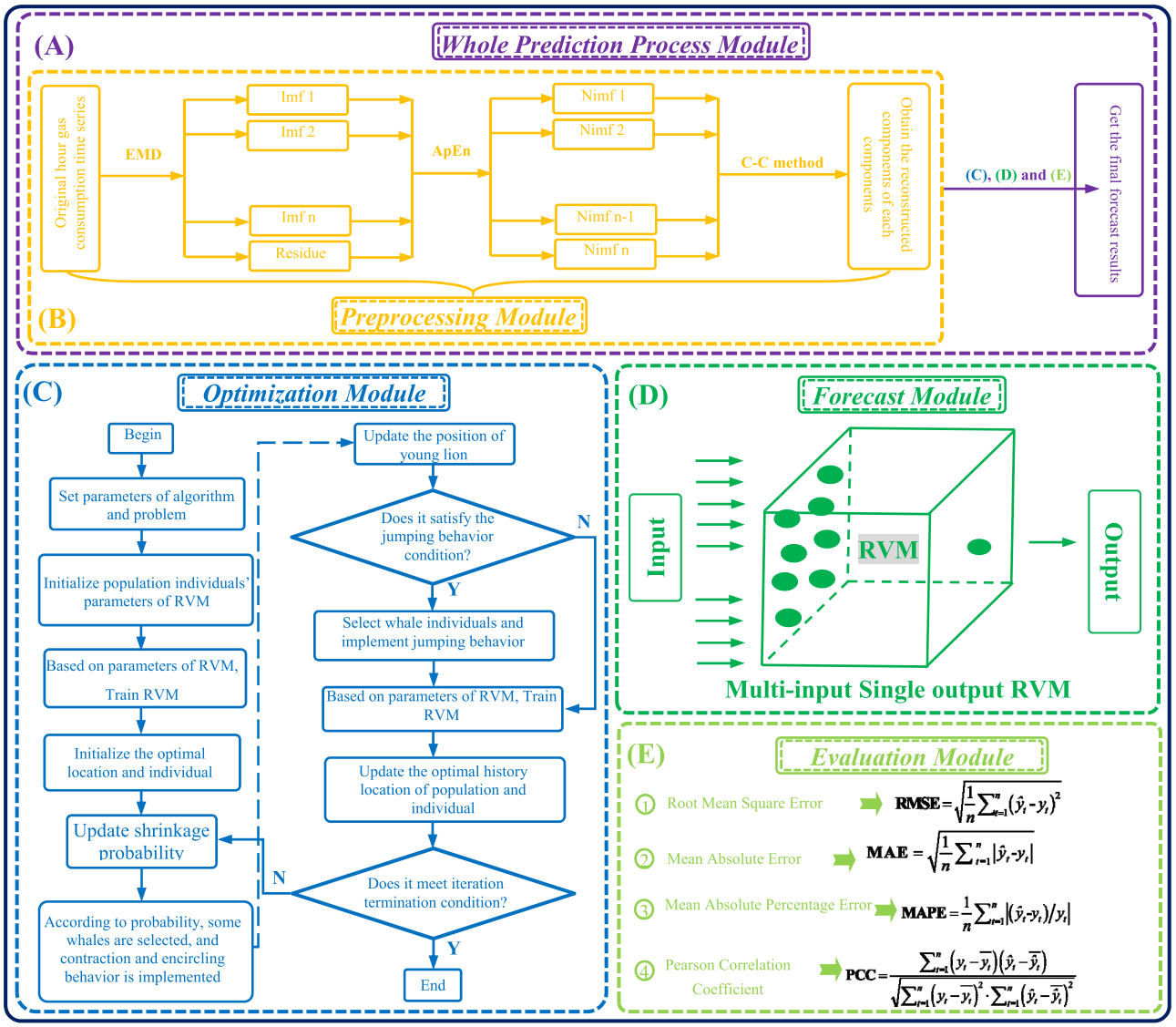

FIGURE 1. Overall forecasting model.

of updating the spiral position of the t-th generation; t is the current number of iterations; $T_{\max}$ is the maximum iterations; $S = 2$.

Step 2: Introducing jump behavior

In order to improve the search ability of WOA from various aspects, this paper introduces the jumping behavior of whales. Although whales that are some distance away from the true optimal solution may be subject to interference from local optimal solutions, the fitness value will not change. When the whale's jumping behavior is introduced, the whale tries to separate the current local optimum from the region by randomly changing the position of the whale to reach a local minimum.

The whale's jumping behavior is achieved by: after N iterations of the WOA, if the difference in the objective function value of the optimal whale is less than the present value between the two iterations, then some whales are randomly selected for jumping. The formula is:

$$X_i(t+1) = X_i(t) + c\,(1-2\mathbf{rand})\,(\max(\mathbf{X_{all}}) - \max(\mathbf{X_{all}}))/2 \quad (13)$$

where c is the jumping coefficient; $\mathbf{X_{all}}$ are all whale sets.

C. EMD, APEN AND C-C METHOD

This paper introduces EMD, ApEn and C-C method to assist in the implementation of the model. Among them, EMD can handle static signals and non-stationary signals, and has strong adaptive decomposition ability. Therefore, it has been widely used in many fields such as electric power, medicine, transportation and machinery. Its related theory and implementation methods are shown in the literature [31]. ApEn is more common in engineering applications and is an important nonlinear dynamic parameter proposed by Pincus

TABLE 1. Pseudo code of IWOA algorithm.

Input:	$y_t^{(0)}=\left(y^{(0)}(1),y^{(0)}(2),\cdots,y^{(0)}(q)\right)$——the training samples $y_r^{(0)}=\left(y^{(0)}(q+1),y^{(0)}(q+2),\cdots,y^{(0)}(q+l)\right)$——the test samples
Output	$\hat{y}_t^{(0)}=\left(\hat{y}^{(0)}(q+1),\hat{y}^{(0)}(q+2),\cdots,\hat{y}^{(0)}(q+l)\right)$——the forecast dataset
Parameters:	
	Pmin— Minimum probability of adaptive shrink wrapping mechanism
	P0— Initial probability of adaptive shrink wrapping mechanism
	Bound— Scope of X
	m— population
	T— Maximum iterations
	b— Whale algorithm constant
	s— Adaptive coefficient
1	**/* Set parameters of IWOA. */**
2	**/* Initial population. */**
3	X=initial Population ($bound_{Sstate}$, $bound_{\eta}$, $bound_{L}$, popsize)
4	**/* Calculating the fitness. */**
5	Fit=calculate Fitness(X, F_{tr}, seed, m)
6	**/* Update the optimal position of history individual. */**
7	[X_p, Fit_p] = updataXp(X, Fit)
8	**For each i in 1:T**
9	**/* Update Adaptive Probability. */**
10	a=exp(-30/(i/T)^s)
11	**If i==1**
12	Temp_p=P0
13	**Else**
14	Temp_p=Temp_p*a+ Pmin
15	**END (If)**
16	**/* Update population individuals. */**
17	**For each j in 1: Popsize**
18	**If rand<Temp p**
19	**/*Contraction encirclement. */**
20	New_X(i)=shrink(X(i))
21	**Else**
22	**/*Spiral upward. */**
23	New_X(i)=spiralism(X(i))
24	**END (if)**
25	**END (for)**
26	**/* To determine whether the algorithm has converged. */**
27	**If variation (Fit_p, Last_fit_P)< min_change && i>1**
28	**/*In the population, some X is selected randomly for jumping behavior. */**
29	New_X=jump(New_X)
30	**END (if)**
31	**/* Recalculate fitness. */**
32	New_fit = calculateFitness(New_X, Ftr, seed, m);
33	Last_fit_p=fit_p;
34	**/* Update the optimal position of history individual. */**
35	[X_p, Fit_p] = updataXp(New_X, New_fit, X_p, Fit_p);
36	**END (for)**

for measuring the complexity of time series. The smaller the sample data, the more stable the value. The approximate entropy is related to the complexity of the time series. The greater the probability of generating a new pattern, the more complex the time series, and the greater the entropy. In the predictive model, ApEn is used to re-integrate the

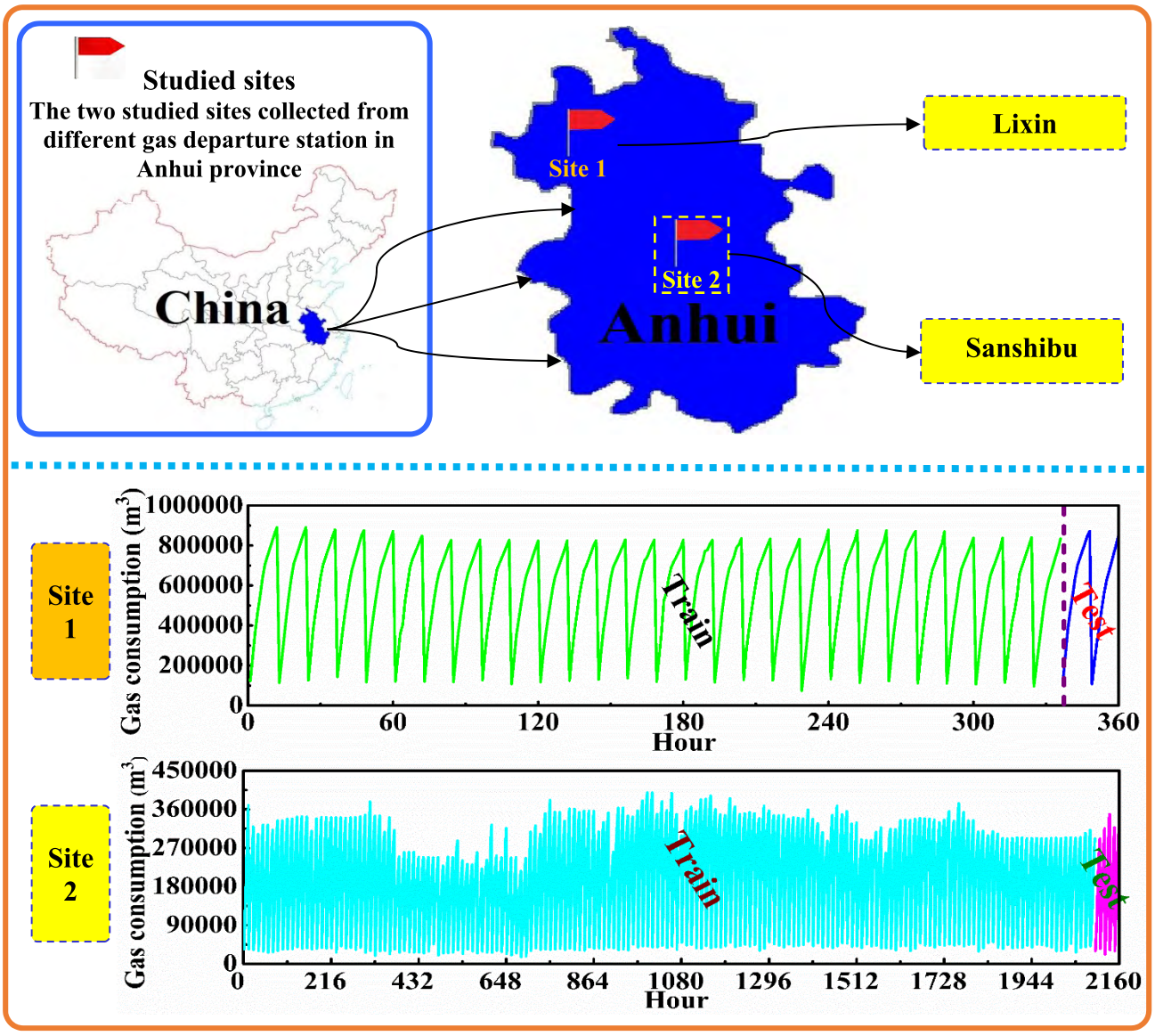

FIGURE 2. Data from two natural gas gate stations. (a) Lixin station; (b) Sanshibu station.

decomposed components and reconstruct the phase space of the re-integrated components [32]. C-C method is used to determine the optimal delay time and embedding dimension, and then the original time series can be reconstructed in phase space [33]. The prediction model architecture is shown in Fig. 1.

D. ERROR EVALUATION

In order to quantitatively evaluate the validity and accuracy of the new hybrid prediction model, RMSE (root mean square error), MAE (mean absolute error) and MAPE (average absolute error percentage) are considered respectively.

$$RSME = \sqrt{\sum_{t=1}^{N} \left(\hat{y}(t) - y(t)\right)^2 / N} \tag{14}$$

$$MAE = \sqrt{\sum_{t=1}^{N} \left|\hat{y}(t) - y(t)\right| / N} \tag{15}$$

$$MAPE = \left(\sum_{t=1}^{N} \left|\left(\hat{y}(t) - y(t)\right) / y(t)\right|\right) / N \tag{16}$$

where N is the total number of training or test set; $\hat{y}(t)$ and $y(t)$ is the forecast value and the actual value, respectively.

III. CASE STUDY

A. BASIC DATA

Data from Lixin (33.16, 116.19) and Sanshibu (31.86, 117.32) gas gate stations in Anhui Province, China are

TABLE 2. Statistical indicators of two groups of data.

Departure station	Data set	Number	Mean (m^3)	Maximum (m^3)	Minimum (m^3)	Standard deviation (m^3)
Lixin station	Training	336	574465	890500	73800	228838.4
	Test	24	568637.9	872400	107900	238499.5
	Total	360	574076	890500	73800	229156.1
Sanshibu station	Training	2100	199043.9	399022	16161	98515.06
	Test	60	182433.5	348604	22004	98938.72
	Total	2160	198582.5	399022	16161	98541.68

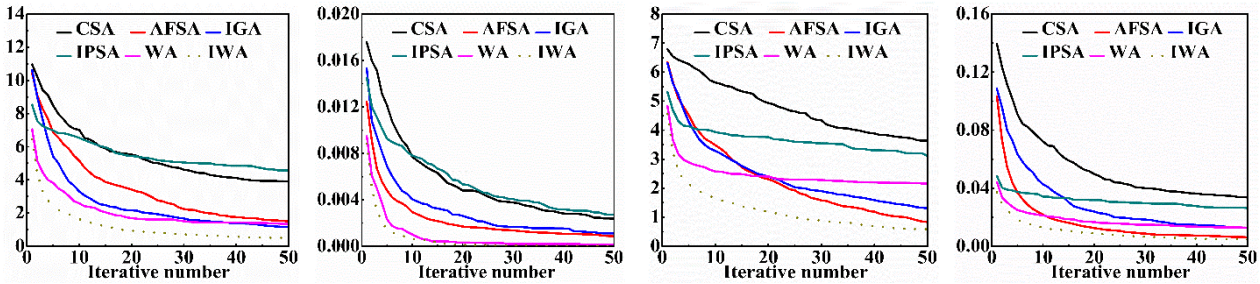

FIGURE 3. Convergence curves of six algorithms for four functions. (a) RF; (b) SF; (c) AF; (d) GF.

TABLE 3. The result comparison of four test functions by using CSA, AFSA, IGOA, IPSOA, WOA AND IWOA.

Function	Indicator	CSA	AFSA	IGA	IPSOA	WOA	IWOA
RF	AMV	3.77935	1.51303	1.13302	4.38338	1.48403	**0.51382**
	AIN	24.18	33.81	34.04	14.77	15.06	34.58
	ACT (s)	0.02086	0.00119	0.00064	0.00034	0.00049	0.00087
SF	AMV	0.00263	0.00076	0.00111	0.00241	0.00026	**0.00002**
	AIN	24.51	44.58	24.46	14.7	13.87	47.59
	ACT (s)	0.02063	0.00103	0.00076	0.00040	0.00052	0.00101
AF	AMV	3.60384	0.95467	1.46908	3.03626	2.25983	**0.52124**
	AIN	24.85	38.22	31.39	17.55	15.48	36.11
	ACT (s)	0.02169	0.00103	0.00046	0.00033	0.00047	0.00101
GF	AMV	0.02762	0.00622	0.01032	0.02603	0.01262	**0.00318**
	AIN	25.69	42.76	32.62	13.36	16.5	42.88
	ACT (s)	0.02256	0.00138	0.00076	0.00045	0.00049	0.00093

Note: (1) The bold numbers are the minimum values in the algorithm comparison. (2) AMV represents average minimum value, AIN represents average iteration number, ACT represents average calculation time.

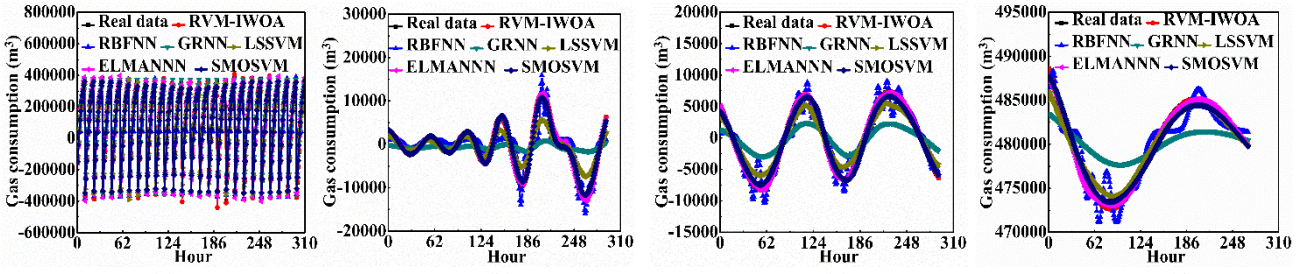

FIGURE 4. Prediction results of components of Lixin gate station training set. (a) Nimf 1; (b) Nimf 2; (c) Nimf 3; (d) Nimf 4.

collected in this paper. Lixin gate station has a small amount of data. Its training set size is 336, covering 28 days, and the test set size is 24, covering 2 days. The other set of data is large, from Sanshibu gate station, the training set size is 2040, covering 175 days, and the test set size is 120, covering 10 days. The two groups of data are shown in Fig. 2, and their statistical indicators are shown in Table 2.

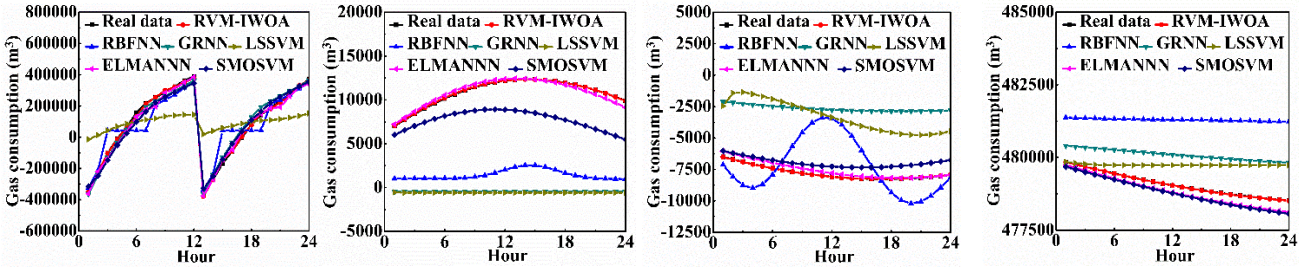

FIGURE 5. Prediction results of components of Lixin gate station test set. (a) Nimf 1; (b) Nimf 2; (c) Nimf 3; (d) Nimf.

TABLE 4. Entropy of each component.

Gate station	Original components	Entropy value	New Components
Lixin	imf 1	0.1545	Nimf 1
	imf 2	0.4584	
	imf 3	0.3122	Nimf 2
	imf 4	0.1249	Nimf 3
	imf 5	0.0565	Nimf 4
	Residue	0.0030	
Sanshibu	imf 1	0.3988	Nimf 1
	imf 2	0.3785	Nimf 2
	imf 3	0.4648	
	imf 4	0.2294	Nimf 3
	imf 5	0.0897	
	imf 6	0.0346	Nimf 4
	imf 7	0.0125	
	Residue	0.000509	

B. EXPERIMENTAL OBJECTIVE AND TOOL

This paper conducts two experiments for different objectives. The first (**experiment 1**) is to test the improved WOA, which includes convergence speed and calculation time. In this paper, four functions are used to test the optimization performance of the algorithms, the test functions include: Rastrigin function (RF), Schaffer function (SF), Ackley function (AF) and Griewank function (GF). The test results were compared with the other five algorithms, including: whale optimization algorithm (WOA), cat swarm algorithm (CSA), artificial fish swarm algorithm (AFSA), improved genetic algorithm (IGA), and improved particle swarm optimization algorithm (IPSOA). In order to ensure fairness, the six algorithms use the same experimental parameters, that is, the population size is 16 and the maximum number of iterations is 50. The specific parameters are shown in Appendix 1. In addition, 100 experiments are performed due to the randomness of the algorithm in the calculation process.

The second (**experiment 2**) is to test the accuracy of the prediction model, based on the data of the two gate stations, and compare the prediction error with the other five models, including: radical basis function neural network (RBFNN), general regression neural network (GRNN), least squares support vector machine (LSSVM), Elman neural network (ELMANNN) and sequential minimal optimization support vector machine (SMOSVM). In this experiment, considering the natural gas load has a high degree of nonlinearity with time, the EMD method is used to decompose the data into a series of imf and a residual trend term. In order to make predictions more efficiently, the complexity of each imf component and residual is evaluated by ApEn, and each imf component is combined with one remainder according to the calculation result of entropy. The parameters of EMD and APEN are shown in Appendix 2.

Before making predictions, in order to reduce the impact between different data dimensions, normalization is required and denormalization is performed before the final prediction results are generated. The formula is as follows:

$$x_{anor} = \frac{(y_{\max} - y_{\min})\,(x_{bnor} - x_{bnor\,\min})}{(x_{bnor\,\max} - x_{bnor\,\min})} + y_{\min} \quad (17)$$

where $y_{\max}$ and $y_{\min}$ is 1 and -1, respectively, x_{anor} is the data after normalization; $x_{bnor\max}$ is the maximum value before normalization; $x_{bnor\min}$ is the minimum value before normalization.

TABLE 5. Results of six algorithms for the test set of the lixin gate station (large amount of data).

Indicator	IWOA-RVM	RNFNN	GRNN	ELMANNN	LSSVM	SMOSVM
RSME (m^3)	9064.71	81604.28	30843.87	25861.79	190428.32	36382.81
MAE (m^3)	82.51	247.54	167.70	141.98	405.24	182.93
MAPE (%)	0.02	0.14	0.06	0.04	0.54	0.09

TABLE 6. Results of six algorithms for the test set of the sanshibu gate station (small amount of data).

Indicator	IWOA-RVM	RNFNN	GRNN	ELMANNN	LSSVM	SMOSVM
RSME (m^3)	2515.06	28415.94	18805.44	13055.82	97098.69	12066.91
MAE (m^3)	45.60	155.38	123.71	96.61	291.57	97.71
MAPE (%)	0.02	0.20	0.14	0.10	0.99	0.13

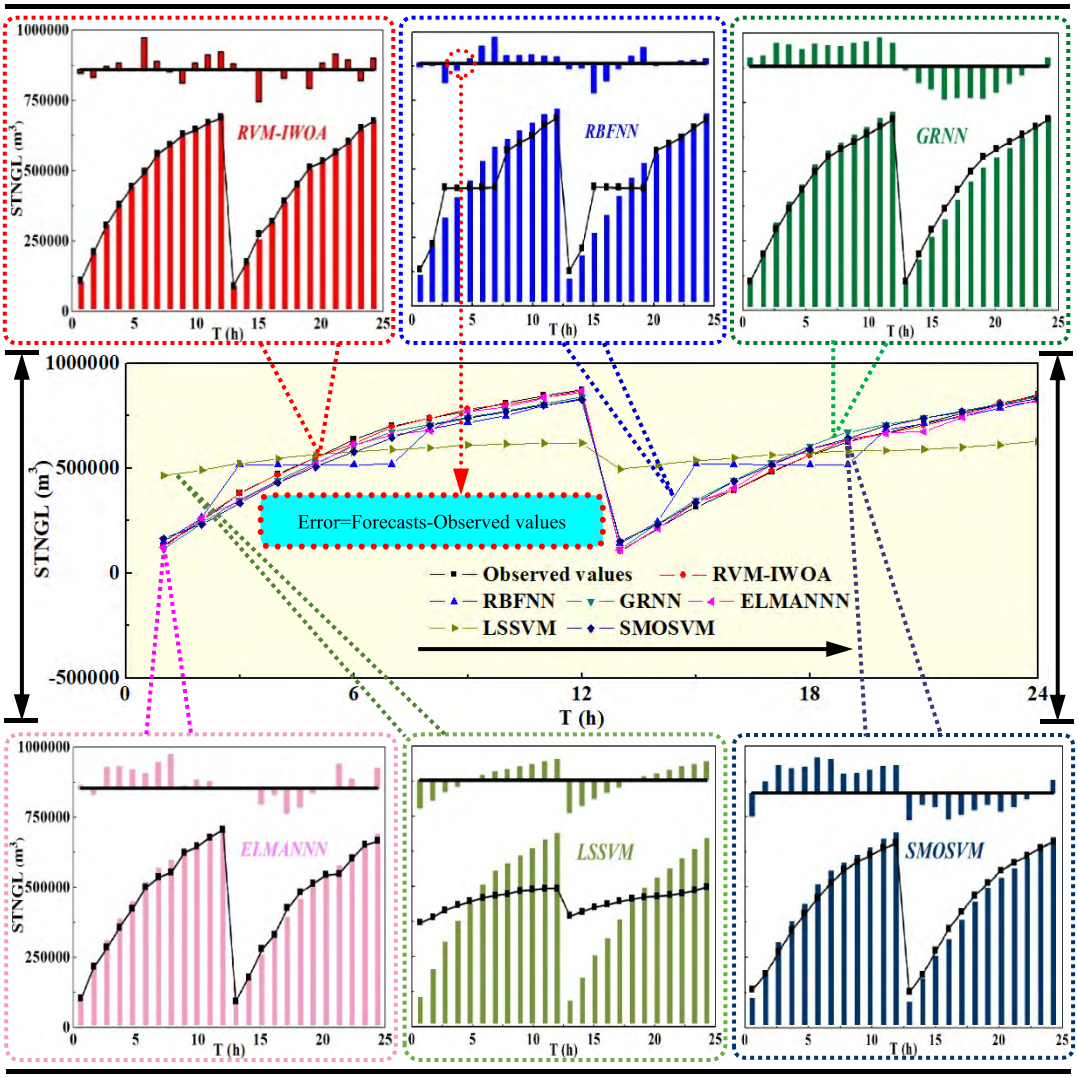

FIGURE 6. Prediction results after data integration in Lixin gate station test set.

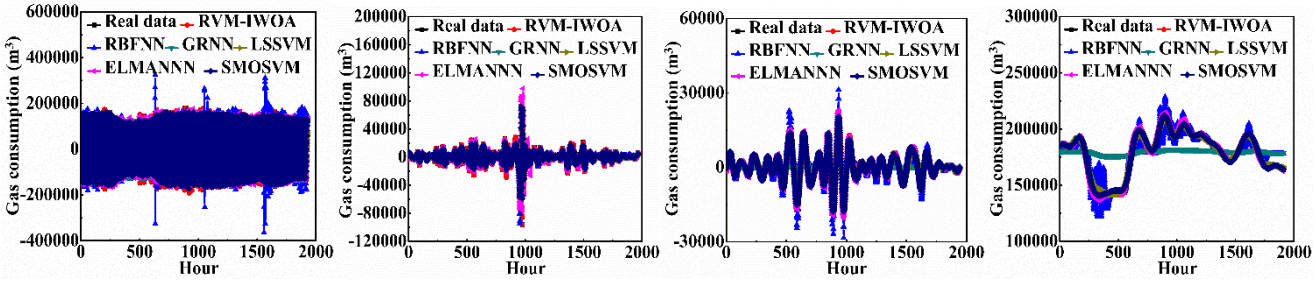

FIGURE 7. Prediction results of components of Sanshibu gate station training set. (a) Nimf 1; (b) Nimf 2; (c) Nimf 3; (d) Nimf 4.

All experiments are completed in MATLAB 2018. The computer is a Windows 10 (64 bit) operating system. It is configured as Intel Core at 2.60 GHz, i7 6700 processor, and random-access memory at 32.00 GB.

IV. EXPERIMENT RESULTS AND DISCUSSIONS

A. EXPERIMENT 1 (ALGORITHM TEST)

Fig. 3 and Table 3 show the test results of these six algorithms for the four functions. It can be seen that:

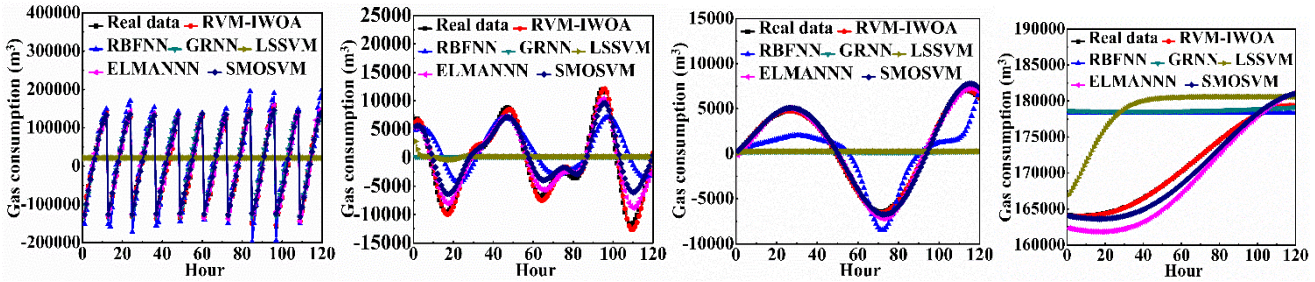

FIGURE 8. Prediction results of components of Sanshibu gate station test set. (a) Nimf 1; (b) Nimf 2; (c) Nimf 3; (d) Nimf 4.

(1) The optimal solution of the four functions obtained by the IWOA algorithm is closest to the true optimal solution, and the error in the solution of the Schaffer function is less than 0001.

(2) When the number of iterations is less than 20, the IPSOA and WOA algorithms can converge, but the obtained optimal solutions are still very different from the real optimal solutions. In this case, they can be considered to have premature convergence problems. The other four algorithms do not have this problem.

(3) IWOA's calculation time for the four functions is shorter than CSA and AFSA, but longer than IGA, IPSOA and WOA, indicating that the calculation speed is at a medium level.

(4) The convergence curve of IWOA is steeper than that of WOA, CSA, AFSA, IGA and IPSOA, and the convergence accuracy is higher than the other five algorithms.

B. EXPERIMENT 2 (PREDICTION PERFORMANCE)

The data of Lixin gate station and Sanshibu gate station are decomposed by EMD and ApEn methods. It can be seen that the approximate entropy of each instrinsic mode function (imf) component decreases with the decrease of the frequency of the imf component, which indicates that the complexity of the components from high frequency to low frequency decreases, and the approximate entropy of some adjacent imf components has little difference. Therefore, in order to reduce the computational scale, the adjacent imf components with little difference in approximate entropy are merged to form new instrinsic mode function (Nimf), and the results are shown in Table 4.

Fig. 4 and Fig. 5 show the training set and test set prediction results of each part of the Lixin gate station after data decomposition. Fig. 7 and Fig. 8 show the training set and test set prediction results of each part of the Sanshibu gate station after data decomposition. Tables 5-6 and Figs. 6, 9 summarize the error and correlation analysis of the test set data of the Lixin gate station and the Sanshibu gate station after integration. The following conclusions can be drawn:

(1) For the Lixin gate station, it can be seen from Figs. 4 (training set for Lixin gate station) and 5 (test set for Lixin gate station) that the predicted results of the RBFNN are more volatile than IWOA-RVM, ELMANNN, LSSVM, and SMOSVM.

(2) It can be seen from Table 5 (test set for Lixin gate station) and Table 6 (test set for Sanshibu gate station) that IWOA-RVM has the smallest RSME, MAE and MAPE in the six models, and has the largest PCC, indicating that the new model can make more accurate predictions when the amount of data is small or large.

(3) It can be seen from Fig. 7 (training set for Sanshibu gate station) that the RBFNN has a severe deviation at some points in the prediction of Nimf 1 and Nimf 4, while the GRNN remains unchanged in the prediction of Nimfs 2-4. However, it can be seen from Fig. 8 (test set for Sanshibu gate station) that LSSVM, RBFNN and GRNN will remain unchanged in the prediction of different data sets, that is, they are quite different from the actual data.

TABLE 7. Calculating time of gas load prediction in lixin gate station.

Component	IWOA-RVM	RBFNN	GRNN	ELMANNN	LSSVM	SMOSVM
Nimf 1	104.7	1.1	1.1	1.9	0.2	0.6
Nimf 2	107.5	1.0	1.1	3.7	0.3	0.8
Nimf 3	105.8	0.4	1.1	5.8	0.3	0.8
Nimf 4	169.5	0.6	1.1	6.3	0.3	0.7

TABLE 8. Calculating time of gas load prediction in sanshibu gate station.

Component	IWOA-RVM	RBFNN	GRNN	ELMANNN	LSSVM	SMOSVM
Nimf 1	14032.9	10.8	6.4	15.5	0.9	1.7
Nimf 2	10350.1	3.9	4.4	3.5	0.6	0.9
Nimf 3	4193.2	8.8	3.2	12.0	0.8	1.1
Nimf 4	3209.9	2.8	2.0	6.0	0.6	1.0

C. DISCUSSION ON MODEL EFFICIENCY

From IV.B, it is concluded that the IWOA-RMV model has higher prediction accuracy. This paper also needs to discuss

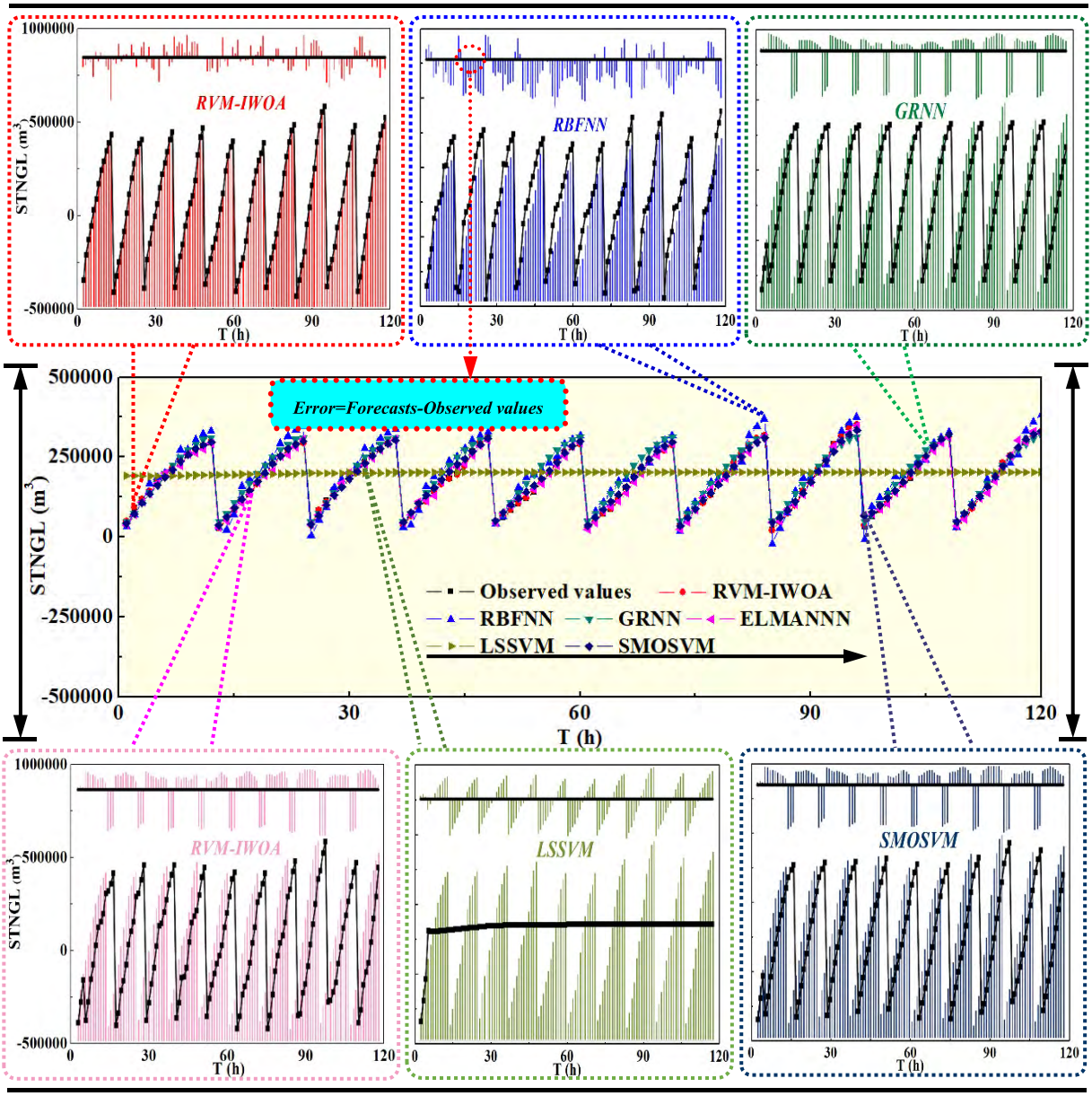

FIGURE 9. Prediction results after data integration in Sanshibu gate station test set.

TABLE 9. Test functions.

RF	SF
$Ras(\mathbf{x}) = 20 + x_1^2 + x_2^2 - 10(\cos 2\pi x_1 + \cos 2\pi x_2)$	$Sch(\mathbf{x}) = 0.5 + \left(\sin^2\left(x_1^2 - x_2^2\right) - 0.5\right) \Big/ \left(\left[1 + 0.001\left(x_1^2 + x_2^2\right)\right]^2\right)$
where: x_1 is belongs to [-5, 5] and x_2 is belongs to [-5, 5].	where: x_1 is belongs to [-5, 5] and x_2 is belongs to [-5, 5].
AF	**GF**
$Ack(\mathbf{x}) = -a\exp\left(-b\sqrt{\frac{1}{d}\sum_{i=1}^{d} x_i^2}\right) - \exp\left(\frac{1}{d}\sum_{i=1}^{d}\cos(cx_i)\right) + a + \exp(1)$	$Gre(\mathbf{x}) = \sum_{i=1}^{N}\frac{x_i^2}{4000} - \prod_{i=1}^{N}\cos\left(x_i/\sqrt{i}\right) + 1$
where: i is equal to 1 and 2, a=20, b=0.2, c=2π, d=2, x_1 is belongs to [-10, 10] and x_2 is belongs to [-10, 10].	where: i is equal to 1 and 2, x_1 is belongs to [-10, 10] and x_2 is belongs to [-10, 10].

the efficiency of the model, because the computational efficiency is also a very important factor in the prediction. Tables 7 and 8 show the calculating time required for the prediction of natural gas load at two gate stations. It can be seen that the computating time of IWOA-RVM model is much longer than that of the other five algorithms, but this time (several hours) is acceptable in engineering.

TABLE 10. The parameters in 5.2.

Algorithm and function	Parameter setting	Value	Parameter setting	Value
IWOA	maximum iterations	50	population	16
	jumping behavior probability	0.1	b	1
	minimum error of jumping	0.0001		
WOA	maximum iterations	50	population	16
	b	1		
COA	maximum iterations	50	population	16
	step size control variable	0.2	inertia weight	5
	beta	2	sigma	0.6966
FOA	maximum iterations	50	population	16
	View	0.2	step	0.8
	predation attempts	10		
IGOA	maximum iterations	50	population	16
	cross probability	0.8	mutation probability	0.1
IPSOA	maximum iterations	50	population	16
	learning factor 1	1	learning factor 2	1
	maximum weight	0.8	minimum weight	0.5

V. CONCLUSIONS AND FUTURE WORK

This paper proposes a model that combines IWOA and RVM for short-time gas demand prediction. The improved WOA algorithm solves the problem of premature convergence of the original WOA and can obtain the global optimal solution. Through the algorithm test and the comparison of prediction results, the following conclusions are drawn: (1) Through the test of four functions, IWOA has better ability to jump out of local optimum, has higher optimization performance, and the calculation speed is at a medium level. (2) Compared with other models, the IOWA-RVM model has higher prediction accuracy when the amount of data is larger or smaller, but the calculation time is relatively longer, but the calculation time is acceptable in engineering.

From the research of this paper, it is found that although the calculation time of the proposed algorithm can meet the engineering requirements, it is still necessary to focus on improving the computational efficiency in future work. In addition, although the model proposed in this paper has higher prediction accuracy in the case of larger and smaller data amount, it does not discuss the most suitable data amount range, which is also the future work content.

APPENDIX

APPENDIX 1

See Table. 9.

APPENDIX 2

See Table. 10.

REFERENCES

[1] H. Lu, K. Huang, M. Azimi, and L. Guo, "Blockchain technology in the oil and gas industry: A review of applications, opportunities, challenges, and risks," *IEEE Access*, vol. 7, pp. 41426–41444, 2019. doi: 10.1109/ACCESS.2019.2907695.

[2] E. Liu, L. Lv, Q. Ma, J. Kuang, and L. Zhang, "Steady-state optimization operation of the west–east gas pipeline," *Adv. Mech. Eng.*, vol. 11, no. 1, 2019, Art. no. 1687814018821746.

[3] B. Soldo, P. Potočnik, G. Šimunović, T. Šarić, and E. Govekar, "Improving the residential natural gas consumption forecasting models by using solar radiation," *Energy Buildings*, vol. 69, no. 3, pp. 498–506, Feb. 2014.

[4] F. Taşpinar, Çelebi, and N. Tutkun, "Forecasting of daily natural gas consumption on regional basis in Turkey using various computational methods," *Energ Buildings*, vol. 56, no. 1, pp. 23–31, 2013.

[5] M. Brabec, O. Konár, E. Pelikán, and M. Malý, "A nonlinear mixed effects model for the prediction of natural gas consumption by individual customers," *Int. J. Forecasting*, vol. 24, no. 4, pp. 659–678, 2008.

[6] B. C. Ervural, O. F. Beyca, and S. Zaim, "Model estimation of ARMA using genetic algorithms: A case study of forecasting natural gas consumption," *Procedia-Social Behav. Sci.*, vol. 235, pp. 537–545, Nov. 2016.

[7] A. Azadeh, S. M. Asadzadeh, and A. Ghanbari, "An adaptive network-based fuzzy inference system for short-term natural gas demand estimation: Uncertain and complex environments," *Energy Policy*, vol. 38, no. 3, pp. 1529–1536, 2010.

[8] C. Zhang, Y. Liu, and H. Zhang, "Research on short-term gas load forecasting based on support vector machine model," in *Proc. Int. Conf. Life Syst. Modeling Simulation*, Berlin, Germany, 2010, pp. 390–399.

[9] P. Potočnik, B. Soldo, G. Šimunović, T. Šarić, A. Jeromen, and E. Govekar, "Comparison of static and adaptive models for short-term residential natural gas forecasting in Croatia," *Appl. Energy*, vol. 129, no. 129, pp. 94–103, 2014.

[10] R. Kizilaslan and B. Karlik, "Comparison neural networks models for short term forecasting of natural gas consumption in Istanbul," in *Proc. 1st Int. Conf. Appl. Digit. Inf. Web Technol. (ICADIWT)*, Ostrava, Czech Republic, Aug. 2008, pp. 448–453.

[11] H. Zhou, G. Su, and G. Li, "Forecasting daily gas load with OIHF-Elman neural network," *Procedia Comput. Sci.*, vol. 5, pp. 754–758, Jan. 2011.

[12] D. Chaudhari, A. R. Joshi, and N. Gera, "Short term natural gas demand forecasting using neural networks," in *Proc. Int. Conf. New Vistas Entrepreneurial Endeavors*. 2012.

[13] A. Azari, M. Shariaty-Niassar, and M. Alborzi, "Short-term and medium-term gas demand load forecasting by neural networks," *Iranian J. Chem. Chem. Eng.*, vol. 31, no. 4, pp. 77–84, 2012.

[14] W. Y. Shi, "Research on gas load forecasting using artificial neural network," *Adv. Mater. Res.*, vol. 717, pp. 423–427, Jul. 2013.

[15] T. Fu, "Gas short-term load forecasting based on BP AdaBoost model," *B Sci. Technol. Soc.*, vol. 29, no. 10, pp. 55–57, 2013.

[16] H. Lu, M. Azimi, and T. Iseley, "Short-term load forecasting of urban gas using a hybrid model based on improved fruit fly optimization algorithm and support vector machine," *Energy Rep.*, vol. 5, pp. 666–677, Nov. 2019.

[17] L. Zhu, M. S. Li, Q. H. Wu, and L. Jiang, "Short-term natural gas demand prediction based on support vector regression with false neighbours filtered," *Energy*, vol. 80, pp. 428–436, Feb. 2015.

[18] Y. Bai and C. Li, "Daily natural gas consumption forecasting based on a structure-calibrated support vector regression approach," *Energ Buildings*, vol. 127, pp. 571–579, Sep. 2016.

[19] Z. Liu, W. Li, and W. Sun, "A novel method of short-term load forecasting based on multiwavelet transform and multiple neural networks," *Neural Comput. Appl.*, vol. 22, no. 2, pp. 271–277, 2013.

[20] H. Karimi and J. Dastranj, "Artificial neural network-based genetic algorithm to predict natural gas consumption," *Energy Syst.*, vol. 5, no. 3, pp. 571–581, Sep. 2014.
[21] F. Yu and X. Xu, "A short-term load forecasting model of natural gas based on optimized genetic algorithm and improved BP neural network," *Appl. Energy*, vol. 134, no. 134, pp. 102–113, Dec. 2014.
[22] F. Yu, Z. Q. Wang, and X. Z. Xu, "Short-term gas load forecasting based on wavelet BP neural network optimized by genetic algorithm," *Appl. Mech. Mater.*, vols. 631–632, pp. 79–85, Sep. 2014.
[23] J. Szoplik, "Forecasting of natural gas consumption with artificial neural networks," *Energy*, vol. 85, pp. 208–220, Jun. 2015.
[24] X. Ma, X. Mei, W. Wu, X. Wu, and B. Zeng, "A novel fractional time delayed grey model with Grey Wolf optimizer and its applications in forecasting the natural gas and coal consumption in Chongqing China," *Energy*, vol. 178, pp. 487–507, Jul. 2019. doi: 10.1016/j.energy.2019.04.096.
[25] M. Akpinar, M. F. Adak, and N. Yumusak, "Forecasting natural gas consumption with hybrid neural networks—Artificial bee colony," in *Proc. 2nd Int. Conf. Intell. Energy Power Syst. (IEPS)*, Jun. 2016, pp. 1–6.
[26] I. P. Panapakidis and A. S. Dagoumas, "Day-ahead natural gas demand forecasting based on the combination of wavelet transform and ANFIS/genetic algorithm/neural network model," *Energy*, vol. 118, pp. 231–245, Jan. 2017.
[27] S. Mirjalili and A. Lewis, "The whale optimization algorithm," *Adv. Eng. Softw.*, vol. 95, pp. 51–67, May 2016.
[28] I. Aljarah, H. Faris, and S. Mirjalili, "Optimizing connection weights in neural networks using the whale optimization algorithm," *Soft Comput.*, vol. 22, no. 1, pp. 1–15, 2016.
[29] M. Mafarja and S. Mirjalili, "Whale optimization approaches for wrapper feature selection," *Appl. Soft Comput.*, vol. 62, pp. 441–453, Jan. 2018.
[30] H. H. Mehne and S. Mirjalili, "A parallel numerical method for solving optimal control problems based on whale optimization algorithm," *Knowl.-Based Syst.*, vol. 151, pp. 114–123, Jul. 2018.
[31] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proc. Roy. Soc. London. A, Math., Phys. Eng. Sci.*, vol. 454, no. 1971, pp. 903–995, Mar. 1998.
[32] S. M. Pincus and R. A. Neidorff, "Approximate entropy," U.S. Patent 5 191 524 A, Mar. 2, 1993.
[33] L. Cao, "Practical method for determining the minimum embedding dimension of a scalar time series," *Phys. D, Nonlinear Phenomena*, vol. 110, nos. 1–2, pp. 43–50, Dec. 1997.