

Received March 5, 2021, accepted March 23, 2021, date of publication March 29, 2021, date of current version April 9, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3069280

# Optimal Variational Mode Decomposition and Integrated Extreme Learning Machine for Network Traffic Prediction

JINMEI SHI<sup>1,2</sup>, (Member, IEEE), YU-BENG LEAU<sup>1</sup>, (Member, IEEE),  
KUN LI<sup>3</sup>, AND HUANDONG CHEN<sup>4</sup>

<sup>1</sup>Faculty of Computing and Informatics, Universiti Malaysia Sabah, Kota Kinabalu 88400, Malaysia

<sup>2</sup>College of Information Engineering, Hainan Vocational University of Science and Technology, Haikou 571126, China

<sup>3</sup>Faculty of Electrical and Control Engineering, Liaoning Technical University, Liaoning 125105, China

<sup>4</sup>College of Information Science and Technology, Hainan Normal University, Haikou 571158, China

Corresponding authors: Yu-Beng Leau (lybeng@ums.edu.my) and Kun Li (bhulikun@163.com)

This work was supported by Universiti Malaysia Sabah and Hainan Vocational University of Science and Technology.

**ABSTRACT** Network traffic prediction plays a vital role in effective network management, load evaluation and security warning. Extreme learning machine has the advantages of fast convergence speed and strong generalization ability. Also, it does not easily fall into local optima. The evolutionary algorithm can be used to optimize the number of its hidden layer nodes. However, most of the existing evolutionary algorithms have some adjustable parameters which depend on subjective experience or prior knowledge. Hence, this can affect the model prediction accuracy. Given this context, this paper proposes a network traffic prediction mechanism based on optimized Variational Mode Decomposition (VMD) and Integrated Extreme Learning Machine (ELM). A Scalable Artificial Bee Colony (SABC) algorithm which has fewer adjustable parameters and can thus guarantee the accuracy and stability of the prediction mechanism is also proposed. It can be used in the optimization selection of VMD, Phase Space Reconstruction (PSR) and ELM to achieve higher prediction performance. Finally, we utilize Mackey-Glass, Lorenz chaotic time series of recognized benchmark and a WIDE backbone actual network traffic data to prove the validity of the proposed network traffic prediction mechanism.

**INDEX TERMS** Network traffic, time series analysis, variational mode decomposition, extreme learning machine, artificial bee colony algorithm.

## I. INTRODUCTION

### A. BACKGROUND

Network traffic is the amount of data transmitted over the network. The size of network traffic is of great significance to the design of network architecture. With the continuous expansion of the Internet structure, various network types have consistently emerged in recent years. The constant improvement of the scale of network users has significantly increased the network traffic data, which in turn bring new challenges to the effective management, load evaluation and security warning of the network. Network traffic is time-related data with chaotic characteristics, which can be transformed into a time series problem [1]. Network traffic prediction is to use

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

the network traffic data of the past period to predict the data of a specific moment in the future to grasp the development trend of the incoming network and improve the management ability of network load.

At the beginning of the development of the computer network, the amount of data transmitted by the network was small. Hence, Poisson model and Markov model could be effectively used to describe network traffic data and establish prediction model. In the 1990s, Leland [2] *et al.* found that self-similar characteristics existed in traffic data; a finding which opened a new chapter of traffic prediction. Some self-similar and linear models were thus used to model and predict network traffic data. Nowadays, with the rapid development of network scale, network traffic proved to be a typical nonlinear time series with the characteristics of time-varying, long correlation, self-similarity, sudden, chaotic and

so on [3], [4]. Given that, the traditional linear models are unable to accurately describe network traffic characteristics, leading to poor prediction performance and low accuracy of prediction problems. Thus, researchers shifted from linear model to nonlinear model based on Neural Network (NN), Chaotic Theory (CT) and Statistical Theory (ST), particularly the nonlinear analysis which is mainly based on Neural Network. With the recent research development, the problems of some neural network models were gradually exposed. These issues include gradient explosion of Back Propagation neural network [5], slow training speed and low generalization ability of Radial Basis Function neural network [6], more adjustable parameters of ESN network [7] and so on. In light of the aforementioned, a type of feed forward neural network, Extreme Learning Machine (ELM), was proposed by Huang *et al.* The ELM model has the advantages of fast convergence speed, strong generalization ability, and not easily falling into the local optima, etc [8]. Hence, it is not surprising that in recent years, time series-based prediction model of ELM has attracted wide research attention. This resulted in an ELM network traffic prediction model based on Map Reduce, proposed by Liu *et al.* [9]. Tian *et al.* [10] introduced an improved Artificial Bee Colony algorithm to optimize the number of nodes in the hidden layer of ELM for network traffic prediction. In the same year, he proposed a prediction model based on Empirical Mode Decomposition and ELM [11]. In 2017, Li *et al.* [12] put forward an OS-ELM prediction model based on down-hole working condition. In the following year, Li *et al.* advocated a time series decomposition and optimization of ELM short-term wind speed prediction method [13].

Along with the complexity of network traffic data in the current large network, the randomness and non-stationarity could make the actual data have strong nonlinearity, which would decrease the computational accuracy of the prediction model. The time-series decomposition (abbreviated by TSD) analysis seems to be an effective method to handle this problem [13]. In recent years, some TSD methods, such as: empirical mode decomposition (EMD) [14]–[16], local mean decomposition (LMD) [17], variational mode decomposition (VMD) [18]–[20], etc. have caught many scholar's attentions. Variational Mode Decomposition (VMD) is a signal processing method which transfers the signal decomposition process adaptively to a variational framework. It then realizes the signal decomposition by searching for the optimal solution of the constrained variational model [21]. VMD splits the signal into several modal components to minimise the sum of estimated bandwidth of each mode. VMD is an improved version of EMD which has better robustness for the data decomposition [22]; compared to EMD, VMD has a superior denoising property and ability to separate tones of similar frequencies [23]; compared to wavelet decomposition (WD) and EMD, VMD is foremost in decomposing signals non-recursively by adaptively determining the frequency center of the processing data and thereby optimally fixing each subcomponents limit [24].

Moreover, in our research, the Phase Space Reconstruction (PSR) is employed to reflect the real change of system by the original data sequence. The dynamic change process of any variables in the system can reflect the influence of other variables related to it, and the phase space obtained through reconstruction can reflect the change rule of the system state. So, after VMD decomposition, the original signal sequence is decomposed into  $K$  sub-sequence; before ELM modeling each sub-sequence, PSR can be carried out to construct the input and output variables of the ELM model.

In the combination of VMD, PSR and ELM, there is one crucial problem needs to be better solved. That is, the value of some tuning parameters in the model should be accurately set, such as: the number of modes and the iterative factor in VMD, the embedding dimension and the delay time in PSR, and the number of hidden layer nodes in ELM. In some applications, the value of these adjustable parameters is highly dependent on subjective experience or prior knowledge. Hence, when the setting is not reasonable it will negatively affect the optimization result of model parameters and the calculation accuracy and stability of the model.

Evolutionary Computation is a research direction of Computational Intelligence that involves combinational optimization. Its algorithm is influenced by the natural selection mechanism of "survival of the fittest" and the transmission law of genetic information in the process of biological evolution. It simulates this process through program iteration, treats the problem to be solved as the environment, and seeks for the optimal solution through a natural evolution in a population composed of some possible solutions. At present, Particle Swarm Optimization (PSO) [25], improved Artificial Bee Colony algorithm (IABC) [10], improved Harmony Search algorithm (IHS) [26], improved Free Search algorithm (IFS) [27], improved Gravity Search algorithm (IGSA) [13], Levy-Cloud-Model Fruit Fly Optimization algorithm (LVCMFOA) [28] and other different types of evolutionary algorithms have attracted the attention of researchers. ABC algorithm is a swarm intelligence optimization algorithm which imitates the honey gathering mechanism of bees. Its principle and implementation are relatively simple so as to it has been studied and applied in many fields. ABC algorithm [29] involves three kinds of bee colonies, namely, lead bee, follow bee and scout bee. One leading bee corresponds to a food source and searches the food source first. Follow bees are to select dancing areas and visit food sources. Scout bees conduct random searches of food sources. When the food source is exhausted, the lead bee for the intended food source is transformed into a scout bee.

## B. RELATED WORKS

In this paper, in order to decrease the computation amount when conducting many repeated times' VMD decomposition and ELM modeling, an improved Artificial Bee Colony algorithm, named Scalable Artificial Bee Colony (SABC) algorithm, is proposed to ensure the reliability of the selection of different tuning parameters. In this regard, this study

optimizes the VMD and an integrated ELM to establish the prediction model mechanism. The study also utilizes the SABC algorithm to realize the optimization selection of VMD, PSR and various adjustable parameters in ELM. The proposed prediction mechanism can be applied to network traffic data prediction to ensure that the prediction results are less affected by adjustable parameters and have higher stability and prediction accuracy.

In recent years, the time series prediction based on VMD and ELM has caught many scholars' attentions. Literature [30] used Robust Kernel based Extreme Learning Machine (RKELM) integrated with VMD to predict stock price and movement, but they did not consider the phase space reconstruction of each subset and the optimized selection of two control parameters. Literature [31] combined the VMD, Singular Spectrum Analysis (SSA), Long Short Term Memory (LSTM) network and ELM to forecast wind speed, but many tuning parameters of the VMD, ELM were not optimized for better choice. In Literature [23], VMD and a new low rank robust kernel based Extreme Learning Machine (RKELM) were combined to forecast solar irradiation, but they did not consider the optimized selection of different tuning parameters. In Literature [32], a hybrid mode decomposition (HMD) method (comprised of VMD, sample entropy (SE) and wavelet packet decomposition (WPD)) and online sequential outlier robust extreme learning machine (OSORELM) was proposed to predict wind speed, but they did not use the phase space reconstruction to reflect characteristics of the real system, and meanwhile the optimized selection of different control parameters was not conducted. Literature [24], proposed a combination of an adaptive regularized extreme learning machine (ARELM) and an improved VMD, and used the Ant Colony Optimization (ACO) algorithm to select the optimal value of some tuning parameters, but they did not consider a synchronous optimization of several control parameters in VMD, PSR and ELM. In Literature [22], VMD and an optimized outlier-robust ELM was used for point and interval forecasting of metal prices, but GWO was only used to optimize several tuning parameters in DFs and ORELM, and a synchronous optimization of several control parameters in VMD, PSR and ELM was not considered.

From the above analysis, main contributions of our work are summarized as follows:

- (1) The synchronous optimization of VMD, PSR and ELM parameter is used to improve the calculation accuracy and stability of the prediction model.
- (2) A Scalable Artificial Bee Colony (SABC) algorithm is designed to improve the convergence accuracy and speed of the algorithm through the new solution generation mechanism and the addition of fine-tuning disturbance to the new solution.

The rest of our works is organized as follows. Section II has a brief overview of VMD, ELM and PSR. In Section III, main steps of the classical Artificial Bee Colony algorithm are firstly given, and then a new Scalable Artificial Bee

Colony (SABC) algorithm is introduced, including some improvements. Section IV gives the proposed prediction mechanism for the network traffic data, in which a synchronous optimization of VMD, ELM and PSR is used. In Section V, the validity of the proposed mechanism is verified using three datasets. Finally, Section VI concludes our works.

## II. NETWORK TRAFFIC PREDICTION MODEL

### A. VARIATIONAL MODE DECOMPOSITION METHOD

According to VMD [21], its constrained variational model is:

$$\left\{ \begin{array}{l} \min_{\{u_K\}, \{\omega_K\}} \left\{ \sum_K \left\| \partial_t [(\delta(t) + \frac{j}{\pi t}) * u_K(t)] e^{-j\omega_K t} \right\|_2^2 \right\} \\ \text{s.t.} \quad \sum_K u_K = f \end{array} \right. \quad (1)$$

where,  $u_K$  is  $K$  modal components,  $\omega_K$  is the centre frequency of each modal component,  $\delta(t)$  is the impulse function.  $\|\cdot\|_2^2$  denotes the Euclidean norm.

To address the limitations of the constrained variational model as mentioned earlier, two parameters are introduced; they are quadratic penalty factor  $\alpha$  and Lagrange multiplier  $\lambda$ . The constrained variational problem is transformed into an unconstrained variational problem. The transformed unconstrained variational model is:

$$\begin{aligned} L(\{u_K\}, \{\omega_K\}, \lambda) : \\ = \alpha \sum_K \left\| \partial_t [(\delta(t) + \frac{j}{\pi t}) * u_K(t)] e^{-j\omega_K t} \right\|_2^2 \\ + \left\| f(t) - \sum_K u_K(t) \right\|_2^2 + \left\langle \lambda(t), f(t) - \sum_K u_K(t) \right\rangle \end{aligned} \quad (2)$$

The main steps for solving the unconstrained variational model of Equation (2) are as follows:

Step 1: Initialise each modal component, centre frequency and Lagrange multiplier, denoised as:  $u_K^1, \omega_K^1$  and  $\lambda^1$ . Let the counter  $n = 0$ .

Step 2: Transform each variable from the time domain to the frequency domain. For the  $n+1$  count, in the non-negative frequency interval,  $K$  modal component  $u_K$  is updated as follows:

$$\hat{u}_K^{n+1}(\omega) \leftarrow \frac{\hat{f}(\omega) - \sum_{i \neq K} \hat{u}_i(\omega) + \frac{\hat{\lambda}(\omega)}{2}}{1 + 2\alpha(\omega - \omega_K)^2} \quad (3)$$

where,  $\hat{u}_K(\omega)$ ,  $\hat{f}(\omega)$  and  $\hat{\lambda}(\omega)$  are the Fourier transforms of  $u_K, f(t)$  and  $\lambda$ , respectively.

Step 3: For the  $n+1$  count, update the centre frequency  $\omega_K$  of each modal component as follows:

$$\omega_K^{n+1} \leftarrow \frac{\int_0^\infty \omega |\hat{u}_K(\omega)|^2 d\omega}{\int_0^\infty |\hat{u}_K(\omega)|^2 d\omega} \quad (4)$$

Step 4: For the  $n + 1$  count, update the Lagrange multiplier as follows:

$$\hat{\lambda}^{n+1}(\omega) \leftarrow \hat{\lambda}^n(\omega) + \tau_1(\hat{f}(\omega) - \sum_K \hat{u}_K^{n+1}(\omega)) \quad (5)$$

Among them,  $\tau_1$  is the iteration factor.

Step 5: Set the convergence accuracy  $\varepsilon > 0$ , iteration termination conditions as follows:

$$\frac{\sum_K \left\| \hat{u}_K^{n+1} - \hat{u}_K^n \right\|_2^2}{\left\| \hat{u}_K^n \right\|_2^2} < \varepsilon \quad (6)$$

If equation (6) is true, the iteration is stopped and the result is output. Otherwise, return to Step 2 to continue.

The VMD method involves two main controllable parameters: the number of modes  $K$  and the iterative factor  $\tau_1$ . The values of the signals need to be pre-set before they can be decomposed. For the purpose of reducing the influence of artificial subjective experience and prior knowledge onset values, an optimization method is used to realize their adaptive selection to adapt to the decomposition of different signals. Therefore, after the signal is decomposed into the optimized number of  $K$  modes, the ELM method is then introduced to model each mode.

### B. EXTREME LEARNING MACHINE

Extreme Learning Machine (ELM) is a neural Network training method and the purpose is to solve the back propagation algorithm learning efficiency which is low and the parameter setting of complicated issues. ELM can better improve the learning speed of network structure, which avoids the problems of local minimum, iteration number and performance index. Determine the training sample set as  $\{(x_i, y_i)\}$ ,  $i = 1, 2, \dots, N$ ,  $N$  is the number of samples in the sample set. The mathematical model of the Extreme Learning Machine (ELM) can be described by Equation (7):

$$t_i = \sum_{j=1}^L \beta_{ji} g(\omega_j, b_j, x_i) \quad (7)$$

where,  $t_i$  is the  $i^{\text{th}}$  training output,  $\beta_{ji}$  is the connection weight between the hidden layer neuron and the  $i^{\text{th}}$  output neuron,  $\omega_j$  is the connection weight between the hidden layer neuron and the input neuron,  $b_j$  is the bias of the  $j^{\text{th}}$  hidden layer neuron,  $L$  is the number of hidden layer node;  $g(\cdot)$  is the activation function.

Equation (7) can be expressed as the following matrix form:

$$t = G \cdot \beta \quad (8)$$

where  $t = [t_1, \dots, t_N]^T$ ,  $g_l = [g(\omega_1, b_1, x_1), \dots, g(\omega_N, b_N, x_N)]^T$ ,  $G = [g_1 | \dots | g_L]^T$ , and  $\beta = [\beta_1, \dots, \beta_L]^T$ .  $G$  is the hidden layer output matrix, connection weight  $\omega$  and bias  $b$  are randomly given when the system is initialized, and remain unchanged in the training process.

The purpose of ELM training is to find the optimal output weight matrix  $[\beta_1 \ \beta_2 \ \dots \ \beta_L]^T$  so that the following formula is true.

$$\|t - y\| = \min_{\beta} \|t - y\| \quad (9)$$

where  $y = [y_1, \dots, y_N]^T$  and  $\|\cdot\|$  denotes the Euclidean norm.

By solving equation (9), the output weight matrix  $\hat{\beta}$  is obtained as follows:

$$\hat{\beta} = G^+ \cdot y \quad (10)$$

where  $G^+$  is the pseudo-inverse of the output matrix of the hidden layer.

Since  $\omega$  and  $b$  do not change during the model training, the output of hidden layer is only related to the input  $x_i$  and the number of hidden layer nodes  $L$ . The output of the ELM model can be described by the following formula:

$$y = \sum_{i=1}^L \hat{\beta}_i \cdot g(\omega_i, b_i, x) \quad (11)$$

The output accuracy of ELM model is closely related to the number of hidden layer nodes  $L$ . ELM modeling is conducted for the  $K$  modes after VMD decomposition, involving the setting of the number of  $K$  hidden layer nodes  $L$ . Therefore, an optimization method is suggested to realize the adaptive selection in order to minimize the influence of artificial subjective experience and prior knowledge onset values.

### C. PHASE SPACE RECONSTRUCTION

After VMD decomposition, the original signal sequence is decomposed into  $K$  sub-sequence. Before ELM modeling each sub-sequence, Phase Space Reconstruction (PSR) can be carried out to construct the input and output variables of the ELM model. Let the chaotic time series be  $\{x_i, i = 1, 2, \dots, N\}$ , where  $N$  is the number of samples in the sequence. Then, the input variables after the reconstruction of the phase space can be expressed in the following form [33]:

$$x(N\tau_2) = (x(N\tau_2), \ x((N-1)\tau_2), \ \dots, \ x((N-m)\tau_2)) \quad (12)$$

And the training set can be expressed by

$$\{(x((N-1)\tau_2), \ x(N\tau_2)) : N = m + 1, \dots, N_0\} \quad (13)$$

where  $x((N-1)\tau_2)$  is the input vector and  $x(N\tau_2)$  is the target,  $N_0$  is an arbitrary value of  $N$ .

Among them,  $m$  is the embedding dimension and  $\tau_2$  is the delay time. ELM modeling can be completed according to equations (12) and (13). In this paper, a one-step prediction is adopted, that is, to predict the value of the next time point from a period of the data sequence.

$m$  and  $\tau_2$  are two critical parameters in phase space reconstruction. With reasonable values, they can make the reconstructed phase space close to the original system to a certain extent. In traditional studies, C-C method [34] and

G-P method [35] are often used to determine their values, but the advance calculation is required for specific time series. In this paper, the VMD method is used to decompose the signal, and the phase space reconstruction parameters of each mode cannot be calculated in advance. Therefore, to enable  $m$  and  $\tau_2$  to have an adaptive capacity for each model after VMD decomposition, the optimization method can be adopted in the training process.

### III. ARTIFICIAL BEE COLONY ALGORITHM

In this study, the cluster intelligent optimization algorithm is used to optimize several critical parameters in the network traffic prediction model. These parameters are the number of modes  $K$ , iterative factor  $\tau_1$ , the number of hidden layer nodes  $L$  of each ELM model, and the phase-space reconstruction parameters  $m$  and  $\tau_2$  after decomposition. Although some intelligent optimization algorithms can be used, such as: PSO, IABC, IHS, IFS, IGSA, and so on, there are some adjustable parameters involved. That are, learning factor in PSO, elasticity coefficient of velocity update and population update, parameter iteration velocity coefficient, and forth; In IABC, control parameters of the food source, random step size and location update coefficient moving towards the optimal solution are given up; In IHS, the harmony number, the retention probability of the harmony memory bank, the likelihood of memory disturbance, the minimum bandwidth, the maximum bandwidth, etc.; In IFS, search small steps, search radius, etc; and coefficient of gravity, coefficient of iteration, percentage of attraction in IGSA, etc. These adjustable parameters could reduce the stability and accuracy of the prediction model. Therefore, in order to solve this problem, the study advocates a Scalable Artificial Bee Colony (SABC) algorithm to minimize the problem of uncertainty and poor stability of the prediction model.

#### A. CLASSICAL ARTIFICIAL BEE COLONY ALGORITHM

The ABC algorithm realizes the search of the food source (the optimal solution) through the cooperation of the three kinds of bee colonies. The search process of the algorithm is as follows:

(1) Randomly generate initial solution  $x_i = \{x_{i1}, x_{i2}, \dots, x_{id}\}$ ,  $i = 1, 2, \dots, M$ ,  $M$  is the number of honey sources,  $d$  is the dimension of solution. The initial solution is generated as follows:

$$x_i = x_i^{down} + (x_i^{up} - x_i^{down}) \cdot \text{rand}() \quad (14)$$

where,  $[x_i^{down}, x_i^{up}]$  is the value range of  $x_i$ , and  $\text{rand}()$  is the random number within  $[0,1]$ .

(2) In the initial stage of the search, each lead bee generates a new solution as follows:

$$\text{new}x_{i,j} = x_{i,j} + (x_{i,j} - x_{k,j}) \cdot \text{rands}() \quad (15)$$

Among them, the  $k \in \{1, 2, \dots, M\}$ ,  $j \in \{1, 2, \dots, d\}$ , and  $k \neq i$ ,  $\text{rands}()$  is the random number in  $[-1, 1]$ .

(3) After all the lead bees have completed the search process, they will share the solution information with the follow bees in the recruitment area.

The follow bee calculates the selection probability of each solution as follows:

$$P_i = \frac{fit_i}{\sum_{k=1}^M fit_k} \quad (16)$$

where the  $fit$  is the fitness function value. Then, a random number is generated in the interval  $[-1,1]$ . If  $P_i$  is greater than the random number, the follow bee will generate a new solution from Equation (15). If the  $fit_i$  of the new solution is better than the previous one, the follow bee will remember the new solution and forget the old solution; otherwise, the old solution will be retained.

(4) After all the follow bees have completed the search process, if a solution has not been updated after several cycles, the honey source will be abandoned (considered to be trapped in the local optima). When a honey source is abandoned, its corresponding lead bee turns into a scout bee, generating a new honey source from Equation (14).

Artificial Bee Colony algorithm finds the optimal solution through the above cyclic search. The main steps of Artificial Bee Colony algorithm are as follows:

Step 1: The initialization phase. Set parameters such as colony number, maximum iteration times, solution search range, etc., and then generate the initial solution within the solution search range according to Equation (14). The number of which is half of the population.

Step 2: Calculate the fitness function value of each initial solution, then sort the advantages and disadvantages according to the results. Take the first 50% as the lead bee and the second 50% as the follow bee.

Step 3: Set the loop conditions and begin the search process.

Step 4: For the leading bees, generate a new solution according to Equation (15), and calculate its fitness value.

Step 5: Evaluate the fitness function value of the new solution. If it is better than the old solution, the bees will be led to remember the new solution and forget the old solution. Otherwise, the old solution will still be kept.

Step 6: Calculate the selection probability of each solution according to Equation (16). The following bee selects the honey source according to probability  $P_i$ , and produces a new solution according to Equation (15), and calculates its fitness value. If the fitness function value of the new solution is better than the old solution, the bees will be guided to remember the new solution and forget the old solution; otherwise, the old solution will be retained.

Step 7: Judge whether there are any solutions to be given up. If there are, the scout bees will randomly generate new solutions according to Equation (14) and replace them.

Step 8: Record the optimal solution for each iteration.

Step 9: Determine whether the loop termination condition is met; if so, end the loop and output the optimal solution; otherwise, return to Step 4 to continue the search.

**B. NEW SCALABLE ARTIFICIAL BEE COLONY ALGORITHM**

The search of Artificial Bee Colony algorithm involves three main processes: initial population evaluation, lead bee population renewal, and reconnaissance honey source renewal. Decomposition and modeling calculation will be repeated many times when VMD decomposition and ELM prediction is applied, resulting in the complex calculation and long optimization process time. Also, in the process of lead bee populations' update, there is a need to update each lead through several neighboring lead bees. When applied in predicting VMD decomposition and ELM, every solution of the decomposition and modeling calculation has to be updated. If there is too much focus on local waste and its calculations overlap computing resources, it will take a long time to calculate the optimization. In this regard, this paper improves the search mechanism of artificial bee colony algorithm by reducing the number of iterations of VMD decomposition and ELM prediction required. New Scalable Artificial Bee Colony Algorithm (SABC) was proposed to improve the convergence speed and prediction accuracy.

**C. UPDATE MODEL WITH NEW SOLUTIONS**

According to the population dimension, each dimension is traversed successively, and an individual lead bee in the lead bee population (the first 50% of the whole population) is randomly selected for the updating every time.

We update the selection formula of the solution as follows:

$$\begin{cases} newx_{i,j} = x_{i,j} \\ j = \text{ceil}(\text{rand}() \cdot \frac{\text{popsize}}{2}) \end{cases} \quad (17)$$

where  $i$  represents the  $i^{\text{th}}$  dimension variable,  $j$  represents the  $j^{\text{th}}$  selected individual, and  $\text{popsize}$  means the population size.

The updated and revised formulae for the new solution are as follows:

$$newx_{i,j} = x_{i,j} + \delta \cdot (x_{i,j} - \text{neig}x_{i,k}) \cdot \text{rand} \text{ s}() \quad (18)$$

$$k = \begin{cases} \text{fix}(\text{rand}() \cdot \frac{\text{popsize}}{2}), & k \neq j \\ \text{fix}(\text{rand}() \cdot \frac{\text{popsize}}{2}) + 1, & k = j \end{cases} \quad (19)$$

where  $\text{neig}x_{i,k}$  is the  $K^{\text{th}}$  nearest neighbour solution of  $x_{i,j}$ ,  $\delta$  is the search step size.

**D. FINE-TUNE DISTURBANCES THAT GENERATE NEW SOLUTIONS**

After the new solution is generated, falling into local optimum is a concern which should be addressed. Hence, set the probability of fine-tuning disturbance  $PA$ , and randomly generate a number in the interval  $[0,1]$ . When it is less than  $PA$ , a positive fine-tuning disturbance will be generated for

the new solution, with the following relationship:

$$newx_{i,j} = newx_{i,j} + \sigma \cdot \text{rand}(), \quad \text{rand} < PA \quad (20)$$

where  $\sigma$  is the disturbance step size.

When the random number generated in the interval of  $[0, 1]$  is greater than or equal to  $PA$ , a negative fine-tuning disturbance is generated to arrive at the new solution which has the following relationship:

$$newx_{i,j} = newx_{i,j} - \sigma \cdot \text{rand}(), \quad \text{rand} \geq PA \quad (21)$$

The choice of the  $PA$  value of the fine-tuning perturbation probability can be set according to the specific application problem. A larger  $PA$  value can accelerate the speed of searching solution space and make the optimization process converge as soon as possible. Smaller  $PA$  values focus on local searches, particularly in more detailed searches around the global optimal solution.

*Remark:* According to equations (17)-(21), the update mechanism of our algorithm can be implemented by the following procedures:

1.  $j \leftarrow I(1, \lceil \frac{\text{popsize}}{2} \rceil)$ , where  $I(1, \lceil \frac{\text{popsize}}{2} \rceil)$  is an integer randomly chosen from the interval  $[1, \lceil \frac{\text{popsize}}{2} \rceil]$  (with uniform probability);
2.  $k \leftarrow I(1, \lceil \frac{\text{popsize}}{2} \rceil)$ ;
3. If  $k = j$ , then  $k \leftarrow k+1$ ;
4.  $newx_{i,j} \leftarrow x_{i,j} + \delta \cdot (x_{i,j} - \text{neig}x_{i,k}) \cdot U(-1, 1)$ , where  $U(-1,1)$  is a uniform random variable on  $(-1,1)$ ;
5. If  $U[0, 1] < PA$ ,  $newx_{i,j} \leftarrow newx_{i,j} + \sigma \cdot U[0, 1]$ ; else  $newx_{i,j} \leftarrow newx_{i,j} - \sigma \cdot U[0, 1]$ .

**IV. PROPOSED PREDICTION MECHANISM**

We begin by utilizing the VMD method to decompose the data sequence to obtain  $K$  modes corresponding to  $K$  data subset sequences. Then, the data subsets in each mode are reconstructed in phase space to get the data sequences in high dimensional space. Finally, the reconstructed data of  $K$  group phase space is substituted into the  $K$  ELM models as input to obtain the predicted output. During the whole prediction process, the proposed Scalable Artificial Bee Colony (SABC) algorithm is used to optimize the selection of different model parameters. It includes the control parameters  $K$  and  $\tau_1$  for VMD decomposition, control parameters  $m$  and  $\tau_2$  for phase space reconstruction and the number of hidden layer nodes  $L$  in the ELM model.

The main calculation steps of the proposed prediction method are as follows:

Step 1: Initialization. Set the population size, the maximum number of iterations, and the range of control parameters to be optimized.

Step 2: Make the initial population and each individual in the population to be the 5-dimensional variable, which is represented as  $[L, m, \tau_2, K, \tau_1]$ . Each individual of the initial population is randomly assigned within its range.

Step 3: VMD decomposition is carried out for the data sequence; phase space reconstruction is carried out for each

data subset sequence after decomposition. Then  $K$  ELM models are established, and the training samples are used for training.

Step 4: Evaluate each individual in the initial population by the fitness function defined in Equation (22) to determine the optimal individual and global optimal solution in the initial population.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N_{data}} (y_i - y'_i)^2}{N_{data}}} \quad (22)$$

where  $N_{data}$  is the number of samples,  $y_i$  and  $y'_i$  are the actual value and predicted value of the  $i^{th}$  sample, respectively.

Step 5: Following the main stages of the Scalable Artificial Bee Colony (SABC) algorithm and Equations (17)-(21), update the population using the SABC and obtain the revised new solution.

Step 6: Calculate the fitness function value of the updated new solution according to Equation (22) and compare it with the historical global optimal solution. If it is better than the historical global optimal solution, update the historical global optimal solution; otherwise, keep the historical global optimal solution.

Step 7: If the search termination condition is met, terminate the iteration and output the global optimal solution; otherwise, return to Step 5 to continue the iteration.

The flow chart of the prediction method is shown in Figure 1.

**V. THE SIMULATION VERIFICATION**

The validity of the proposed mechanism is verified using three datasets. They are Mackey-glass chaotic time series, Lorenz chaotic time series and WIDE backbone actual network traffic data of MAWI Working Group. The experiments are conducted in a simulated testbed. Six swarm intelligence optimization algorithms (IGSA [13], IABC [10], IFS [27], IHS [26], LVCMFOA [28] and PSO [25]) are applied to construct the optimal modeling for VMD and ELM, respectively. The results are used to compare the findings of our proposed algorithm, Scalable Artificial Bee Colony (SABC) algorithm. The adjustable parameters setting in different swarm intelligence algorithms are shown in Table 1.

The different values in the parameters may have some impact on the performance of the algorithm. To avoid the bias, the experiments are conducted repeatedly by randomly selecting the parameter value within the range. The computer configuration is as follows: “Windows 7” system, “Matlab 2015a”, Intel core i7 4.00GHz CPU, 16GB RAM.

**A. MACKEY-GLASS CHAOTIC TIME SERIES**

The following differential equations describe Mackey-Glass chaotic sequences:

$$\frac{dx(t)}{dt} = \frac{0.2x(t - \eta)}{\eta} - 0.1x(t) \quad (23)$$

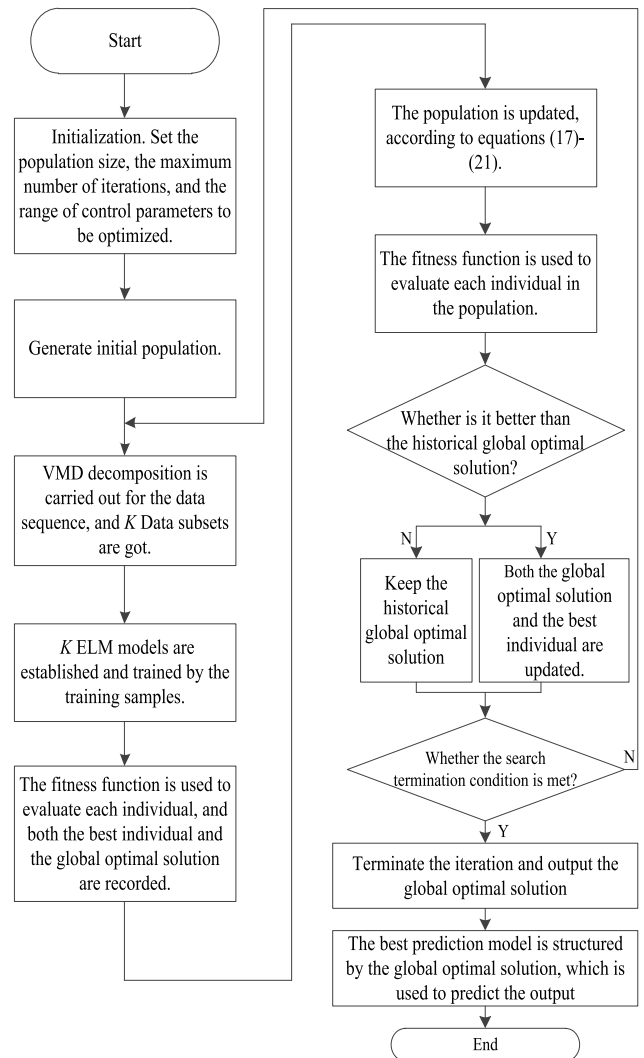


FIGURE 1. Flow chart of the prediction method.

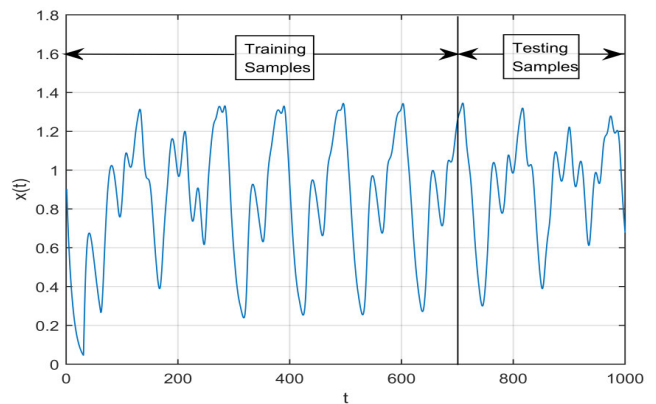


FIGURE 2. Mackey-Glass chaotic time series.

When  $\eta > 17$ , the Mackey-Glass sequence has chaotic behavior. In this paper, the  $\eta = 30$  and the initial value  $x(t)|_{t=0} = 0.9$  are selected. One thousand data points are generated, as shown in Figure 2.

TABLE 1. Parameters setting.

Algorithm	Parameters
IGSA <sup>[13]</sup>	Two adjustable parameters are involved in the gravitational constant. The gravitational coefficient is randomly assigned within [20,200], and the iteration coefficient is randomly assigned within [1,50]. The acceleration constant involves an adjustable parameter, and the attraction percentage is randomly assigned within [1,10].
IABC <sup>[10]</sup>	The control parameters of the abandoned food source are randomly assigned within [10,100]. The random step length is randomly assigned within [-1,1]. The update coefficient of the position moving toward the optimal solution is randomly assigned within [0,1]. The scaling factor and exponential factor of the adaptive iterative calculation are randomly assigned within [10,100] and [2,10], respectively.
IFS <sup>[27]</sup>	The number of small search steps is randomly assigned within [1,50], while the search radius is randomly assigned within [0.1,1].
IHS <sup>[26]</sup>	The harmony number of the harmony memory bank is randomly assigned within [20,50]. The retention probability of the harmony memory bank is randomly assigned within [0.5,0.999]. The memory disturbance probability was randomly assigned within [0.01,0.5]. The minimum bandwidth is randomly assigned within [0.001,0.005], whereas the maximum bandwidth is randomly assigned within [0.5,1].
LVCMFOA <sup>[28]</sup>	The three adjustable coefficients in the cloud model are randomly assigned within [0.001,0.1], [0.2,4] and [0.1,10], respectively.
PSO <sup>[25]</sup>	The two learning factors are randomly assigned within [0,2]. The elasticity coefficient of velocity update and population update is randomly assigned within [0.8,1.2], and the parameter iteration velocity coefficient is randomly assigned within [0.1,1.0].
SABC (proposed)	The control parameter of abandoning food source is randomly assigned within [10,100], optimized parameter is $L \in [2,500]$ , $m \in [1,n/30]$ , $\tau_2 \in [1,n/50]$ , $K \in [2,15]$ , $\tau_1 \in [0.1,3]$ . $n$ is the number of samples.

As shown in Figure 2, 1000 data points are generated; the first 700 groups are taken as training samples, while the last 300 groups are taken as test samples. Seven swarm intelligence optimization algorithms (IGSA [13], IABC [10], IFS [27], IHS [26], LVCMFOA [28], PSO [25] and SABC) are used to conduct optimization modeling for VMD and ELM, respectively. Each model is run repeatedly for several times. The optimal value and worst value predicted by different data sets under the three indicators of Root Mean Squared Error (RMSE), Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) as well as the average training time and test time of the model are accordingly recorded. The average fitness curves of different models at the training stage are displayed in Figure 3.

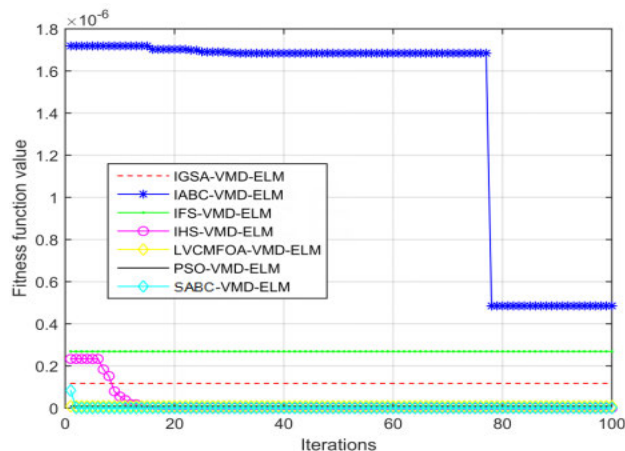


FIGURE 3. Average fitness curve of different models in the training stage (Mackey-Glass).

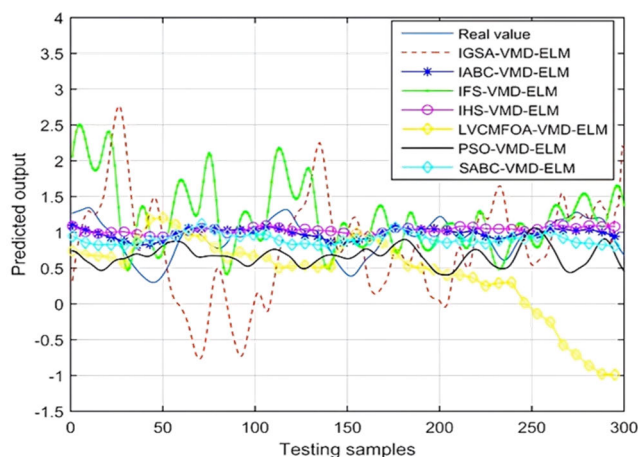


FIGURE 4. Average predicted output curves of different models in the test phase (Mackey-Glass).

As seen in Figure 3, the convergence performance of IABC-VMD-ELM is the worst in Mackey-Glass chaotic time series. Meanwhile, the IFS-VMD-ELM and IGSA-VMD-ELM fall into the local optima state. The convergence speed of IHS-VMD-ELM is relatively slow. IFS-VMD-ELM, LVCMFOA-VMD-ELM and SABC-VMD-ELM all have a good convergence performance, but the SABC-VMD-ELM has a good effect on the convergence accuracy.

The average output curves of different models in the test stage are shown in Figure 4.

According to Figure 4, IFS-VMD-ELM, IGSA-VMD-ELM and LVCMFOA-VMD-ELM all have poor generalization ability. However, SABC-VMD-ELM can track the actual output better and has better prediction ability comparatively. The predictive performance metrics (RMSE, MAE, and MAPE) for several models run repeatedly are shown in Table 2.

The findings in Table 2 depict that IGSA-VMD-ELM, IFS-VMD-ELM, LVCMFOA-VMD-ELM and PSO-VMD-ELM



TABLE 2. Comparison of predictive performance metrics of the models (Mackey-Glass).

Models	RMSE		MAE		MAPE	
	Optimal value	Worst value	Optimal value	Worst value	Optimal value	Worst value
IGSA-VMD-ELM	0.2732	4.0931	0.2179	3.3409	33.6650	376.5902
IABC-VMD-ELM	0.2488	0.2689	0.1920	0.2042	28.5558	31.5123
IFS-VMD-ELM	0.3267	3.4922	0.2789	2.7709	39.0579	311.8475
IHS-VMD-ELM	0.2379	0.3308	0.1810	0.2550	27.9956	40.7154
LVCMFOA-VMD-ELM	0.3027	2.9169	0.2329	1.9930	35.6722	207.8941
PSO-VMD-ELM	0.2584	0.9506	0.2027	0.7440	29.4848	78.1168
SABC-VMD-ELM	0.2191	0.2659	0.1737	0.2217	25.2589	27.1613

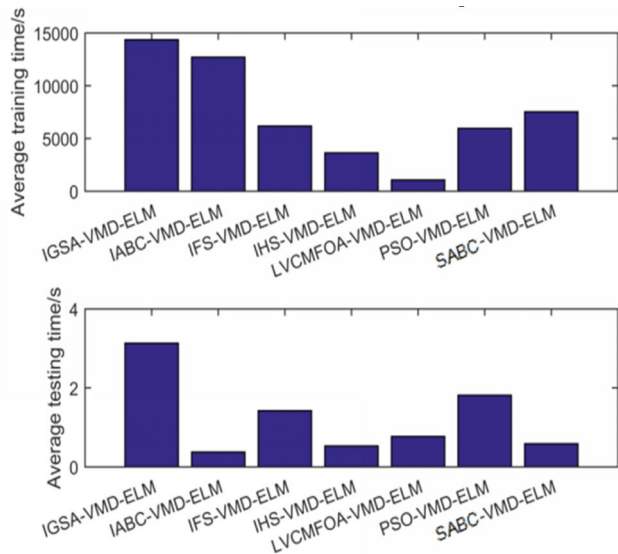


FIGURE 5. Calculation time of different models.

have a wide range of variations. It indicates that different values of adjustable model parameters in the intelligent optimization algorithms will affect the prediction results. The optimal predictive performance metrics of IHS-VMD-ELM is better than that of IABC-VMD-ELM. Yet its overall predictive performance and stability are weaker than that of IABC-VMD-ELM. Compared with the other six models, SABC-VMD-ELM has better prediction performance, overall prediction performance and stability.

The comparison of the average training time and average test time of the seven models is shown in Figure 5.

Figure 5 shows that in comparison with the IABC-VMD-ELM model, SABC-VMD-ELM can significantly reduce the training time. Although SABC-VMD-ELM does not have the least training time among all models, it is acceptable when overall predictive performance and stability are taken into account.

B. LORENZ CHAOTIC TIME SERIES

The differential equation of the Lorenz equation is described as follows:

$$\begin{cases} \frac{dx}{dt} = A \cdot (y(t) - x(t)) \\ \frac{dy}{dt} = B \cdot x(t) - y(t) - x(t) \cdot z(t) \\ \frac{dz}{dt} = x(t) \cdot y(t) - C \cdot z(t) \end{cases} \quad (24)$$

where  $A = 10$ ,  $B = 2$  and  $C = 8/3$ , and the initial values of  $x(0)$ ,  $y(0)$  and  $z(0)$  are 12, 2 and 9, respectively.

One thousand sets of  $x(t)$  component data are generated, as shown in Figure 6. The first 700 groups are taken as training samples, while the last 300 groups are taken as test samples.

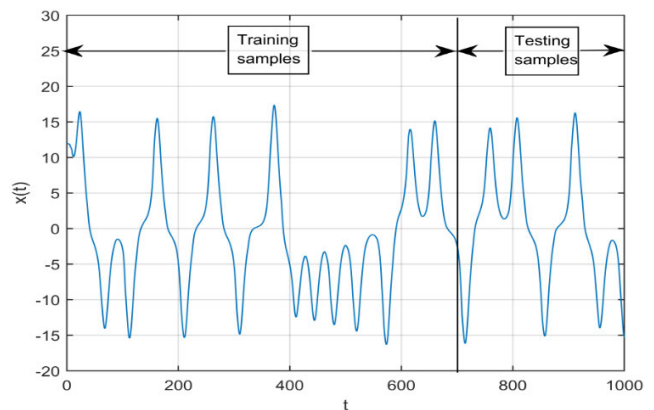


FIGURE 6. Lorenz chaotic time series.

The seven models are run repeatedly for several times. The optimal value and the worst value predicted by different data sets under RMSE, MAE and MAPE as well as the model's average training time and test time, are recorded, respectively.

The average fitness curves of different models at the training stage are shown in Figure 7.

TABLE 3. Comparison of predictive performance metrics of the models (Lorenz).

Models	RMSE		MAE		MAPE	
	Optimal value	Worst value	Optimal value	Worst value	Optimal value	Worst value
IGSA-VMD-ELM	8.9190	44.8147	7.3178	36.1740	521.0107	2925.48835
IABC-VMD-ELM	2.2789	14.7667	1.0092	11.8931	41.7971	670.4422
IFS-VMD-ELM	8.4385	14.8419	6.8399	12.0576	367.7279	941.3042
IHS-VMD-ELM	8.1711	9.3338	6.4582	7.3426	198.6465	269.6382
LVCMFOA-VMD-ELM	8.0015	88.5849	6.4334	23.8704	211.7773	2082.8568
PSO-VMD-ELM	8.2883	13.1922	6.5549	10.3343	210.7426	913.6893
SABC-VMD-ELM	4.1681	4.5772	2.3803	3.0253	88.8466	115.5435

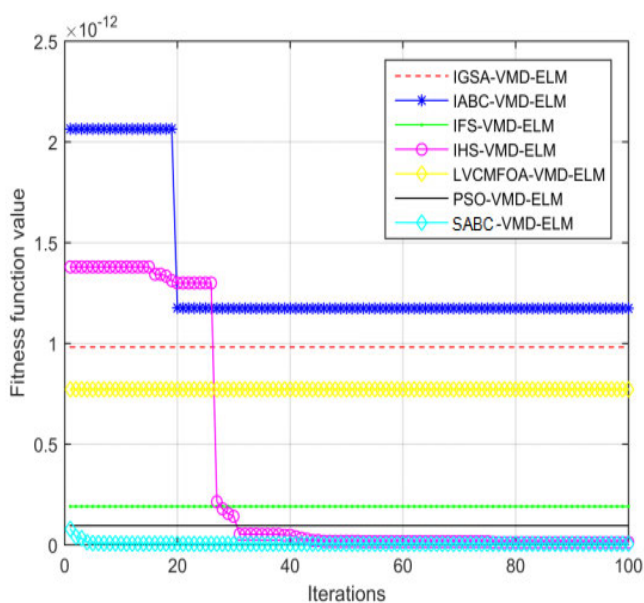


FIGURE 7. Mean fitness curve of different models in the training stage (Lorenz).

As shown in Figure 7, the convergence performance of IABC-VMD-ELM is the worst for Lorenz chaotic time series. Whereas IGSA-VMD-ELM, LVCMFOA-VMD-ELM, IFS-VMD-ELM and PSO-VMD-ELM fall into the local optima in the initial stage. IHS-VMD-ELM has good convergence accuracy, but its convergence speed is slow. On the contrary, the SABC-VMD-ELM has a good effect in both convergence speed and convergence accuracy.

The average output curves of different models in the test stage are shown in Figure 8.

It can be seen in Figure 8 that the random value of adjustable parameters will affect the stability of the predicted output of the model when repeated many times. Compared with the other models, SABC-VMD-ELM is the least affected and has certain advantages in prediction

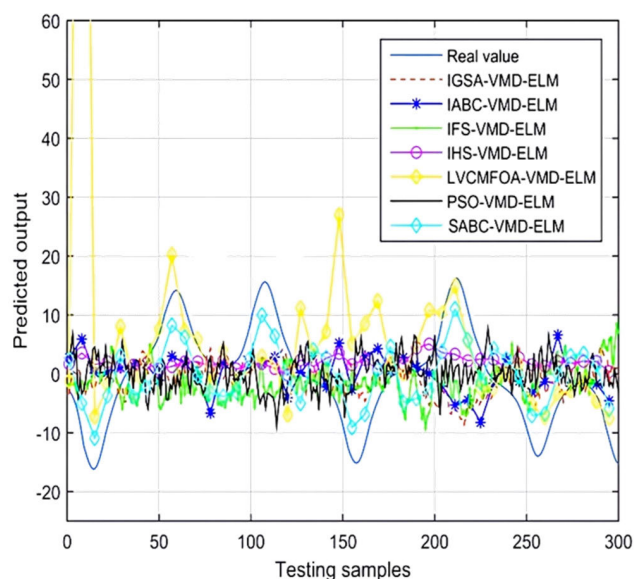


FIGURE 8. Average predicted output curve of different models in the test phase (Lorenz).

accuracy and stability. The predictive performance metrics (RMSE, MAE, and MAPE) for several models run repeatedly are shown in Table 3.

Based on Table 3, it can be concluded that after many repeated operations, the predictive performance metrics of IGSA-VMD-ELM and LVCMFOA-VMD-ELM have a broad range of variations, indicating that the predictive performance of the two models is greatly affected by the random value of adjustable parameters. IHS-VMD-ELM, IFS-VMD-ELM and PSO-VMD-ELM have good predictive stability, but the prediction accuracy is low. Although IABC-VMD-ELM proves to have the best prediction accuracy, its prediction stability is still poor. Compared with the other models, SABC-VMD-ELM demonstrates better prediction performance and stability on the whole.

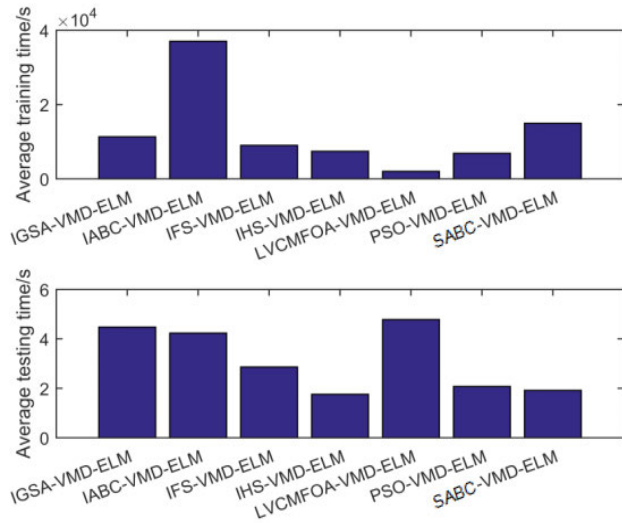


FIGURE 9. Calculation time of different models (Lorenz).

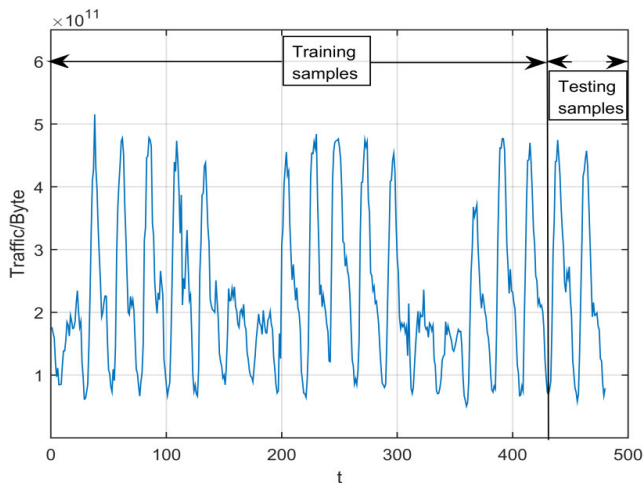


FIGURE 10. Network traffic data set of "Hour."

The comparison of the average training time and average test time of the seven models is shown in Figure 9.

As shown in Figure 9, SABC-VMD-ELM can significantly reduce the training time when compared with the IABC-VMD-ELM model. Moreover, SABC-VMD-ELM has an advantage over the latter when considering the overall prediction performance, stability and training time.

C. WIDE BACKBONE NETWORK TRAFFIC DATA

WIDE backbone network traffic data is MAWI Working Group which investigates the Internet service records of network traffic datasets (<http://mawi.wide.ad.jp/mawi/>) in routine work. From July 1 to July 21 2018, 480 data points in the data set were collected and processed and used for simulation experiment. The sampling period of the data was 1 Hour and recorded as "Hour" data set [28]. The 480 data points in the data are shown in Figure 10, in which the first 430 data points

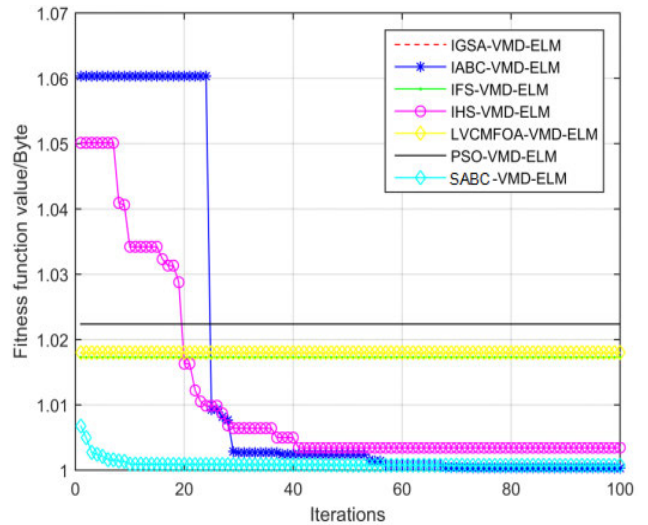


FIGURE 11. Mean fitness curves of different models at the training stage ("Hour").

are used as training samples, and the last 50 groups of data are used as test samples.

The 7 models are run repeatedly for several times. The optimal value and the worst value predicted by different data sets under RMSE, MAE and MAPE as well as the model's average training time and test time are recorded, respectively.

The average fitness curves of different models at the training stage are shown in Figure 11.

As shown in Figure 11, for the "Hour" data set, IFS-VMD-ELM, LVCMFOA-VMD-ELM and PSO-VMD-ELM fall into the local optima state at the initial stage as well as demonstrate poor convergence effect. Although IHS-VMD-ELM gradually converges, it does not achieve the best convergence accuracy. The convergence precision of IABC-VMD-ELM is better than that of IHS-VMD-ELM, but the convergence speed is slow. Compared with the other models, SABC-VMD-ELM has a better effect on convergence speed and convergence precision.

The average output curves of different models in the test stage are shown in Figure 12.

Figure 12 shows that IGSA-VMD-ELM, IFS-VMD-ELM, IHS-VMD-ELM, LVCMFOA-VMD-ELM and PSO-VMD-ELM have low generalization ability when the adjustable parameters are randomly taken in this actual network traffic data. IABC-VMD-ELM and SABC-VMD-ELM can track the actual output better, but the latter has an advantage in prediction accuracy.

The predictive performance metrics (RMSE, MAE, and MAPE) for several models run repeatedly are shown in Table 4.

Based on the results in table 4 and on the actual network traffic data, it can be surmised that the random value of adjustable parameters will have a great impact on the generalization ability of IGSA-VMD-ELM,

TABLE 4. Comparison of predictive performance metrics of the models ("Hour").

Models	RMSE		MAE		MAPE	
	Optimal value	Worst value	Optimal value	Worst value	Optimal value	Worst value
IGSA-VMD-ELM	5.3e+29	6.5e+49	8.3e+18	1.1e+35	3.8e+50	1.3e+93
IABC-VMD-ELM	1.0953	1.1461	1.0584	1.0996	1.1635	1.2891
IFS-VMD-ELM	9.2e+51	1.8e+80	5.3e+31	8.8e+61	8.8e+83	3.7e+165
IHS-VMD-ELM	1.2e+127	5.1e+205	8.4e+91	3.1e+130	3.1e+245	1.5e+256
LVCMFOA-VMD-ELM	2.0e+15	3.4e+55	4.9e+11	1.5e+41	1.4e+31	2.1e+109
PSO-VMD-ELM	44.6999	8.0e+55	25.4607	2.3e+39	5.4e+03	6.4e+103
SABC-VMD-ELM	1.0272	1.0407	1.0202	1.0272	1.0546	1.0743

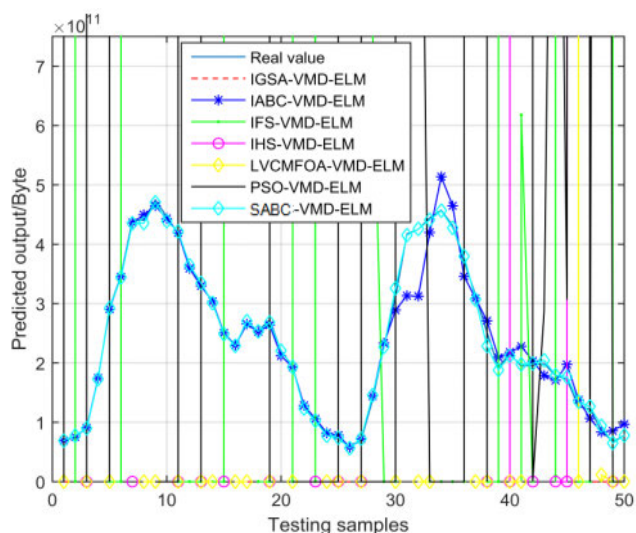


FIGURE 12. Average predicted output curves of different models in the test phase ("hour").

IFS-VMD-ELM, IHS-VMD-ELM and LVCMFOA-VMD-ELM, resulting in a large deviation in the predicted output. Although the optimal value of PSO-VMD-ELM prediction is acceptable, the overall prediction performance is unstable. On the other hand, IABC-VMD-ELM and SABC-VMD-ELM can track the actual value well and are less affected by the random value of adjustable parameters. At the same time, SABC-VMD-ELM has better prediction accuracy and stability.

The comparison of the average training time and average test time of the seven models is shown in Figure 13.

As can be seen from Figure 13, SABC-VMD-ELM significantly reduces the training time when compared with IABC-VMD-ELM. However, in comparison with the other five models, it has a longer training time; yet it has a clear

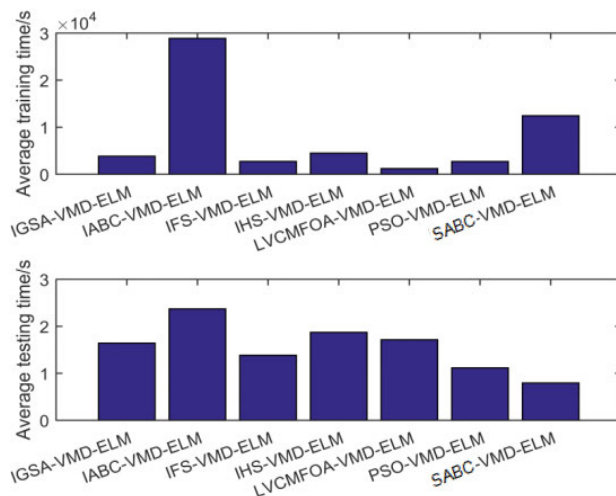


FIGURE 13. Calculation time of different models "Hour".

advantage in testing time. In sum, the prediction performance of SABC-VMD-ELM is the best when considering the prediction accuracy, stability and calculation time.

### VI. CONCLUSION

This paper proposed a prediction mechanism based on optimal variational mode decomposition and integrated extreme learning machine to predict the network traffic. To reduce the influence of different scale characteristics of network traffic on the prediction accuracy, firstly, the variational mode decomposition method is used to decompose the network traffic data. Then, for the sub-data set corresponding to each mode after decomposition, the predictive sub-models are established respectively using the integrated extreme learning machine. Finally, the outputs of all the sub-models are integrated to get the outcome [36]. In the process of

designing a prediction model, several optimized parameters are involved in VMD, PSR and ELM. A Scalable Artificial Bee Colony (SABC) algorithm is thus proposed to train their optimal values. Compared with other evolutionary algorithms, the proposed SABC algorithm has less adjustable parameters. Another advantage is its optimization accuracy and stability. It improves the performance of network traffic prediction model by optimizing multiple parameters in the prediction model. Simulation experiments are carried out with Mackey-Glass chaotic time series, Lorenz chaotic time series and WIDE backbone-network traffic data. The results proved that the proposed model demonstrates better accuracy in predicting network traffic.

## ACKNOWLEDGMENT

The authors would like to thank the Faculty of Computing and Informatics, Universiti Malaysia Sabah. The funding of the project supported by Hainan Provincial Natural Science Foundation of China, NO. 620RC669, and regional science foundation project approved by 2016 Natural Science Foundation of China, NO. 61663007.

## REFERENCES

- Z. D. Tian, S. J. Li, Y. H. Wang, and X. D. Wang, "Network traffic multi-step prediction based on chaos theory and improved echo state network," *J. Commun.*, vol. 37, no. 3, pp. 55–70, 2016.
- W. E. Leland et al., "On the self-similar nature of Ethernet traffic (extended version)," *IEEE/ACM Trans. Netw.*, vol. 2, no. 1, pp. 1–15, 1994.
- Q. Li and R.-C. Lin, "A new approach for chaotic time series prediction using recurrent neural network," *Math. Problems Eng.*, vol. 2016, pp. 1–9, Dec. 2016.
- D. Wei, "Network traffic prediction based on RBF neural network optimized by improved gravitation search algorithm," *Neural Comput. Appl.*, vol. 28, no. 8, pp. 2303–2312, Aug. 2017.
- W. Guo, Y. Qiao, and H. Hou, "BP neural network optimized with PSO algorithm and its application in forecasting," in *Proc. IEEE Int. Conf. Inf. Acquisition*, Aug. 2006, pp. 617–621.
- K. Li, W. Xu, Y. Han, F. Ge, and Y. Wang, "A hybrid modeling method for interval time prediction of the intermittent pumping well based on IBSO-KELM," *Measurement*, vol. 151, pp. 1–15, Feb. 2020.
- Y. Han and K. Li, "Improved FOA-ESN method using opposition-based learning mechanism for the network traffic prediction with multiple steps," in *Proc. Chin. Autom. Congr. (CAC)*, Oct. 2017, pp. 7183–7187.
- G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing*, vol. 70, nos. 1–3, pp. 489–501, Dec. 2006.
- Q. Liu, W. Cai, J. Shen, Z. Fu, and N. Linge, "A novel neural computing model for fast predicting network traffic," *J. Comput. Theor. Nanosci.*, vol. 12, no. 12, pp. 6056–6062, Dec. 2015.
- Z. Tian et al., "Network traffic prediction method based on improved ABC algorithm optimized EM-ELM," *J. China Univ. Posts Telecommun.*, vol. 25, no. 3, pp. 37–48, 2018.
- T. Zhongda, M. Chengcheng, W. Gang, and R. Yi, "Approach for time series prediction based on empirical mode decomposition and extreme learning machine," in *Proc. Chin. Control Decis. Conf. (CCDC)*, Jun. 2018, pp. 3119–3123.
- K. Li, Y. Han, and T. Wang, "A novel prediction method for down-hole working conditions of the beam pumping unit based on 8-directions chain codes and online sequential extreme learning machine," *J. Petroleum Sci. Eng.*, vol. 160, pp. 285–301, Jan. 2018.
- C. Li, Z. Xiao, X. Xia, W. Zou, and C. Zhang, "A hybrid model based on synchronous optimisation for multi-step short-term wind speed forecasting," *Appl. Energy*, vol. 215, pp. 131–144, Apr. 2018.
- Z. Tian, "Approach for short-term traffic flow prediction based on empirical mode decomposition and combination model fusion," *IEEE Trans. Intell. Transp. Syst.*, early access, May 8, 2020, doi: 10.1109/TITS.2020.2987909.
- R. Guo, Y. Wang, H. Zhang, and G. Zhang, "Remaining useful life prediction for rolling bearings using EMD-RISI-LSTM," *IEEE Trans. Instrum. Meas.*, vol. 70, Jan. 2021, Art. no. 3509812.
- O. Abedinia, M. Lotfi, M. Bagheri, B. Sobhani, M. Shafie-khah, and J. P. S. Catalao, "Improved EMD-based complex prediction model for wind power forecasting," *IEEE Trans. Sustain. Energy*, vol. 11, no. 4, pp. 2790–2802, Oct. 2020.
- Z. Tian, "Short-term wind speed prediction based on LMD and improved FA optimized combined kernel function LSSVM," *Eng. Appl. Artif. Intell.*, vol. 91, May 2020, Art. no. 103573.
- G. Li, W. Chang, and H. Yang, "Monthly mean meteorological temperature prediction based on VMD-DSE and Volterra adaptive model," *Adv. Meteorol.*, vol. 2020, pp. 1–17, Mar. 2020.
- H. Hu, J. Zhang, and T. Li, "A comparative study of VMD-based hybrid forecasting model for nonstationary daily streamflow time series," *Complexity*, vol. 2020, pp. 1–21, Jul. 2020.
- H. Niu, K. Xu, and W. Wang, "A hybrid stock price index forecasting model based on variational mode decomposition and LSTM network," *Appl. Intell.*, vol. 50, pp. 4296–4309, Jul. 2020.
- K. Dragomiretskiy and D. Zosso, "Variational mode decomposition," *IEEE Trans. Signal Process.*, vol. 62, no. 3, pp. 531–544, Feb. 2014.
- P. Du, J. Wang, W. Yang, and T. Niu, "Point and interval forecasting for metal prices based on variational mode decomposition and an optimized outlier-robust extreme learning machine," *Resour. Policy*, vol. 69, Dec. 2020, Art. no. 101881.
- I. Majumder, P. K. Dash, and R. Bisoi, "Variational mode decomposition based low rank robust kernel extreme learning machine for solar irradiation forecasting," *Energy Convers. Manage.*, vol. 171, pp. 787–806, Sep. 2018.
- S. N. Deepa, B. Gobu, S. Jaikumar, N. Arulmozhi, and P. Kanimozhi, "Adaptive regularized ELM and improved VMD method for multi-step ahead electricity price forecasting," in *Proc. 17th IEEE Int. Conf. Mach. Learn. Appl. (ICMLA)*, Dec. 2018, pp. 1255–1260.
- K. Li, X. Gao, Z. Tian, and Z. Qiu, "Using the curve moment and the PSO-SVM method to diagnose downhole conditions of a sucker rod pumping unit," *Petroleum Sci.*, vol. 10, no. 1, pp. 73–80, Mar. 2013.
- Z. Tian, S. Li, Y. Wang, and X. Wang, "A network traffic hybrid prediction model optimized by improved harmony search algorithm," *Neural Netw. World*, vol. 25, no. 6, pp. 669–686, 2015.
- Z. Tian and S. Li, "A network traffic prediction method based on IFS algorithm optimised LSSVM," *J. Eng. Syst. Model. Simul.*, vol. 19, no. 4, pp. 200–213, 2017.
- Y. Han, Y. Jing, K. Li, and G. M. Dimirovski, "Network traffic prediction using variational mode decomposition and multi-reservoirs echo state network," *IEEE Access*, vol. 7, pp. 138364–138377, 2019.
- D. Karaboga, "An idea based on honey bee swarm for numerical optimization," Dept. Comput. Eng., Fac. Eng., Erciyes Univ., Kayseri, Turkey, Tech. Rep. tr06, 2005, vol. 200, pp. 1–10.
- R. Bisoi, P. K. Dash, and A. K. Parida, "Hybrid variational mode decomposition and evolutionary robust kernel extreme learning machine for stock price and movement prediction on daily basis," *Appl. Soft Comput.*, vol. 74, pp. 652–678, Jan. 2019.
- H. Liu, X. Mi, and Y. Li, "Smart multi-step deep learning model for wind speed forecasting based on variational mode decomposition, singular spectrum analysis, LSTM network and ELM," *Energy Convers. Manage.*, vol. 159, pp. 54–64, Mar. 2018.
- D. Zhang, X. Peng, K. Pan, and Y. Liu, "A novel wind speed forecasting based on hybrid decomposition and online sequential outlier robust extreme learning machine," *Energy Convers. Manage.*, vol. 180, pp. 338–357, Jan. 2019.
- R. Singh and S. Balasundaram, "Application of extreme learning machine method for time series analysis," *Int. J. Electr. Comput. Eng.*, vol. 2, no. 8, pp. 549–555, 2007.
- H. S. Kim, R. Eykholt, and J. D. Salas, "Nonlinear dynamics, delay times, and embedding windows," *Phys. D, Nonlinear Phenomena*, vol. 127, nos. 1–2, pp. 48–60, Mar. 1999.
- A. M. Albano, J. Muench, C. Schwartz, A. I. Mees, and P. E. Rapp, "Singular-value decomposition and the grassberger-procaccia algorithm," *Phys. Rev. A, Gen. Phys.*, vol. 38, no. 6, pp. 3017–3026, Sep. 1988.
- J. Shi, Y.-B. Leau, K. Li, Y.-J. Park, and Z. Yan, "Optimization and decomposition methods in network traffic prediction model: A review and discussion," *IEEE Access*, vol. 8, pp. 202858–202871, 2020.



**JINMEI SHI** (Member, IEEE) is currently pursuing the Ph.D. degree in computer science with the Faculty of Computing and Informatics, Universiti Malaysia Sabah. She has been engaged in research and teaching for 13 years at Hainan Vocational University of Science and Technology. She has always been in the forefront of scientific research. She plays an active role in the construction of computer science and personnel training, and has good practical experience and academic foundation. Her research interests include network traffic prediction, algorithm analysis, and software. In recent years, she has published ten articles, three sponsored projects, three books, and ten awards above the provincial level.



**KUN LI** received the B.Sc. degree from the Shandong University of Science and Technology, in 2005, the M.Sc. degree from Liaoning Technical University, in 2008, and the Ph.D. degree from Northeastern University, Shenyang, China, in 2013. He is currently a Professor with the Faculty of Electrical and Control Engineering, Liaoning Technical University. His current research interests include complex industrial process modeling, intelligent optimal control, machine learning, and their application.



**YU-BENG LEAU** (Member, IEEE) received the B.Sc. degree in multimedia technology from Universiti Malaysia Sabah, Malaysia, the M.Sc. degree in information security from University Teknologi Malaysia, and the Ph.D. degree in internet infrastructures security from University Sains Malaysia. He is currently a Senior Lecturer of computer science with the Faculty of Computing and Informatics, Universiti Malaysia Sabah. His current research interests include intrusion alert detection and prediction, network security situation awareness, IPv6 security, the Internet of Things (IoT), and information centric networks (ICNs).



**HUANDONG CHEN** received the bachelor's degree in mathematics from Hainan Normal University, China, in 1993, and the master's degree in communication engineering from Hainan University, China, in 2007. He is currently a Professor with the College of Information Science and Technology, Hainan Normal University. His research interests include data processing, network management, remote sensing image analysis, and education technology.

...