A Hybrid Model for Financial Time Series Forecasting–Integration of EWT, ARIMA With The Improved ABC Optimized ELM

HE YU[ID]1, LI JING MING2, RUAN SUMEI2, AND ZHAO SHUPING3
[1]Anhui Vocational College of Press and Publishing, Hefei, China
[2]Anhui University of Finance and Economics, Bengbu 233030, China
[3]Hefei University of Technology, Hefei, China

Corresponding author: Li Jing Ming (homer998@163.com)

ABSTRACT The practical significances and complexities of financial time series analysis induce highly demand more reliable hybrid model that denoised the data efficiently, handled with both linear and nonlinear patterns in the data, to achieve more accurate results. This paper suggests a new forecasting hybrid model for financial time series data combined Empirical Wavelet Transform (EWT) technique with improved Artificial Bee Colony (ABC) algorithm, Extreme Learning Machine (ELM) neural network, and Auto-Regressive Integrated Moving Average (ARIMA) linear analysis algorithm. The EWT is used to decompose and denoise the data to reconstruct the data more suitable for forecast. The improvement of the ABC algorithm is according to the Good Point Sets (GPS) theory and adaptive Elite-based Opposition (EO) strategy (GPS-EO-ABC) to overcome the drawbacks of the original algorithm and enhance the optimization performance. The optimized ELM with GPS-EO-ABC, as well as the ARIMA, are utilized independently to generated different forecasting results and combined by the weight-based procession. We testify the performance of the proposed improved ABC algorithm by ten benchmark functions, simulating the proposed forecasting models by three financial time-series datasets. The results indicate that: (1) The proposed algorithm shows outstanding capacities in parameter optimization. The optimized ELM generated more stable and precise results compared with original ELM, ABC-ELM, single LSTM, and ANN; (2) The proposed hybrid model has not only effectiveness but also efficiencies in denoising data, correcting outliers, coordinating both linear and nonlinear patterns, its performance in financial time series forecasting is more excellent than existing hybrid models.

INDEX TERMS Financial time-series forecasting, EWT data prepossessing, ARIMA model, GPS-EO-ABC ELM model, hybrid model methodology.

I. INTRODUCTION

The statistical tools or techniques that reveal the rules and forecast the futures of phenomena through financial time-series data have a guiding signature for both governments and enterprises to predict revenues and costs, to evade financial risks available. Financial time-series analysis has always been the frontier field of financial engineering and enterprise risk management [1]. As a peculiar time-series data, financial time series data which derived from stock prices, interest rates, and so on, commonly include the following characteristics [1]–[3]:

(1) The randomness and complexity in the generation process;
(2) Most data collected with noise;
(3) Inherent non-linear relationships among data.

Thus, data decomposing, data denoising, and dealing with nonlinear data patterns properly are core points of putting financial time-series models into the application. Traditional time series analysis model such as Auto-Regressive Integrated Moving Average (ARIMA) [4], dynamics models such

The associate editor coordinating the review of this manuscript and approving it for publication was Sudipta Roy[ID].

as Generalized Auto Regressive Conditional Heteroskedasticity (GARCH) [5], and machine learning model such as Artificial Neural Network (ANN) [2] are not solely sufficient to address these problems [6]–[8]. Many studies have attempted to use the hybrid model methodology to denoise the data, increase the chance to capture the linear patterns with nonlinear patterns in the data, and improve the performance of the forecasting results [6]. The hybrid model in works of literature mainly comprises three categories as follows.

Hybrid models combined the aforementioned independent models. Since it reported in 1969, the hybrid methodology that a combination of forecasts has been attracting a lot of interest [9]. During the recent three years, the most commonly seen in combined linear time series model with machine learning model for financial time series data forecasts are ARIMA with ANN [10], ARIMA with SVM [11], ARIMA with DLN [12]. There are also many studies combined nonlinear dynamics models with machine learning models such as GARCH with SVM [13], *etc.* When employing this type of hybrid modeling, one model is based on the residual of the other one [6], [8]. If the subsequent model failed to model the residual of the prior model, the whole generalized performance of the hybrid model would be declined. Therefore, it is still needed further studies for the model architecture and the performance improvement by other techniques.

Hybrid models integrated the data decomposing and denoising module. This type mainly utilizes the data preprocessing approaches such as Wavelet Transform (WT), Empirical Mode Decomposition (EMD), Chaos Theory (CT), and so on, to establish reasonable prior information which made analysis grounding on statistical inferences obtain sensitivity [14]. For example, Bao *et al.* [15] proposed a hybrid model based on deep learning framework consists of WT and DLN. The stock price series data decomposed by the WT, which made it more suitable for forecasting. Likewise, the studies of Wadia and Ismail [16] and Raimundo and Okamoto [17] applied the WT into ARIMA and SVM, respectively. Awajan *et al.* [18] combined EMD with ARIMA to decompose the nonstationary and nonlinear stock market time-series data into Intrinsic Mode Functions (IMF) for ARIMA. The performance of the result is better than the original ARIMA model. Ravi *et al.* [19] used the CT to construct phase space followed by invoking Multi-Layer Perceptron (MLP), which testified that CT is an efficient way of eliminating volatility in financial time series data.

Hybrid models mixed the swarm intelligence algorithms. This type focus on exert the advantages of swarm intelligence algorithms. In recent decades, the swarm intelligence algorithm involved in financial linear time series models or machine learning models include Genetic Algorithm (GA) [20], Particle Swarm Optimization (PSO) algorithm [21], [22], Fireflies Algorithms (FA) [23], Ant Colony Optimization (ACO) algorithm [24], etc. Among the most popular swarm intelligence algorithms, Artificial Bee Colony (ABC) is on the basis of the particular intelligent

behavior of artificial honeybee swarms [25]–[27]. Compared with other algorithms, it has the advantages of fewer parameters and easy implementation [28]. The ABC algorithm and its optimized models have also been applied to the time series estimation or analysis [29]–[31]. However, the process of ABC optimization is easily fell into under-fitting solutions due to the uneven distribution of initial solutions in the data space [32]–[34], leading to the solutions of algorithm lacked stability and robustness. Besides, like other swarm intelligence algorithms, techniques to enhance the global search, as well as the local search ability of the whole algorithm is subject to high demanding [28], [35], which aims to avoid phenomena such as precocious convergence, and trapping into local optima solution. Hence, although the hybrid models depending on ABC or other swarm intelligence algorithms have some improvement, there still have the space to make performance better.

In summary, hybrid models mentioned above have certified the abilities to synthesize the information in financial time series parameters, reveal the rules among data, and generate more accurate predictions. Hence, this study proposed a hybrid financial time series forecasting models combining data preprocessing techniques Empirical Wavelet Transform (EWT), machine learning model Extreme Learning Machine (ELM), linear model ARIMA, and improved ABC algorithms. This assignment has explained the central importance of hybrid model methodology in financial time series analysis. The overall structure of the study takes the form of five sections, including the introduction section. Section II, Proposing the novelty and basic idea in the proposed hybrid model methodology. Section III, Introducing individual models and the necessary modules in the proposed hybrid model. Section IV, illustrating the testing experiment of the proposed algorithms and the simulation study on the proposed model. Conclusions from this paper presented in Section V.

II. METHODOLOGY

As in FIGURE 1 illuminated, the basic idea and data process route of the proposed hybrid model is: Firstly, the financial time series data submitted to the data-prepossessing model by using of EWT. Then, the reconstructed data yielded to individual forecasting models, including ARIMA, and the ELM improved by GPS-EO-ABC to obtain individual outcomes. At last, by using (1), the corresponding predictions combined
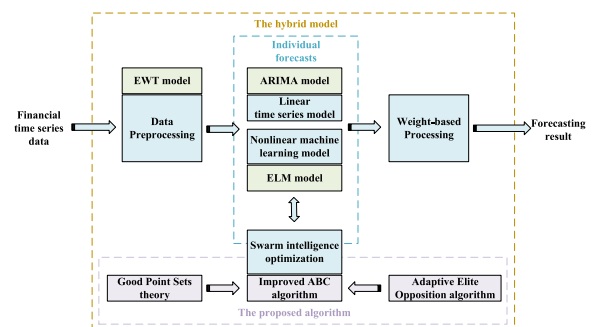

FIGURE 1. The basic idea of the proposed hybrid model.

to get the final forecasts Y.

$$Y = \sum_{t=1}^{N} \sum_{j}^{2} w_j e_{jt}, \quad j = 1, 2 \tag{1}$$

where w_j is the weighting of different models, and $\sum w = 1$, e_{jt} is the j th forecasting results of the time series dataset $\{X(t)\}$ in the t th plot. Details are illustrated in next section.

The novelties of the proposed hybrid model are as follows:

(1) EWT [15]–[17], [36], [37] is a novel data prepossessing model which has the formalism of WT and the adaptability of EMD [37], [38]. The prime advantage of EWT is to remedy the drawbacks that the primary WT requires specification of prior wavelet basis and parameters, as well as lack of self-adaptive method [39]. In this study, the EWT is applied to denoise and decompose the financial time series data to eliminate the impact of outliers, which made the time series data more suitable for the ARIMA and ELM forecast.

(2) In the existing literature, the ANN [2], [8], [29], [40]–[43] mainly based on backpropagation algorithms, it is generally far slower than required. As a Single Layer Feedforward Neural (SLFN) network, ELM has the advantages in learning rate and relatively accurate results [40], [41]. Previous studies had proved its superiority in fast time series analysis [44]. However, its application in financial time series forecasts usually do not have good prediction owing to the randomness and intermittence in the data, the enhancement for forecasting accuracy and reliability of individual ELM still need further study.

(3) The process of ABC optimization is easily fell into under-fitting solutions due to the uneven distribution of initial solutions in the data space [32]–[34], leading to the results of the algorithm lacked with stability and robustness. Besides, like other swarm intelligence algorithms, techniques to enhance the global search, as well as the local search ability of the whole algorithm, is subject to high demanding [28], [35], which aims to avoid phenomena such as precocious convergence, and trapping into local optima solution. Therefore, this study proposed an improved ABC based on Good Point Sets (GPS) theory and adaptive Elite Opposition-based (EO) strategy [25]–[31], [33] to address these problems.

(4) The aforementioned hybrid models have generated more precise and reliable forecasting results. Nevertheless, few studies have combined data preprocessing, linear time series model, machine learning model with improved swarm intelligence algorithms. This study would fill this gap.

III. MODEL

A. EWT MODEL

WT represents any function as a superimposition of a set of wavelets that provide crucial information about both time and frequency domains at the located positions [36]. When employing the WT technology to decompose the time series data, it is important to note that performing wavelet decomposition on time series must be independent for each partition period in order to avoid incorporating information from future data [45]. As a combination of WT and EMD, Empirical Wavelet Transform (EWT) proposed by J. Gilles [37], its core point is a self-adaptive improvement for traditional WT. The brief introductions are [38]: in signal space $L^2(R)$, the EWT consists of the empirical scaling function spectrum $\hat{\phi}_n$ and the empirical wavelets spectrum $\hat{\psi}_n$, which mathematically expressed by using (2) and (3) respectively.

$$\hat{\phi}_n = \begin{cases} 1 & \text{if } |\omega| \leq (1-\gamma)\omega_n \\ \cos[\frac{\pi}{2}\beta(\frac{1}{2\gamma\omega_n}|\omega| - (1-\gamma)\omega_n)] \\ & \text{if } (1-\gamma)\omega_n \leq |\omega| \leq (1+\gamma)\omega_n \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

$$\hat{\psi}_n = \begin{cases} 1 & \text{if } (1+\gamma)\omega_n \leq |\omega| \leq (1-\gamma)\omega_{n+1} \\ \cos[\frac{\pi}{2}\beta(\frac{1}{2\gamma\omega_{n+1}}|\omega| - (1-\gamma)\omega_{n+1})] \\ & \text{if } (1-\gamma)\omega_{n+1} \leq |\omega| \leq (1+\gamma)\omega_{n+1} \\ \sin[\frac{\pi}{2}\beta(\frac{1}{2\gamma\omega_n}|\omega| - (1-\gamma)\omega_n)] \\ & \text{if } (1-\gamma)\omega_n \leq |\omega| \leq (1+\gamma)\omega_n \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

where ω_n is the nth maxima of Fourier spectrum, γ is the real number belong to $[0, 1]$, $\beta(X)$ is an arbitrary $C^k([0, 1])$ function which satisfies the following properties:

$$\forall X \in [0, 1], \quad \beta(X) = \begin{cases} 0, & X \leq 0 \\ 1, & X \geq 0 \end{cases}, \beta(X) + \beta(1-X) = 1,$$

when

$$\gamma < \min_n(\frac{\omega_{n+1} - \omega_n}{\omega_{n+1} + \omega_n}) \tag{4}$$

the set $\left\{\phi_1(t), \{\psi_n(t)\}_{n=1}^{N}\right\}$ is an orthonormal basis of $L^2(R)$.

Like the classical wavelet transform, EWT can be implemented arising from the built of orthonormal basis. Therefore, the detailed coefficients between signal and empirical wavelet are calculated according to (5),

$$w_f^\varepsilon(n, t) = \langle f, \psi_n \rangle = \int_{-\infty}^{+\infty} f(\tau)\overline{\psi_n(\tau - t)}d\tau = \left(\hat{f}(\omega)\overline{\hat{\psi}_n(\omega)}\right)^\vee \tag{5}$$

the approximation coefficients between them are calculated by convention $w_f^\varepsilon(0, t)$ using (6):

$$w_f^\varepsilon(0, t) = \langle f, \{\phi_1\} \rangle = \int_{-\infty}^{+\infty} f(\tau)\overline{\phi_1(\tau - t)}d\tau = \left(\hat{f}(\omega)\overline{\hat{\phi}_1(\omega)}\right)^\vee \tag{6}$$

Algorithm 1 Data Preprocessing by EWT

1. Set the separation rate α;
2. Extend the financial time series data by Fourier transform, get the Fourier spectrum $\{X(t)\}_{t=1}^{N}$;
3. Detect the boundaries:
   Normalized $\{X(t)\}_{t=1}^{N}$ to [0,1];
   $a_{max} = \max\left(\{X(t)\}_{t=1}^{N}\right)$, $a_{min} = \min\left(\{X(t)\}_{t=1}^{N}\right)$;
4. Extract the boundaries:
   $a_t = a_{min} + \alpha \times (a_{max} - a_{min})$;
   $M = 0, J \in R^N$;
   For i = 1 : N;
     If $a_i \geq a_t$
       $M = M + 1$;
       $J(M) = w_i$;
     End If
   End For
   For j = 1 : M−1;
     $w_j = (J(j) + J(j+1)/2)$;
   End for
   Return w_j, j = 1 : M−1;
5. Establish the filter bank according to ***Equ. (5)***, ***Equ. (6)*** when ***Equ. (4)*** is satisfied;
6. Reconstructed the data according to ***Equ. (7), (8), (9)***.

---

where $f(T)$ is the function mapping signal, τ is the half the length of the "transition phase", and $\tau_n = \gamma \omega_n$, by using (7), the reconstruction is obtained:

$$f(T) = w_f^\varepsilon(0, t) \star \phi_1(t) + \sum_{n=1}^{N} w_f^\varepsilon(n, t) \star \psi_n(t)$$

$$f(T) = \left(\widehat{w_f^\varepsilon}(0, \omega) \star \widehat{\phi}_1(\omega) + \sum_{n=1}^{N} \widehat{w_f^\varepsilon}(n, \omega) \star \widehat{\psi}_n(\omega)\right)^\vee \tag{7}$$

Following this formalism, the empirical mode is given by (8) and (9) [46]:

$$f_0(t) = w_f^\varepsilon(0, t) \star \psi_1(t) \tag{8}$$

$$f_k(t) = w_f^\varepsilon(k, t) \star \psi_k(t) \tag{9}$$

Thus, the EWT is adaptively decomposition of the Fourier line, it does not necessarily correspond to the position of different modes. By using of the local max-min method, the period boundary of Fourier supports can be detected adaptively from the information contained in the processed spectrum.

Previous studies [15], [46], [47] have proved that data-decomposition level greater than or equal to three is sufficient to describe the data in a meaningful way. In this study, the data are decomposed to four levels. Relative pseudocode is listed in Algorithm 1.

B. ARIMA MODEL

The ARIMA pioneered by Box and Jenkins are the most widespread linear time series analysis model [1], [48]. It integrates the Auto Regressive (AR) model and Moving Average (MA) model to forecast from the parameters. Briefly introductions are as followings [4]:

Given time series with N data $\{x(t)\}_{t=1}^{N}$, an ARIMA (p, d, q) model can be defined as:

$$\phi(B)(1 - B)^d x(t) = \theta(B)\varepsilon_t \tag{10}$$

where p, q are the model orders, d is the difference order. B is the lag which made $(B)x(t) = x(t-1)$.

Partial Auto Correlation Function (PACF) is the commonly used lag identification functions, ε_t is the white noise that the mathematical expectation is zero, and the variance is one. $\phi(B)$ is the autoregressive model of order p and $\theta(B)$ is the moving average function of order q, which can be mathematically expressed by (11) and (12) respectively.

$$\phi(B) = 1 - \phi_1(B)^1 - \cdots - \phi_p(B)^p \tag{11}$$

$$\theta(B) = 1 - \theta_1(B)^1 - \cdots - \theta_q(B)^q \tag{12}$$

C. IMPROVED ABC

The ABC algorithm was proposed by Turkish academic Karaboga in 2005 [25]. It inspired by simulating bees honey-foraging behavior to solve multidimensional optimization problems, and it consists of three categories of artificial honey bees: employed foragers, onlookers, and scouts [26], [27]. In ABC, each food source regarded as a solution in the data space. The employed foragers and the onlookers are responsible for the exploitation of food sources, and the scouts are in charge of avoiding too few kinds of food sources [28]. The process of artificial honey bees foraging is essentially a solving process. The ABC algorithm steps can describe as follows [25]–[27]:

Step 1: Randomly generate initial honey bee population with the scale of M, the first half is the employed foragers, and the second half is the onlookers, each artificial bee x is a D dimensional vector;

Step 2: The employed foragers sent onto the food sources, and every forager updated the *fitness* of food sources according to (13):

$$new\, X_{ij} = x_{ij} + \gamma \times (x_{ij} - x_{kj}) \quad i, k = 1, \cdots, M/2,$$
$$j = 1, \cdots, D \tag{13}$$

where $newX_{ij}$ is the new location of food source, x_{ij} is the jth dimension of forager or food source X_i, x_{kj} is the jth dimension of random selected X_k, k is not equal to j, γ is a random number between $[-1,1]$;

Step 3: Calculate the probability that the food sources with which the onlookers prefer them, according to (14):

$$P = fitness_{ij} / \sum_{i=0}^{\frac{M}{2}} fitness_{ij} \tag{14}$$

Step 4: Send the scouts into the search area for avoiding too few food sources, memorize the best food source. If the requirements met, ABC ends. Otherwise, back to *Step 1*.

The principle of the proposed GPS-EO-ABC algorithms is to improve the initialization of artificial bee colony according to Good Point Sets theory to enhance the global searching

abilities. Then, to improve the way of food sources and bees updating based on Elite-based Opposition algorithms to enhance the local search abilities. Moreover, the inertia weights function is introduced, which can adaptively adjust the search step length to avoid the phenomenon of oscillation in the iterative late.

1) GOOD POINT SETS THEORY

The studies of Liu *et al.* [33], Li *et al.* [49], and Zhang [50] have proved that ABC, FA, and GA based on good point sets theory will provide better target solving data space and generate more diversified initial population, which can enhance the global search abilities and avoid quickly falling into optimal local value. The definitions of GPS theory are as below [51].

Definition 1: Let D dimensional Euclidean space, the unit cube: $\{G_D|xi\} = (x_1, x_2, \cdots, x_D)$, the good point: $r = \{r_1, r_2, \cdots, r_D\}$, $r \in G_D$, Compare r with sample sets $P_n(i)$ one by one, and the deviation sets is obtained, where: $P_n(i) = (r_1^{(n)}(i), r_2^{(n)}(i), \ldots, r_D^{(n)}(i))$, $1 \leq i \leq n$. Assemble deviations $\Psi(n)$ satisfied equation: $\Psi(n) = C(r, \varepsilon)n^{-1+\varepsilon}$ in the sets, named them Good Point Sets where $C(r, \varepsilon)$ is an arbitrary constant only related to r and $(\varepsilon > 0)$.
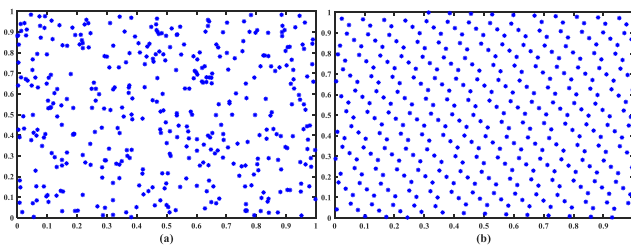

FIGURE 2. Initial population distribution between [0,1], (a) random method, (b) Good Point Sets method.

FIGURE 2 is bidimensional initial population distribution of 200 samples generated by random method compared with the good point sets method. Obviously, the latter is much more well-distributed than the former. In addition, the construction of the good point sets has nothing to do with the dimension of data space. This method can well address optimization problems in the field of multi-dimension functions. Thus, this study proposed a good point sets initialized ABC algorithm, to enhance the global search abilities of traditional ABC to avoid premature or underfitting solutions. The good point sets r_j calculated by using (15):

$$r_j = 2\cos(2\pi \times j/p) \quad (p-3)/2 \geq D \tag{15}$$

where D represents the dimension of the unit cube in the Euclidean space, p means the minimum prime that can make $(p-3)/2 \geq D$ hold. If the coding length of the solution is 20, which means the chromosome dimension of each bee equals to 20, the $p = 43$.

2) ELITE OPPOSITION BASED ALGORITHM

Elite Opposition-based (EO) machine learning algorithm was introduced by Tizhoosh H. R. in 2005 [52]. Its basic idea is

learning from past data or instructions, optimize estimated solutions, and search in large spaces for an existing solution [52], [53]. The studies of improved ABC depended on EO have proved it has the outstanding ability to coordinate global search with local search [54]. Thus, this study regards the food sources preferred by onlookers as the Elites, implementing the EO algorithms to maintain the diversity of population and strengthen the ability of local search. Related approaches are as follows [52]–[54]:

Definition 2: Let the Elites of the food sources (where artificial bees are send onto, they can also regard as the Elites of bees colony) in D dimensional space: $E = (x_{1j}, x_{2j}, \cdots, x_{nj})$, $j = 1, \cdots, D$ and $x_i \in [a_i, b_i]$, $i = 1, \ldots, n$. Where a_i and b_i is the lower and upper boundaries of x_i. The opposition point of Elites defined as $\tilde{x}_i$, where:

$$\tilde{x}_i = a_i + b_i - x_i \tag{16}$$

then calculate the fitness $f(x_i)$ and $f(\tilde{x}_i)$, the ABC algorithm continues with x_i if $f(x_i) \geq f(\tilde{x}_i)$, otherwise, with $\tilde{x}_i$.

This study regards the food sources preferred by onlookers as the Elite points, implementing the EO algorithms to maintain the diversity of population and strengthen the ability of local search. In addition, traditional swarm intelligence algorithms are easily sunk into oscillation near the peak value of function in the iterative late [21], [22], [32], [55], which is related to the search steps of artificial bees. In the iterative late, if the distance between the peak value with the t th times artificial bees is less than the step-length. At the $t+1$ th times iteration, the bees would move to the other side of the peak value, and back to the origin at the $t + 2$ th times iteration. Optimization with adaptive step-length could address this problem well [56], [57].

Taking the steady iterative state problem into account, an adaptive Elite-Opposition based algorithm referred to the inertia weight strategy of particle searching in the PSO algorithm [55] is presented in this study. The inertia weight W of artificial bees searching is calculating by (17):

$$W(t) = \omega - e^{|t/t_{max}|} \tag{17}$$

where t is the number of the iteration, t_{max} is the maximum iteration, ω is the constant coefficient. As t increases, the inertia weight $W(t)$ would decrease. The artificial bees searching steps can be decreased automatically by the application of $W(t)$, it can enhance the searching accuracy around the vicinity of Elite points in the iterative late. In this study, the calculation of the opposition point of Elites is not using (16) but using (18):

$$\tilde{x}_i = W(t) \times (a_i + b_i - x_i) \tag{18}$$

In summary, the pseudocode of GPS-EO-ABC is listed in Algorithm 2.

D. GPS-EO-ABC ELM MODEL

Extreme Learning Machine (ELM) proposed by Huang *et al.* [40], is attracting widespread interest due to

Algorithm 2 GPS-EO-ABC

1. Set the upper boundaries *ub,* lower boundaries *lb,* artificial bees population size *M,* the limits of food sources *Limit,* the maximum iteration t_{max}, the iterative counter $t = 0$; Generate the initial population *P* according to **Definition 1.** and **Equ. (15)**;
2. Send foragers onto the food sources, calculate the fitness $F(x)$;
3. Set the non-updated counter $Bas = 0$ $(Bas \leq Limit)$;
4. $t = 1$;
5. **While**$(t \leq t_{max})$
6.   **For** i $= (1 : M/2)$
7.   Foragers search for new food sources $newX_{ij}$ according to **Equ. (13)**;
8.   **End For**
9.   Calculate the selected possibility *P* according to **Equ. (14)**;
10.   Send onlookers onto the food source:
11.   **For** $l = (1:M/2)$
12.   **If** rand $(0, 1) < P(l)$
13.   Onlookers search for new food sources $newX_{lj}$ according to **Equ. (13)**;
14.   **Do** Elite-Opposition based searching algorithm according to **Equ. (18)**;
15.   Calculate $F(newX_{lj})$, updating the population based on Greedy selection;
16.   **Otherwise** $Bas = Bas + 1$
17.   **End If**
18.   **End For**
19.   Send scouts to find food sources that $Bas = Limit$, generate new one instead of it;
20. $t = t + 1$;
21. **End While**

the extremely learning rate and less intervention [41]. It is based on single layer feedforward artificial neural networks, which are simpler and more efficient than traditional neural networks back propagation algorithms. The standard SLFN with m hidden nodes can be written as [40], [41]:

$$t_j = \sum \beta_i \times g(\omega_i \cdot x_j + b_i) \tag{19}$$

where ω_i is the input weights connecting ith hidden node and the input nodes, $i = 1, \cdots , m$, β_i is the output weights connecting i th hidden node and the output node. b_i is the bias in the i th hidden node, x_j is the *j*th input variables, $j = 1, \cdots ,$ n, t_j is the *j*th output variable, $g(X)$ is the activation function mainly based on Sigmoid. It is expressed compactly by using (20):

$$T = H \times \beta \tag{20}$$

where:

$$H = \begin{bmatrix} g(\omega_1 \cdot x_1 + b_1) & \cdots & g(\omega_m \cdot x_1 + b_n) \\ \vdots & \ddots & \vdots \\ g(\omega_1 \cdot x_n + b_1) & \cdots & g(\omega_m \cdot x_n + b_n) \end{bmatrix},$$

$$T = \begin{bmatrix} t_1^T \\ \vdots \\ t_n^T \end{bmatrix}, \beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_n^T \end{bmatrix} \tag{21}$$

Thus, the output weight β calculated by using (22):

$$\beta = \bar{H} \times T \tag{22}$$

where $\bar{H}$ is the pseudo-inverse matrix of H.

1) PRINCIPLE OF GPS-EO-ABC ELM

The principle of the GPS-EO-ABC ELM model is: Laid down the topology of the model according to the input and output data. Thus, the coding length of artificial bees is determined. Each bee represents one weights and bias terms solution of ELM. Execute the GPS-EO-ABC algorithms as *Algorithm 1.*shown to find the best bee corresponding to the optimal fitness value. Decode the bee and assign the value to ELM, through the training and testing, the model would achieve a more accurate result to suit various modeling conditions. FIGURE 3 gives the flow chart of the GPS-EO-ABC ELM model.
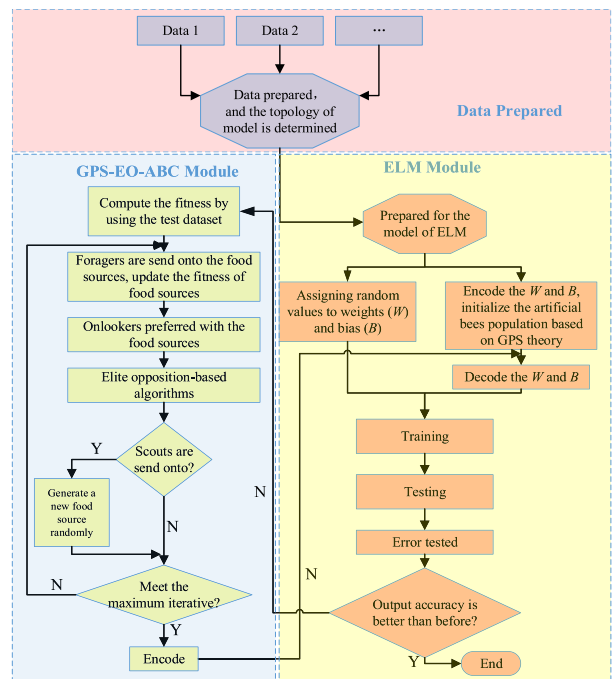
FIGURE 3. Flowchart of the GPS-EO-ABC ELM model.

2) STEPS

Data after decomposing and denoising by EWT are employed as the inputs of ELM. The aforementioned the GPS-EO-ABC ELM algorithm is implemented as follows:

Step 1 (Encoding): Assign random value to input weights and the bias terms *B* between $[-1,1]$, setting hidden neuron number, calculating the output weights and creating the ELM networks according to **Equ. (19)**, **Equ. (20)**, **Equ. (21)**, and **Equ. (22)**, determining the topology of ELM networks.

Encode the initial hidden node number N, weights W, and bias terms B as a whole, each artificial bee, including the value of them.

Step 2 (Parameters Initializing): Set the upper boundary ub, and the lower boundary lb, the size of bee colony M, the first half is the employed foragers, and the second half is the onlookers, the limits of food sources $Limit$, the maximum iteration t_{max}, the iterative counter t; Initialize the artificial bees population-based on GPS theory;

Step 3 (Fitness-Calculating): Decode the artificial bees, training and testing ELM network to get the error indices (MAE, MAPE, and RMSE), using the error indices as the fitness of each bee, and the minima of error indices is the maxima of fitness.

Step 4 (Bees and Food Sources Updating): Update the bees and food sources according to *Algorithm 1*.

Step 5 (Judging): Judge whether the iterative error satisfies the end conditions or not, if it is satisfied, the optima are obtained and given to the ELM network to produce the output result, model ends; Otherwise, judge whether reaches the maximum iteration t_{max} or not, if not, turn to *Step 4*. Otherwise, end.

E. THE ARCHITECTURE OF PROPOSED HYBRID MODEL

The architecture of the proposed hybrid model is displayed in FIGURE 4. And the stages are set as followings:

Stage 1: Utilize the EWT to preprocess the original financial time series data according to Algorithm 1.

Stage 2: Determine input variables by using PACF, operate ARIMA model according to *Equ. (10)*, *Equ. (11)*, *Equ. (12)*, as well as the GPS-EO-ABC ELM model according to aforementioned subsection, to generate forecasts.

Stage 3: Combined the forecasting results according to (1), in this study, the weighting of ARIMA forecasts and the GPS-EO-ABC ELM forecasts are both set to 0.5.

IV. EXPERIMENTS AND SIMULATION

The proposed algorithm and hybrid model have been implemented by using MATLAB R2018b, on a system with an Intel (R) Core (TM) i3-4150CPU @ 3.50 GHz, 4.00GB memory and WIN10 64-bit operating system.

A. EXPERIMENTS FOR PROPOSED ALGORITHM

To evaluate the performance of the improve ABC algorithm based on good point sets theory and adaptive elite-opposed strategy, this study compared it with traditional ABC algorithm, the Good Point Sets optimized ABC algorithm (GPS-ABC). Ten benchmark functions are selected to test the performance as listed in Table 1. Among them, $f_1(x)$ to $f_5(x)$ are bidimensional functions, and $f_6(x)$ to $f_9(x)$ are multi-dimensional functions, the dimension $d = 20$; The parameters of the three ABC-based algorithms are setting as follows: the population size M is set to 100, the maximum iteration t_{max} is set to 100, the limits of food sources $Limit$ is set to 50. In TABLE 2, Best, Worst, Mean and Var are used to record the best solution, the worst solution, the average
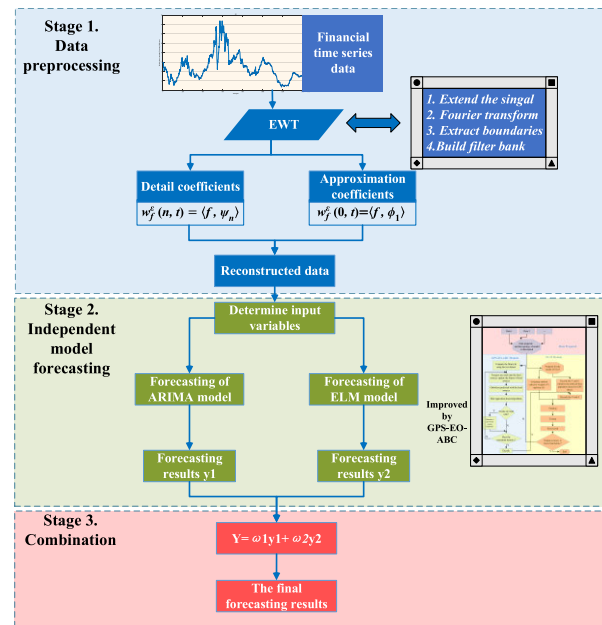

FIGURE 4. The architecture of proposed hybrid model.

TABLE 1. Description of the ten benchmark functions.

Name	Expression
Ackley	$f_1=(1.5-x_1+x_1 \cdot x_2)^2+(2.25-x_1+x_1 \cdot x_2)^2+(2.625- x_1+x_1 \cdot x_2{}^3)^2$
Booth	$f_2=(x_1+2x_2-7)^2+(2x_1+x_2-5)^2$
Bohachecsky	$f_3=x_1{}^2+2x_2{}^2-0.3\times \cos(3\pi x_1)\times \cos(4\pi x_2)+0.3$
Rosonbrock	$f_4=100\times(x_1{}^2-x_2)^2+(1-x_1)^2$
Matyas	$f_5=0.26\times(x_1{}^2+x_2{}^2)-0.482x_1x_2$
Sphere	$f_6=\sum_{i=1}^{d}(x_i{}^2)$
Rastrigrin	$f_7=\sum_{i=1}^{d}(x_i{}^2)-10\cos(3\pi x_i)^2+10$
Griewank	$f_8=\frac{1}{4000}\sum_{i=1}^{d}(x_i{}^2)+\prod_{i=1}^{d}\cos(x_i/i^{\frac{1}{2}})+1$
Dejong	$f_9=\sum_{i=1}^{d-1}100\times(x_{i+1}-x_i{}^2)^2+(1-x_i)^2$
Zakharov	$f_{10}=\sum_{i=1}^{d}(x_i{}^2)+(\sum_{i=1}^{d}0.5ix_i)^2+(\sum_{i=1}^{d}0.5ix_i)^4$

of solutions and the variance of solutions for twenty times independent experiments. FIGURE 5 shows the convergence curve of the three algorithms.

Take $f_6(x)$ for example, comparing with the original ABC, the best solution, the worst solution, the average solution of the twenty times experiments is decreased 3.76×10^{-2}, 7.62×10^{-1}, and 2.52×10^{-1} respectively. Comparing with the ABC only improved by GPS, the corresponding solutions is decreased 1.64×10^{-2}, 1.45×10^{-1}, 5.25×10^{-2} respectively. The performance of the proposed algorithm is the best, which indicates the abilities of GPS-EO-ABC to approximate to the most optimal solutions. Besides, the variance of the solutions is decreased 3.25×10^{-2}, 1.75×10^{-3} for ABC, GPS-ABC respectively, which means the proposed algorithm can generate more accurate solutions stably and robustly. It is because the GPS-EO-ABC is based on the uniform distributed initial population according to Good Point Sets theory and the elite-based opposition strategy introducing
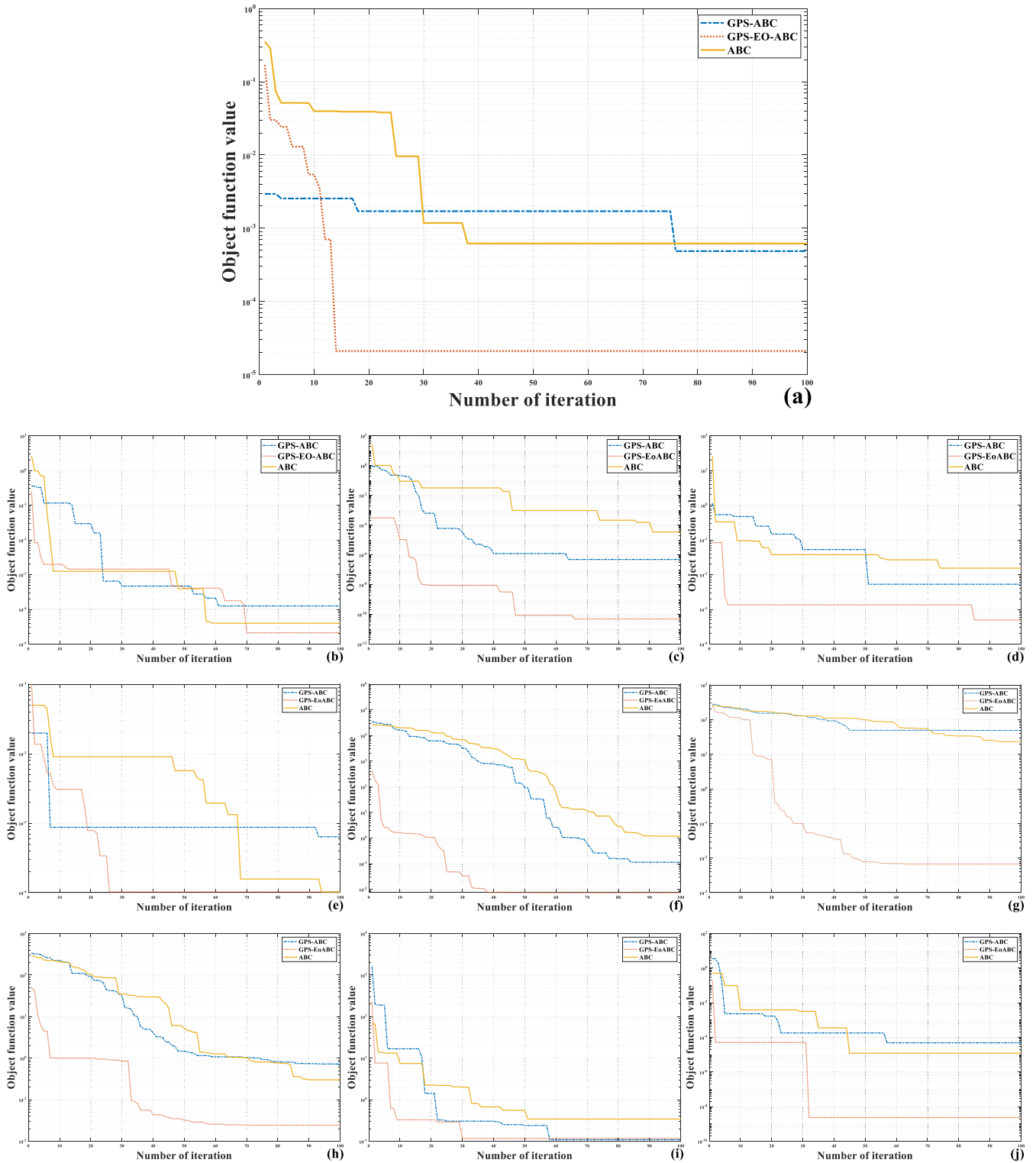
FIGURE 5. Convergence curves of the tested benchmark functions, (a) - (j): $f_1(x)$ - $f_{10}(x)$.

with inertia weights, which made artificial bees in the algorithm keep an adaptive step-length to search for the optima in most cases. Thus, the GPS-EO-ABC has dramatic global search facilities. Furthermore, in view of the converge curve, no matter in bi-dimension functions or in multi-dimension functions, GPS-EO-ABC displays the advantages in convergence speed and constancy. Although the GPS-ABC has some improvement for ABC, the GPS-EO-ABC can approach to the optimal solutions with more stable convergence curve, which demonstrates the superiority of the AEO

FIGURE 6. Financial time series data: (a) stock returns, (b) S&P 500 index, (c) closing index.

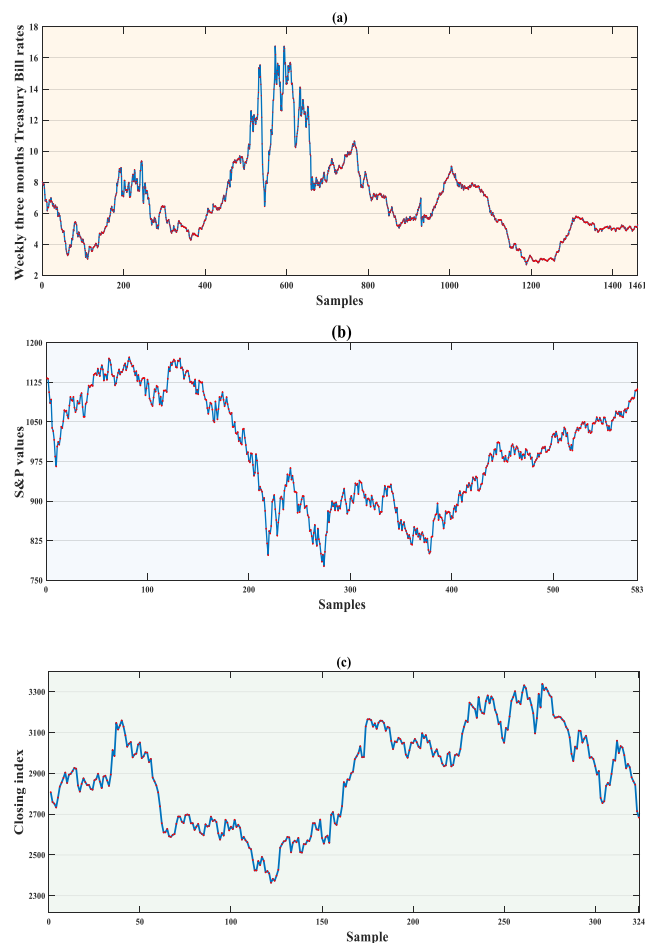**TABLE 2.** Results Of twenty times independent experiments.

	Algorithm	Best	Worst	Mean	Var
$f_1(x)$	ABC	2.2165e-06	3.2000e-03	8.3877e-04	1.0047e-06
	GPS-ABC	3.1733e-06	2.5000e-03	5.7582e-04	4.6617e-07
	GPS-EO-ABC	2.7469e-07	6.6949e-04	1.4819e-04	3.5445e-08
$f_2(x)$	ABC	2.1318e-06	5.4740e-04	9.9392e-05	1.6214e-08
	GPS-ABC	1.4122E-06	5.9236e-04	7.6618e-05	1.9875e-08
	GPS-EO-ABC	1.2561e-06	6.8600e-04	1.7681e-04	3.4341e-08
$f_3(x)$	ABC	3.8066e-11	1.3150e-07	2.5560e-08	2.0430e-08
	GPS-ABC	1.2593e-12	5.4094e-06	2.8129e-07	1.4574e-12
	GPS-EO-ABC	0	3.4040e-10	3.7446e-11	7.8319e-21
$f_4(x)$	ABC	1.6295e-04	1.3500e-02	4.0001e-03	1.5029e-05
	GPS-ABC	1.4431e-07	2.4600e-02	6.1003e-03	4.2712e-05
	GPS-EO-ABC	6.8387e-07	1.6800e-02	2.2000e-03	1.6959e-05
$f_5(x)$	ABC	7.2822e-06	1.3000e-03	3.8224e-04	1.6789e-07
	GPS-ABC	3.3583e-06	1.1000e-03	1.8521e-04	6.8076e-08
	GPS-EO-ABC	2.5848e-11	1.2232e-04	2.3026e-05	8.8168e-10
$f_6(x)$	ABC	3.9000e-02	7.8610e-01	2.6130e-01	3.2600e-02
	GPS-ABC	1.7800e-02	1.6940e-01	6.1600e-02	1.8000e-03
	GPS-EO-ABC	1.4000e-03	2.4600e-02	9.1000e-03	5.0384e-05
$f_7(x)$	ABC	1.0779e+01	5.0334e+01	2.8356e+01	8.5465e+01
	GPS-ABC	1.8727e+01	3.8898e+01	2.8000e+01	3.4214e+01
	GPS-EO-ABC	2.2000e-03	1.8770e-01	4.3201e-02	2.1000e-03
$f_8(x)$	ABC	1.7940e-01	0.1062e+01	6.8730e-01	6.7500e-02
	GPS-ABC	1.7600e-02	0.1169e+01	5.4340e-01	1.1580e-01
	GPS-EO-ABC	1.4700e-02	5.5880e-01	2.1280e-01	2.3300e-02
$f_9(x)$	ABC	4.2000e-03	8.5500e-01	2.2660e-01	6.1100e-02
	GPS-ABC	3.7100e-02	4.0520e-01	1.6010e-01	9.3000e-03
	GPS-EO-ABC	1.1163e-04	1.3000e-01	3.3200e-02	1.2000e-03
$f_{10}(x)$	ABC	1.2534e-08	1.4120e-03	1.2010e-04	9.7444e-08
	GPS-ABC	5.3863e-08	7.1548e-04	1.2031e-04	4.1827e-08
	GPS-EO-ABC	1.0189e-08	5.7656e-05	8.3273e-06	2.2623e-10

algorithm. Therefore, the GPS-EO-ABC could provide a outstanding way to enhance the local search abilities.

B. SIMULATIONS

1) DATA

Three financial time series data are utilized to measure the performance of hybrid model results, the details are displayed in TABLE 3 and FIGURE 6. These data are all wide-sense stationary time series and fluctuate greatly.

The Mean Absolute Error (MAE), Mean Absolute Percent Error (MAPE), and Root Mean Square Error (RMSE) are used to be the optimization functions which are calculated by (23), (24), (25) respectively.

$$MAE = \frac{1}{T}\sum_{t=1}^{T}\left|\frac{Y - Y'}{Y}\right| \quad (23)$$

$$MAPE = \frac{1}{T}\sum_{t=1}^{T}\left|\frac{Y - Y'}{Y}\right| \quad (24)$$

$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(Y - Y')^2} \quad (25)$$

TABLE 3. Details Of the financial time series data.

No.	Time series	Description	Size (training, testing)
Data 1	Treasury bill rates	Weekly 3-month treasury bill rates (Jan., 1970- Dec., 1997)[1]	1461 (75%, 25%)
Data 2	S&P500 values	Daily values of S&P 500 index (Sep., 2001 – Dec., 2003) [1]	583 (75%, 25%)
Data 3	Closing index	Daily closing index of Shanghai composite index (Jan., 2010- Dec., 2010)[2]	324 (75%, 25%)

[1]Datasets in [1], https://faculty.chicagogsb.edu/ruey.tsay/teaching

[2]Datasets from CSMAR research data service, https://www.gtarsc.com

where Y is the observed value, $Y`$ is the predicted value, t is the tth time period in the dataset, T is the total time periods.

2) DATA PREPOSSESSING

As FIGURE 7 illustrated, by using EWT to decompose the data, the information can be segmented in the Fourier spectrum well. In this study, prior to forecasting by ARIMA and GPS-EO-ABC ELM, the financial time series data are decomposed to four levels, the four uncorrelated subseries are extracted and one residual is obtained after extraction. The four subseries are reconstructed and the residual is ignored which helps to denoise the original time series. Besides, the PACF values between each reconstructed data and the antecedent values at lag k to obtain the input variables for individual models are shown in FIGURE 8. It is clearly that the PACF values for Data 2 and Data3 is significant at lag 1 after the elimination of internal correlation, and no PACF plot with 95% confidence bands for Data1, which demonstrated that the data at tth time strongly affect the prediction that occurs one time step ahead, while the effect of the actual data from t - 2th time to $t - k$th time to the one-step ahead prediction is not so obvious. Thus, the values at lag 1 are selected as the input variables for training.

3) PARAMETER SETTING

To verify the efficiency of improved ELM model on basis of GPS-EO-ABC, the original ELM, ABC ELM, GPS-ABC ELM are constructed for financial time series analysis. Besides, to evaluate the performance of the proposed hybrid model, the individual models that ELM, basic ANN, Long Short - Term Memory (LSTM) network, the hybrid models with or without EWT, the hybrid models with or without ARIMA, and the ARIMA-ANN (linear parts predict by ARIMA, and the residuals predict by ANN) are constructed. Parameters of ELM related models are set as follows: the reconstructed data are normalized between [0, 1]. The sigmoid function is used as the activation function, and the hidden neurons number in ELM is set to 10. The parameters of model related to ABC algorithms are set as: the population size M set to 100, the number of employed foragers and onlookers are $M / 2$, the limits of food sources $Limit$ set to 50, the maximum iterative t_{max} is set to 50. For basic ANN that
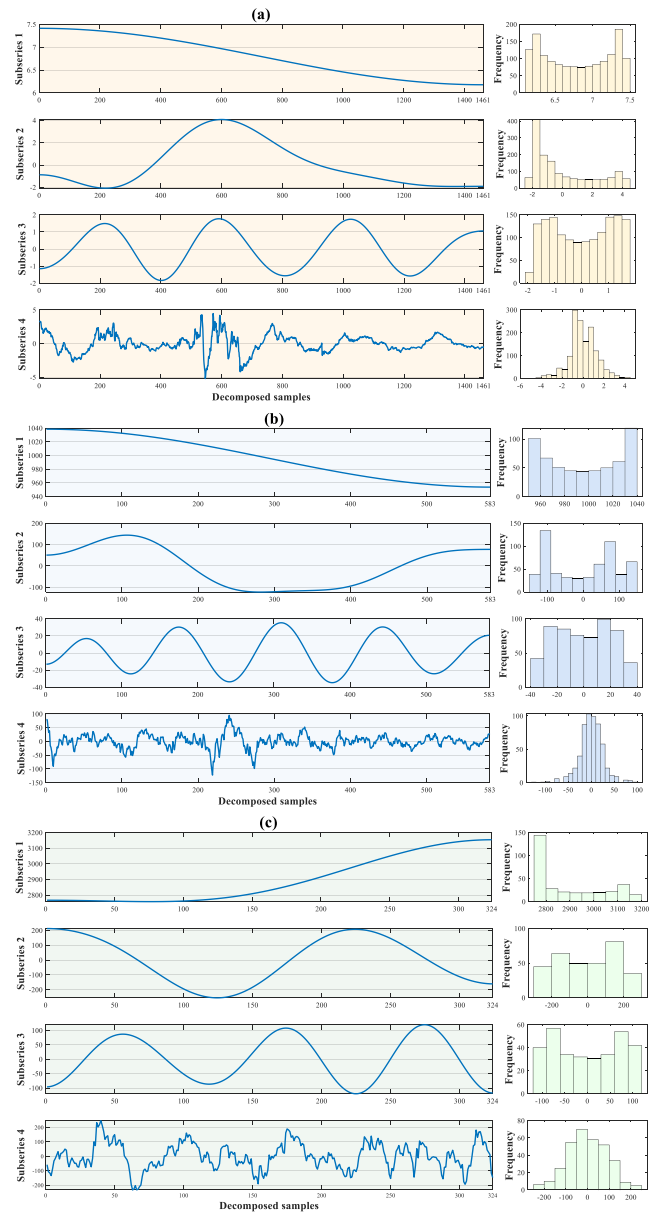

FIGURE 7. Data decomposing by EWT: (a) decomposed subseries and frequency of Data 1, (b) decomposed subseries and frequency of Data 2, (c) decomposed subseries and frequency of Data 3.

used back propagation algorithm, the LM training function is exploited to speed up the training process, the training time is set to 100, the training target is set to 0.01, the learning
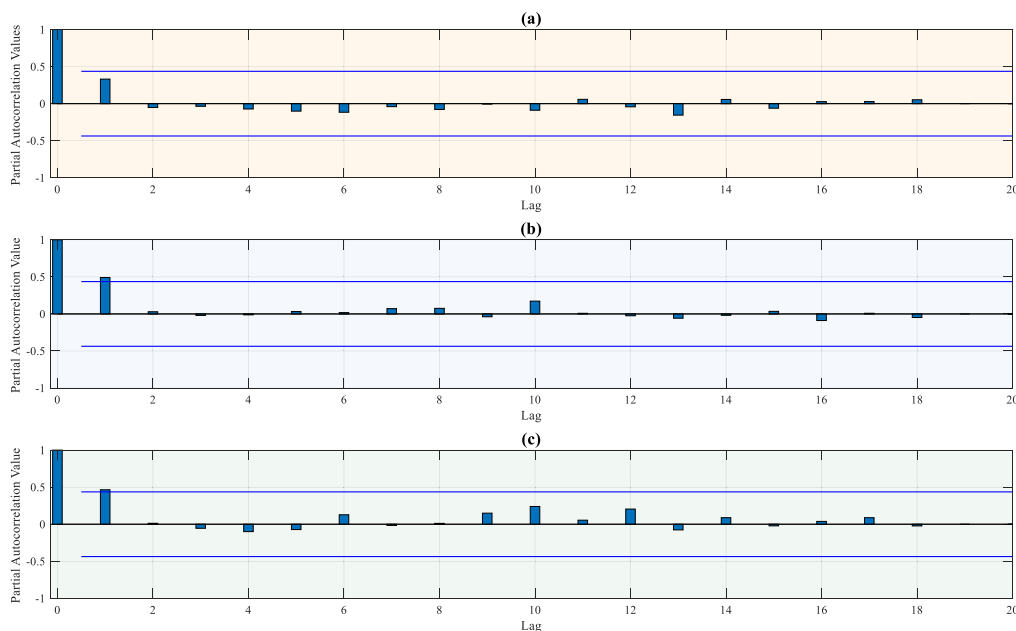
FIGURE 8. PACF values of reconstructed time series: (a) Data1, (b) Data 2, (c) Data3.

TABLE 4. Forecasting results of individual models.

Model	One-step ahead		Two-step ahead		Three-step ahead	
	RMSE	MAPE (%)	RMSE	MAPE (%)	RMSE	MAPE (%)
Data 1						
ELM	0.1354	2.5682	0.1311	2.3968	0.2260	4.7087
ABC ELM	0.1258	2.2491	0.1282	2.3075	0.1698	3.2356
GPS-ABC ELM	0.1262	2.2152	0.1262	2.2176	0.1671	3.1422
GPS-EO-ABC ELM	0.1247	2.1615	0.1246	2.1715	0.1650	3.0614
Data 2						
ELM	8.6330	0.6835	11.2672	0.9019	13.8218	1.1088
ABC ELM	8.5031	0.6723	10.8230	0.8700	13.0563	1.0631
GPS-ABC ELM	8.4890	0.6711	10.7661	0.8656	13.0238	1.0576
GPS-EO-ABC ELM	8.4615	0.6706	10.7647	0.8656	12.9547	1.0460
Data 3						
ELM	56.4723	1.4472	76.3949	2.0392	109.4324	2.9366
ABC ELM	51.6485	1.3157	75.0855	1.9916	91.6051	2.5233
GPS-ABC ELM	51.4533	1.3083	75.6408	2.0058	91.2436	2.5212
GPS-EO-ABC ELM	51.4449	1.3077	74.1179	1.9591	90.7904	2.5001

rate is set to 0.1. For LSTM, the number of hidden units is set to 10, the max epoch is set to 100, the initial learning rate is set to 0.1. Each experiment is repeated ten times, and results with the best accuracy are presented.

4) RESULTS AND COMPARISONS

The forecasting results of abovementioned individual ELM related models for Data 1, Data 2, and Data 3, are listed in TABLE 5. The forecasting results compared with observed time series are displayed in FIGURE 9. The forecasting results of GPS-EO-ABC have the best performance compared with other algorithms. Take Data 3 for example, comparing

with original ELM, the MAPE and RMSE for one-step ahead are decreased by 0.1395%, and 5.0274 respectively, for two-step ahead are decreased 0.0801%, and 2.2770 respectively, for three-step ahead are decreased 0.4365%, and 18.6420 respectively. Take Data 1 for instance, comparing with ABC-ELM, the corresponding indices are decreased by 0.0876%, and 0.0011; 0.1360% and 0.0036; 0.1742% and 0.0048 respectively. These results indicated the optimized ELM model based on GPS and EO have a significant improvement in the accuracy in most cases.

The predicting results of hybrid models for Data1, Data 2, and Data 3, are presented in TABLE 6, TABLE 7, and

TABLE 5. Forecasting results of hybrid models for Data1.

Models	One-step ahead		Two-step ahead		Three-step ahead	
	RMSE	MAPE (%)	RMSE	MAPE (%)	RMSE	MAPE (%)
AR	0.0749	1.2548	0.1184	2.0196	0.1509	2.6652
LSTM	0.4430	10.3869	0.2626	5.7049	0.4962	11.5679
ANN	0.1299	2.3432	0.1559	2.8291	0.1987	3.3813
ELM	0.1354	2.5682	0.1311	2.3968	0.2260	4.7087
GPS-EO-ABC ELM	0.1247	2.1615	0.1246	2.1715	0.1650	3.0614
EWT AR	0.0749	1.2548	0.1184	2.0197	0.1509	2.6652
EWT LSTM	0.2195	4.3815	0.3749	8.5067	0.2129	4.3217
EWT ELM	0.0933	1.7037	0.1496	2.9789	0.1775	3.4936
EWT GPS-EO-ABC ELM	0.0766	1.2966	0.1262	2.2160	0.1670	3.1022
AR LSTM	0.4794	10.8761	0.0925	1.7543	0.4926	10.2023
ARIMA ANN	0.1840	3.5800	0.1320	2.3594	0.5913	11.1653
AR ELM	0.0956	1.8792	0.1220	2.2033	0.1545	2.9508
AR GPS-EO-ABC ELM	0.0754	1.2655	0.1177	2.0062	0.1490	2.7441
Proposed hybrid model	0.0383	0.6487	0.0589	0.9875	0.0747	1.3078

FIGURE 9. The one step ahead forecasting values of individual models versus observed values in the testing samples: (a) Data1, (b) Data 2, (c) Data3.

TABLE 7. Figure 10 shows the one-step ahead forecasts generated by the proposed hybrid model. The first five lines in each table are the results of individual models without data prepossessing by EWT, the proposed optimized ELM not only generated more accurate results than single ELM, but also had good performances comparing with single LSTM and basic ANN. For example, the decrease in RMSE and MAPE of Data 1 for LSTM is 0.3183 and 8.2254%;

0.1380 and 3.5334%; 0.3312 and 8.5065% with different steps ahead. For ANN, the decrease is 0.0052 and 0.1817%; 0.0313 and 0.6576%; 0.0337 and 0.3199%.

The middle five lines are the hybrid models combined EWT with each individual models. Take Data 2 the models GPS-EO-ABC ELM with or without EWT for instance, the error indices of one-step ahead, two-step ahead, and three-step ahead are decreased by 0.0363 and 0.0030%,
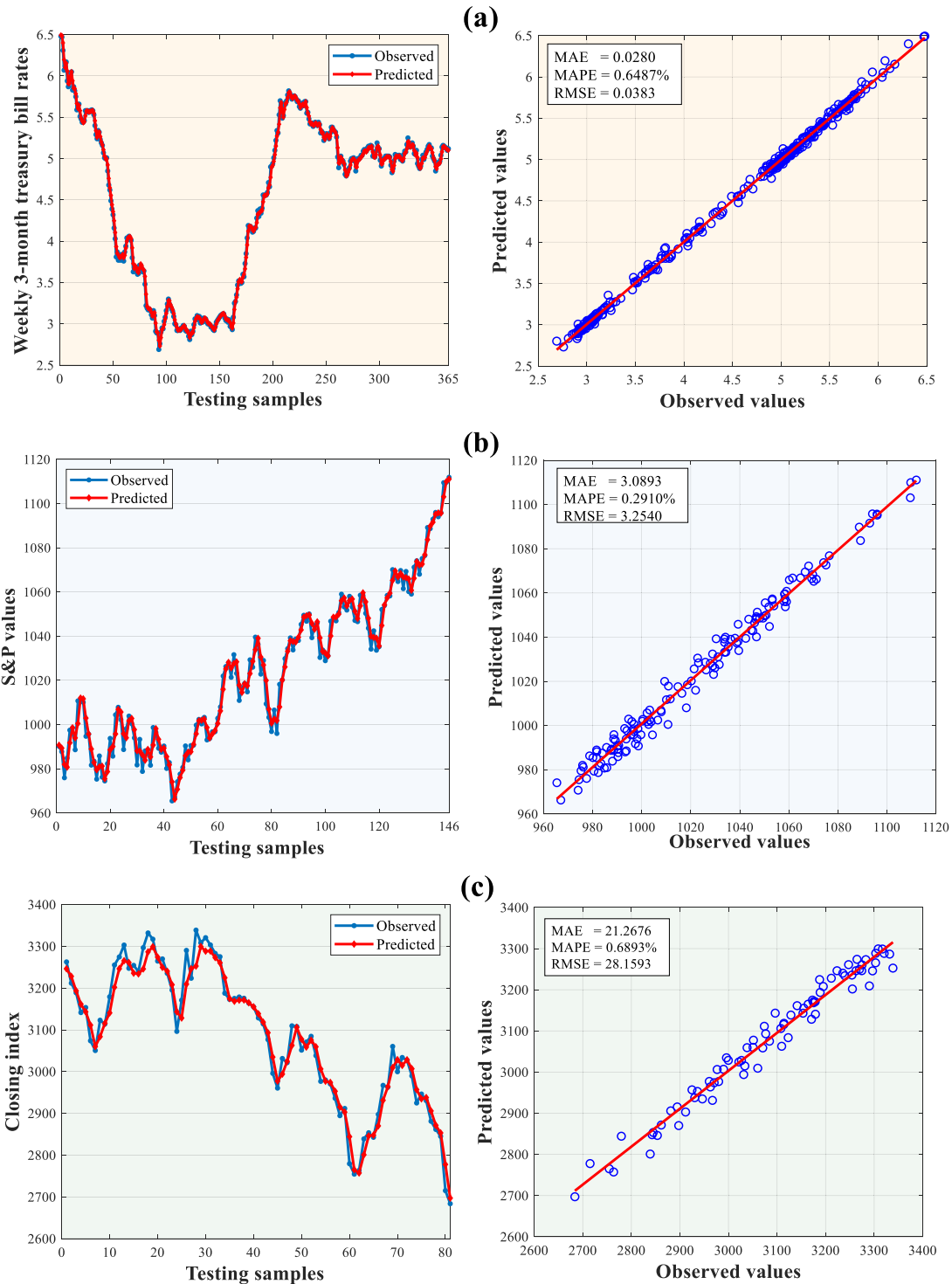
FIGURE 10. One step ahead forecasts of the proposed hybrid model, on the left is the predicted values versus observed values, the right is the scatter plots.

increased by 0.0202 and 0.0015%, decreased by 0.0349 and 0.0030% respectively. The ehancement of accuracy performance by data decomposing and denoising through EWT is not remarkable in two-step ahead. It reflected that the EWT can achieve more accurate results in most environments.

The facilities of EWT technique are also reflected in the last four lines of each table, which lists forecasts of the hybrid models combine linear financial time series analysis model ARIMA with machine learning model ELM, LSTM, and GPS-EO-ABC without EWT, as well as the

TABLE 6. Forecasting results of hybrid models for Data2.

Models	One-step ahead		Two-step ahead		Three-step ahead	
	RMSE	MAPE (%)	RMSE	MAPE (%)	RMSE	MAPE (%)
AR	8.3181	0.6516	10.5518	0.8424	12.7952	1.0377
LSTM	10.5525	0.8317	10.8436	0.8708	13.5926	1.0844
ANN	9.6898	0.7691	13.4744	1.0928	16.2483	1.2938
ELM	8.6330	0.6835	11.2672	0.9019	13.8218	1.1088
GPS-EO-ABC ELM	8.4615	0.6706	10.7647	0.8656	12.9547	1.0460
EWT AR	8.3089	0.6510	10.5518	0.8425	12.7893	1.0376
EWT LSTM	9.5030	0.7583	10.7477	0.8672	12.7908	1.0243
EWT ELM	8.6192	0.6830	11.3071	0.9052	13.9827	1.1116
EWT GPS-EO-ABC ELM	8.4252	0.6676	10.7849	0.8671	12.9198	1.0430
AR LSTM	5.5152	0.4401	6.1314	0.4924	8.1291	0.6702
AR ANN	9.5287	0.7508	11.1210	0.8905	14.3085	1.1639
AR ELM	3.0107	0.2456	6.6621	0.5498	9.8761	0.8388
AR GPS-EO-ABC ELM	4.1040	0.3189	5.2735	0.4227	6.7267	0.5218
Proposed hybrid model	3.2540	0.2910	5.1351	0.4124	6.2704	0.4931

TABLE 7. Predicting results of hybrid models for Data3.

Models	One-step ahead		Two-step ahead		Three-step ahead	
	RMSE	MAPE (%)	RMSE	MAPE (%)	RMSE	MAPE (%)
AR	50.8105	1.2740	72.6813	1.9619	90.2930	2.5239
LSTM	53.8322	1.4077	88.1053	2.3051	108.9569	2.8742
ANN	56.3667	1.4459	74.7542	2.0059	91.3447	2.5422
ELM	56.4723	1.4472	76.3949	2.0392	109.4324	2.9366
GPS-EO-ABC ELM	51.4449	1.3077	74.1179	1.9591	90.7904	2.5001
EWT AR	50.7852	1.2738	72.6888	1.9615	90.2785	2.5237
EWT LSTM	51.3230	1.2860	87.8343	2.0005	102.7978	2.8686
EWT ELM	54.51963	1.4086	77.6023	2.0974	107.8072	2.9123
EWT GPS-EO-ABC ELM	51.2939	1.3084	74.8948	1.9842	90.5755	2.4953
AR LSTM	31.1675	0.8278	34.4220	0.8489	35.1910	0.9430
ARIMA ANN	68.4619	1.6830	73.3863	1.9762	86.9077	2.3516
AR ELM	51.6031	1.3162	72.2850	1.8862	106.3924	2.7696
AR GPS-EO-ABC ELM	28.4102	0.6944	39.3810	1.0124	52.0086	1.3485
Proposed hybrid model	28.1593	0.6893	38.1921	0.9888	51.8239	1.3431

ARIMA-ANN that linear partterns are submitted to ARIMA and nonlinear patterns are submitted to ANN. The results indicate that: The hybrid model methodology combined linear time series model with machine learning model is an relative efficient way to improve the forecasts. Given Data1 GPS-EO-ABC ELM related models as a example. Comparing with individual model, the MAPE and RMSE of AR GPS-EO-ABC ELM are decreased by -1.1737%, and 2.0861 for one-step ahead, by -1.9139%, 2.0538 for two-step ahead, by -2.6172%, and 2.9124 for three-step ahead. Comparing with hybrid models combined with EWT, the corresponding indices are decreased by -1.2098%, and 1.2212; -1.9126%, and 2.0983; -2.6183% and 2.9532. The increasements in MAPE indicates the influence of outliers on the forecasting results. Data denoising by EWT, the outliers could be corrected. Therefore, the proposed hybrid model not only decreased MAE and RMSE further more, the MAPE decreased 0.6168%, 1.0187%, and 1.4363% for different steps ahead respectively, which shows that the most excellent forecasting performance of the proposed hybrid model at each data.

V. CONCLUSION
In this study, we proposed a hybrid model for financial time series forecasts that incorporate with EWT, optimized ELM based on GPS-EO-ABC, as well as the linear time series analysis model ARIMA. The EWT is used to decompose and denoise the data, aiming to eliminate the impact of noise and correct the outliers in the original time series, which made it more suitable for the forecast. The improvement of the ABC based on the GPS theory and adaptive EO algorithm, focusing on enhancing both the global search abilities and local search

facilities of the original ABC algorithm, which made further optimization on the performance of ELM. The reconstructed data by EWT is submitted to ARIMA and the GPS-EO-ABC-ELM model for forecasting respectively, and independent results are combined by weight-based procession, purposing to take the advantage of ELM in extremely learning rate, capture the linear patterns and nonlinear patterns in the time series data adequately, and achieve more accurate results. We used the MAPE and RMSE as the error indices to testify each model. The results of experiments and simulations shown that the proposed algorithm is robust and convergent. Improved ELM model by GPS-EO-ABC can generate more accurate and stable forecasting results compared with the original ELM, ABC-ELM, GPS-ABC, the single LSTM, and basic ANN. The hybrid model, combined with EWT, GPS-EO-ABC ELM, and ARIMA, has the most excellent forecasting performance. The proposed methodology is not only effective but also efficient. Information derived from the proposed model could satisfy the practical requirements of financial time-series analysis.

APPENDIXES

APPENDIX I.

In order to strengthen the theoretical basis of GPS-EO-ABC, this section theoretically analyses the convergence by building the absorbed Markov models.

Definition 3: Given the tth iteration bee colony of GPS-EO-ABC algorithm $X(t)$, n times arbitrary different iteration are set as $t_1 < t_2 < \cdots < t_n \in T$.

If satisfy $P[X(t_n) \le X_n | X(t_{n-1}) = X_n, \ldots, X(t_2) = X_2, X(t_1) = X_1] = P[X(t_n) \le X_n | X(t_{n-1}) = X_{n-1}]$, the $\{X(t)\}_{t=0}^{\infty}$ is a Markov process.

Theorem 4: Let the random process corresponding to the artificial bee colony sequence based on the GPS-EO-ABC algorithm $\{X(t)\}_{t=0}^{\infty}$, the $\{X(t)\}_{t=0}^{\infty}$ has Markov property.

Proof: The Optimization process of GPS-EO-ABC includes four steps. These steps are all independent process, the most optimal forager and food source in the tth iteration is only determined by the t-1th. When $X(0)$ is the initialized population, the selection of initial data located in good point data based on Good Point Sets theory, thus the initial distribution of $X(0)$ bees is uniform in the whole sample space with unchanged values. The state of $X(t)$ bees only determined by $X(t$-1), it has no relationships with $X(t-1), \ldots, X(2), X(1)$, Moreover, the $\{X(t)\}_{t=0}^{\infty}$ is a random procession.

Hence, $P[X(t) | X(t-1), \ldots, X(1), X(0)] = P[X(t) | X(t-1)]$, the $\{X(t)\}_{t=0}^{\infty}$ has Markov property.

Definition 5: Given the Markov process $\{X(t)\}_{t=0}^{\infty}$, and the optimal solution space $Z^* \subset Z$. If satisfy the probability: $P[X(t+1) \notin Z^* | X(t) \in Z^*] = 0$. The $\{X(t)\}_{t=0}^{\infty}$ is an absorbed Markov process.

Theorem 6: Let the random process corresponding to the artificial bee colony sequence based on the GPS-EO-ABC $\{X(t)\}_{t=0}^{\infty}$, the $\{X(t)\}_{t=0}^{\infty}$ is an absorbed Markov process.

Proof: According to Theorem 4, after each iterative, the global optimal solution of the algorithm remains updated or unchanged.

If $X(t) \in Z^*$, then $X(t) \in Z^*$, thus $P[X(t+1) \notin Z^* | X(t) \in Z^*] = 0$, the $\{X(t)\}_{t=0}^{\infty}$ is an absorbed Markov process.

Definition 7: Given the Markov process $\{X(t)\}_{t=0}^{\infty}$, and the optimal solution space Z^*, the probability of the tth iterative reaches the optimal solution as $\varphi(t) = P[X(t) \cap Z^*] \ne \emptyset$. If = 1, the $\{X(t)\}_{t=0}^{\infty}$ converge.

Theorem 8: Given the absorbed Markov process corresponding to GPS-EO-ABC $\{X(t)\}_{t=0}^{\infty}$ and the optimal solution space Z^*, if $P(X(t) \cap Z^* \ne \varphi | P(X(t-1) \cap Z^*) = \varphi) \ge d(t-1) \ge 0$, and $\prod_{i=0}^{t}[1 - d(i)] = 0$, then $\varphi(t) = 1$. Where $\in \varphi(t) = P(X(t)Z*)$.

Proof: According to Theorem 4 and Theorem 6, the $\{X(t)\}_{t=0}^{\infty}$ is an absorbed Markov process. Thus, $P(X(t) \cap Z^* \ne \varphi | X(t-1) \cap Z^* = \varphi) = 1$.

And according to the total probability formula:

$$
\begin{aligned}
\Phi(t) &= [1 - (t-1)] \times P(X(t) \cap Z^* \\
&\ne \varphi | X(t-1) \cap Z^* = \varphi) + (t-1) \times P(X(t) \cap Z^* \\
&\ne \varphi | X(t-1) \cap Z^* = \varphi),
\end{aligned}
$$

It is deduced as follows:

$$
\begin{aligned}
\Phi(t) &= [1 - \varphi(t-1)] \times P(X(t) \cap Z^* \\
&\ne \varphi | X(t-1) \cap Z^* = \varphi) + (t-1) \\
&\Rightarrow 1 - (t) = P(X(t) \cap Z^* \ne \varphi | X(t-1) \cap Z^* = \varphi) \\
&\quad \times [1 - (t-1)] \\
&\Rightarrow 1 - \varphi(t) \le [1 - d(t-1)] \times [1 - \varphi(t-1)] \\
&= [1 - \varphi(0)] \\
&\quad \times \prod_{i=0}^{t}[1 - d(i)] \\
&\Rightarrow \lim_{t \to \infty} \varphi(t) \ge 1 - [1 - \varphi(0)] \times \lim_{t \to \infty} \prod_{i=0}^{t}[1 - d(i)] \\
&\Rightarrow \lim_{t \to \infty} \varphi(t) \ge 1.
\end{aligned}
$$

For $\lim_{t \to \infty} \varphi(t) \le 1$, which means $\lim_{t \to \infty} \varphi(t) = 1$, the GPS-EO-ABC can reach the optimal solutions with the probability of 1, prove up.

APPENDIX II.

The theorical analysis of the financial hybrid model combined EWT with ARIMA and GPS-EO-ABC ELM based on convex quadratic programming Kuhn-Tucker (K-T) conditions.

Definition 9: Let $X = (x_1, x_2, \cdots, x_N)^{\mathrm{T}}$ the points in N-dimensional Euclidean space. For programming:

$$
\begin{cases}
\mathrm{Min}\, f(X) \\
h_i(X) = 0, \quad i = 1, 2, \cdots, m \\
g_j(X) \ge 0, \quad j = 1, 2, \cdots, l.
\end{cases}
$$

If $f(X), g_j(X)$ are convex functions, $h_i(X)$ is linear function, it is a convex programming function.

Theorem 10: Let the financial time series with N data $\{X(t)\}_{t=1}^{N}$, modeling combined ARIMA and GPS-EO-ABC ELM for time series forecasting is a convex programming function.

Proof: The hybrid model chooses MAE, MAPE, and RMSE as the error indices for model accuracy forecasting,

the smaller indices value indicates the better model performance. It is known that the ARIMA is linear function. The GPS-EO-ABC ELM can be written as a quadratic optimization problem:

$$\min \frac{1}{2}\|\beta\|^2 + \frac{1}{2}C\sum_{i=1}^{N}\|\xi_i\|^2$$
$$s.t.\ h(x_i)\beta = t_i^T - \xi_i^T, \quad i = 1, \cdots, N,$$

let the predictions of ARIMA for $\{X(t)\}$ in the tth time x_{1t}, giving weight at w_1, the predictions of GPS-EO-ABC ELM x_{2t}, giving weight at w_2. Let the error indices in the tth time e_t, the corresponding ARIMA error indices e_{1t}, the GPS-EO-ABC ELM error indices e_{2t}. Thus, the proposed hybrid model used greedy selection to obtain final forecasts can be written as:

$$min \sum_{j}^{2}\sum_{t=1}^{N} w_j e_{jt}, \quad j = 1, 2$$

which illustrated the proposed model optimization is a quadratic function, the model combined EWT, ARIMA, and GPS-EO-ABC ELM is a convex programming function.

Definition 11: Let $X = (x_1, x_2, \cdots, x_N)^T$ the points in N-dimensional Euclidean space E, X' one local optima of convex programming function

$$\begin{cases} \min f(X), X \in R \subset E \\ R = \{X|g_j(X) \geq 0, j = 1, 2, \cdots, l\}, \end{cases}$$

J the set of all active constraint subscripts of X, P the feasible direction of X, $f(X)$ is differential at X', and

$$\begin{cases} g_j(X) \text{ is differential at } X' \quad j \in J \\ g_j(X) \text{ is continuity at } X' \quad j \notin J. \end{cases}$$

There is no feasible descent direction at X', thus P is not existed that at the same time satisfied:

$$\begin{cases} \nabla f(X')^{\mathrm{T}}P < 0 \\ \nabla g_j(X')^{\mathrm{T}}P > 0 \end{cases} j \in J$$

Definition 12: Let $A_1, \cdots, A_l$ be l n-dimensional vector, the necessary and sufficient conditions for no feasible descent direction P made $A_j^T P < 0$ hold is existed not all zero numbers $\mu_0, \mu_1, \cdots, \mu_l$ that make $\sum_{j=1}^{l} \mu_j A_j$ hold.

Definition 13: Let X' one local optima of convex programming function:

$$\begin{cases} \min f(X), X \in R \subset E \\ R = \{X|g_j(X) \geq 0, j = 1, 2, \cdots, l\}, \end{cases}$$

$f(X)$ and $g_j(X)$ have continuous first derivative in X'. There must be are not all zero numbers $\mu_0, \mu_1, \cdots, \mu_l$ make

$$\begin{cases} \mu_0\nabla f(X') - \sum_{j=1}^{l} \nabla g_j(X') = 0 \\ \mu_j g_j(X') = 0, j = 1, 2, \cdots, l \\ \mu_j \geq 0, j = 1, 2, \cdots, l \end{cases} \text{hold.}$$

Theorem 14: The convex programming function combined EWT with ARIMA and GPS-EO-ABC ELM must have

local optima X' and not all zero numbers $\mu_0, \mu_1, \cdots, \mu_l$ that make

$$\begin{cases} \mu_0\nabla f(X') - \sum_{j=1}^{l} \nabla g_j(X') = 0 \\ \mu_j g_j(X') = 0, j = 1, 2, \cdots, l \\ \mu_j \geq 0, j = 1, 2, \cdots, l \end{cases} \text{hold.}$$

Proof: Let X' the local optima of proposed model. According to Theorem 10, the hybrid model combined EWT with ARIMA and GPS-EO-ABC ELM is a convex programming function. According to Definition 11, there is no feasible descent direction P at X' satisfied:

$$\begin{cases} \nabla f(X')^{\mathrm{T}}P < 0 \\ -\nabla g_j(X')^{\mathrm{T}}P < 0\ j \in J \end{cases}$$

the necessary and sufficient conditions for $\nabla f(X')^{\mathrm{T}}P < 0$ and $-\nabla g_j(X')^{\mathrm{T}}P < 0$ is existed not all zero numbers $\mu_0 \geq 0$ and $\mu_j \geq 0\ (j \in J)$ made

$$\mu_0\nabla f(X') - \sum_{j\in J} \nabla g_j(X') = 0 \text{ hold},$$

for $j \notin J$, let corresponding $\mu_j = 0$, thus

$$\begin{cases} \mu_0\nabla f(X') - \sum_{j=1}^{l} \nabla g_j(X') = 0 \\ \mu_j g_j(X') = 0, j = 1, 2, \cdots, l \\ \mu_j \geq 0, j = 1, 2, \cdots, l. \end{cases}$$

According to Definition 13, prove up.

REFERENCES
[1] R. S. Tsay, *Analysis of Financial Time Series*. Hoboken, NJ, USA: Wiley, 2005.

[2] S. Yanfeng, L. Yanchun, J. Jingqing, and W. Chunguo, "Neural network method in financial time series prediction," *J. Jilin Univ. (Inf. Sci. Ed.)*, vol. 1, pp. 49–52, 2004.

[3] R. C. Cavalcante, R. C. Brasileiro, V. L. Souza, J. P. Nobrega, and A. L. Oliveira, "Computational intelligence and financial markets: A survey and future directions," *Expert Syst. Appl.*, vol. 55, pp. 194–211, Aug. 2016.

[4] A. A. Ariyo, A. O. Adewumi, and C. K. Ayo, "Stock price prediction using the ARIMA model," in *Proc. UKSim-AMSS 16th Int. Conf. Comput. Model. Simul.*, Mar. 2014, pp. 106–112.

[5] D. M. Al-Najjar, "Modelling and estimation of volatility using ARCH/GARCH models in Jordan's stock market," *Asian J. Finance Accounting*, vol. 8, no. 1, p. 152, 2016.

[6] G. P. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model," *Neurocomputing*, vol. 50, pp. 159–175, Jan. 2003.

[7] X. Yu and J. Zhang, "A comparison of hybrid ARMA-Elman models with single models for forecasting interest rates," in *Proc. 2nd Int. Symp. Intell. Inf. Technol. Appl.*, Dec. 2008, pp. 985–989.

[8] F. Zheng and S. Zhong, "Time series forecasting using an ensemble model incorporating ARIMA and ANN based on combined objectives," in *Proc. 2nd Int. Conf. Artif. Intell., Manage. Sci. Electron. Commerce (AIMSEC)*, Aug. 2011, pp. 2671–2674.

[9] J. M. Bates and C. W. J. Granger, "The combination of forecasts," *J. Oper. Res. Soc.*, vol. 20, no. 4, p. 451, Dec. 1969.

[10] L. Xiong and Y. Lu, "Hybrid ARIMA-BPNN model for time series prediction of the chinese stock market," in *Proc. 3rd Int. Conf. Inf. Manage. (ICIM)*, Apr. 2017, pp. 93–97.

[11] D. S. de O. Santos Júnior, J. F. L. de Oliveira, and P. S. G. de Mattos Neto, "An intelligent hybridization of ARIMA with machine learning models for time series forecasting," *Knowl.-Based Syst.*, vol. 175, pp. 72–86, Jul. 2019.

[12] H. K. Choi, "Stock price correlation coefficient prediction with ARIMA-LSTM hybrid model," 2018, *arXiv:1808.01560*. [Online]. Available: http://arxiv.org/abs/1808.01560

[13] S. S. Chung and S. Zhang, "Volatility estimation using support vector machine: Applications to major foreign exchange rates," *Electron. J. Appl. Stat. Anal.*, vol. 10, no. 2, pp. 499–511, 2017.

[14] F. Zhu, W. Quan, Z. Zheng, and S. Wan, "A Bayesian learning method for financial time-series analysis," *IEEE Access*, vol. 6, pp. 38959–38966, 2018.

[15] W. Bao, J. Yue, and Y. Rao, "A deep learning framework for financial time series using stacked autoencoders and long-short term memory," *PLoS ONE*, vol. 12, no. 7, 2017, Art. no. e0180944.

[16] M. Al Wadia and M. T. Ismail, "Selecting wavelet transforms model in forecasting financial time series data based on ARIMA model," *Appl. Math. Sci.*, vol. 5, no. 7, pp. 315–326, 2011.

[17] M. S. Raimundo and J. Okamoto, "SVR-wavelet adaptive model for forecasting financial time series," in *Proc. Int. Conf. Inf. Comput. Technol. (ICICT)*, Mar. 2018, pp. 111–114.

[18] A. M. Awajan, M. T. Ismail, and S. Al Wadi, "A hybrid EMD-MA for forecasting stock market index," *Italian J. Pure Appl. Math.*, vol. 38, no. 1, pp. 1–20, 2017.

[19] V. Ravi, D. Pradeepkumar, and K. Deb, "Financial time series prediction using hybrids of chaos theory, multi-layer perceptron and multi-objective evolutionary algorithms," *Swarm Evol. Comput.*, vol. 36, pp. 136–149, Oct. 2017.

[20] C.-F. Huang, "A hybrid stock selection model using genetic algorithms and support vector regression," *Appl. Soft Comput.*, vol. 12, no. 2, pp. 807–818, Feb. 2012.

[21] A. Bagheri, H. Mohammadi Peyhani, and M. Akbari, "Financial forecasting using ANFIS networks with quantum-behaved particle swarm optimization," *Expert Syst. Appl.*, vol. 41, no. 14, pp. 6235–6250, Oct. 2014.

[22] D. Pradeepkumar and V. Ravi, "Forecasting financial time series volatility using particle swarm optimization trained quantile regression neural network," *Appl. Soft Comput.*, vol. 58, pp. 35–52, Sep. 2017.

[23] A. Kazem, E. Sharifi, F. K. Hussain, M. Saberi, and O. K. Hussain, "Support vector regression with chaos-based firefly algorithm for stock market price forecasting," *Appl. Soft Comput.*, vol. 13, no. 2, pp. 947–958, Feb. 2013.

[24] Y. Y. Liu, P. Thulasiraman, and R. K. Thulasiram, "Parallelizing active memory ants with MapReduce for clustering financial time series data," in *Proc. IEEE Int. Conferences Big Data Cloud Comput. (BDCloud), Social Comput. Netw. (SocialCom), Sustain. Comput. Commun. (SustainCom) (BDCloud-SocialCom-SustainCom)*, Oct. 2016, pp. 137–144.

[25] D. Karaboga, "An idea based on honey bee swarm for numerical optimization," Eng. Fac. Comput. Eng., Erciyes Univ., Kayseri, Turkey, Tech. Rep. TR06, 2005.

[26] D. Karaboga and B. Basturk, "A powerful and efficient algorithm for numerical function optimization: Artificial bee colony (ABC) algorithm," *J. Global Optim.*, vol. 39, no. 3, pp. 459–471, Oct. 2007.

[27] D. Karaboga and B. Basturk, "On the performance of artificial bee colony (ABC) algorithm," *Appl. Soft Comput.*, vol. 8, no. 1, pp. 687–697, Jan. 2008.

[28] C. Q. Zhang, J. G. Zhang, and X. Wang, "Overview of research on bee colony algorithms," (in Chinese), *Appl. Res. Comput.*, vol. 28, no. 9, pp. 3201–3205 and 3214, 2011.

[29] E. Nourani, A. M. Rahmani, and A. H. Navin, "Forecasting stock prices using a hybrid artificial bee colony based neural network," in *Proc. Int. Conf. Innov. Manage. Technol. Res.*, 2012, pp. 486–490.

[30] R. C. Brasileiro, V. L. F. Souza, B. J. T. Fernandes, and A. L. I. Oliveira, "Automatic method for stock trading combining technical analysis and the artificial bee colony algorithm," in *Proc. IEEE Congr. Evol. Comput.*, Jun. 2013, pp. 1810–1817.

[31] O. Hegazy, O. S. Soliman, and M. Abdul Salam, "LSSVM-ABC algorithm for stock price prediction," 2014, *arXiv:1402.6366*. [Online]. Available: http://arxiv.org/abs/1402.6366

[32] X. J. Bi and W. Y. J, "Improved artificial bee colony algorithm," (in Chinese), *J. Harbin Univ. Eng.*, vol. 33, no. 1, pp. 117–123, 2012.

[33] X.-P. Liu, S.-B. Xuan, and F. Liu, "Artificial bee colony algorithm with good point set and turn process of monkey algorithm," (in Chinese), *Pattern Recognt. Artif. Intell.*, vol. 7, no. 2, pp. 80–89, 2015.

[34] J. Qiu, M. Xu, M. Liu, W. Xu, J. Wang, and S. Su, "A novel search strategy based on gradient and distribution information for artificial bee colony algorithm," *J. Comput. Methods Sci. Eng.*, vol. 17, no. 3, pp. 377–395, Aug. 2017.

[35] J. Kennedy, "Swarm intelligence," in *Handbook of Nature-Inspired and Innovative Computing*. Springer, 2006, pp. 187–219.

[36] Q. Zhang and A. Benveniste, "Wavelet networks," *IEEE Trans. Neural Netw.*, vol. 3, no. 6, pp. 889–898, Nov. 1992.

[37] J. Gilles, "Empirical wavelet transform," *IEEE Trans. Signal Process.*, vol. 61, no. 16, pp. 3999–4010, Aug. 2013.

[38] W. Liu, S. Cao, and Y. Chen, "Seismic time–frequency analysis via empirical wavelet transform," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 1, pp. 28–32, Jan. 2016.

[39] M. İ. Gursoy, S. V. Ustun, and A. S. Yilmaz, "An efficient DWT and EWT feature extraction methods for classification of real data PQ disturbances," *Uluslararası Mühendislik Araştirma Ve Geliştirme Dergisi*, vol. 10, no. 1, pp. 158–171, Jan. 2018.

[40] 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.

[41] G.-B. Huang, H. Zhou, X. Ding, and R. Zhang, "Extreme learning machine for regression and multiclass classification," *IEEE Trans. Syst., Man, Cybern. B. Cybern.*, vol. 42, no. 2, pp. 513–529, Apr. 2012.

[42] S. K. Sunori, P. Kumar Juneja, M. Chaturvedi, A. Rai, and A. Bansal, "ANN modeling for predicting time series," in *Proc. Int. Conf. Adv. Comput., Commun. Control Netw. (ICACCCN)*, Oct. 2018, pp. 792–794.

[43] R. M. Kapila Tharanga Rathnayaka, D. M. K. N. Seneviratna, W. Jianguo, and H. I. Arumawadu, "A hybrid statistical approach for stock market forecasting based on artificial neural network and ARIMA time series models," in *Proc. Int. Conf. Behav., Econ. Socio-Cultural Comput. (BESC)*, Oct. 2015, pp. 54–60.

[44] X. Chen, Z. Y. Dong, K. Meng, Y. Xu, K. P. Wong, and H. W. Ngan, "Electricity price forecasting with extreme learning machine and bootstrapping," *IEEE Trans. Power Syst.*, vol. 27, no. 4, pp. 2055–2062, Nov. 2012.

[45] J. Quilty and J. Adamowski, "Addressing the incorrect usage of wavelet-based hydrological and water resources forecasting models for real-world applications with best practices and a new forecasting framework," *J. Hydrol.*, vol. 563, pp. 336–353, Aug. 2018.

[46] J. Hu and J. Wang, "Short-term wind speed prediction using empirical wavelet transform and Gaussian process regression," *Energy*, vol. 93, pp. 1456–1466, Dec. 2015.

[47] H. Shayeghi and A. Ghasemi, "Day-ahead electricity prices forecasting by a modified CGSA technique and hybrid WT in LSSVM based scheme," *Energy Convers. Manage.*, vol. 74, pp. 482–491, Oct. 2013.

[48] S. J. Taylor, *Modelling Financial Time Series*. Singapore: World Scientific, 2008.

[49] N. Z. W. J. M. Li, X. H. Zhu, and Y. Xu, "Drought prediction model based on GPS-GSO-BPNN parallel ensemble learning algorithm," (in Chinese), *Syst. Eng. Theory Pract.*, vol. 38, no. 5, pp. 1343–1353, 2018.

[50] Z. B. L. Zhang, "Good point set based genetic algorithm," (in Chinese), *Chin. J. Comput.*, vol. 2001, no. 9, pp. 917–922, 2001.

[51] W. Y. Hua and L. G, *The Application of Number Theory in Approximation*. Beijing, China :Science Press (in Chinese), 1978.

[52] H. R. Tizhoosh, "Opposition-based learning: A new scheme for machine intelligence," in *Proc. Int. Conf. Comput. Intell. Modeling, Control Autom. Int. Conf. Intell. Agents, Web Technol. Internet Commerce (CIMCA-IAWTIC)*, vol. 1, 2005, pp. 695–701.

[53] S. Rahnamayan, H. R. Tizhoosh, and M. M. Salama, "Opposition-based differential evolution," *IEEE Trans. Evol. Comput.*, vol. 12, no. 1, pp. 64–79, Feb. 2008.

[54] H. Sharma, J. C. Bansal, and K. V. Arya, "Opposition based Lévy flight artificial bee colony," *Memetic Comput.*, vol. 5, no. 3, pp. 213–227, Sep. 2013.

[55] Y. Jiang, T. Hu, C. Huang, and X. Wu, "An improved particle swarm optimization algorithm," *Appl. Math. Comput.*, vol. 193, no. 1, pp. 231–239, 2007.

[56] P. Niu, Y. Ma, M. Li, S. Yan, and G. Li, "A kind of parameters self-adjusting extreme learning machine," *Neural Process. Lett.*, vol. 44, no. 3, pp. 813–830, Dec. 2016.

[57] Z. Chu, Y. Ma, and J. Cui, "Adaptive reactionless control strategy via the PSO-ELM algorithm for free-floating space robots during manipulation of unknown objects," *Nonlinear Dyn.*, vol. 91, no. 2, pp. 1321–1335, Jan. 2018.

HE YU received the B.B.A. degree from Anhui Normal University (AHNU), in 2010, and the M.A.Sc. degree from Guangxi University (GXU), in 2013. He is currently working with the Anhui Vocational College of Press and Publishing. He has contributed to the design and verification of the proposed hybrid model. He has published over ten articles in journals. His research interests are in artificial intelligence applications in financial engineering, machine learning algorithms, neural networks and statistical learning, and convolutional neuron networks.

LI JING MING received the Ph.D. degree in management science and engineering from the Hefei University of Technology (HFUT), in 2017. He is currently working with the School of Management Science and Engineering, Anhui University of Finance and Economics (AUFE). He has contributed to the design of parallel algorithms for the combination of swarm intelligence optimization algorithm and neural network. These algorithms are usually used for classification problems and are designed by parallel interaction. In the past five years, he has been involved in the algorithm flow of the combination of swarm intelligence optimization algorithm and machine learning method, which are used for solving the problems of agrometeorological disaster prediction, haze warning, and credit evaluation. He also developed an attribute selection method based on the combination of discrete Glowworm Swarm Optimization algorithm and fractal learning. His main research interests include artificial intelligence, machine learning, and financial engineering.

RUAN SUMEI received the Ph.D. degree in business administration from the Hefei University of Technology, Hefei, China, in 2014. She currently serves as a Professor and a Master Tutor at the Finance School, Anhui University of Finance and Economics (AUFE). She has published over 40 articles in journals. Her main research interests are bank management, corporate governance, and internet finance.

ZHAO SHUPING received the Ph.D. degree in business administration from the Hefei University of Technology, Hefei, China, in 2015. He is currently working management information systems with the Hefei University of Technology. He has developed some machine learning algorithms for classification or prediction in recent years. He has published over ten articles in journals. His main research interests are decision-making theory and data mining.

· · ·